



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

João Manoel Cardoso Martinez

Classificação de Imagens Histológicas Explorando Deep-Learned Features de Representações DeepDream e IA Explicável

São José do Rio Preto
2023

João Manoel Cardoso Martinez

Classificação de Imagens Histológicas Explorando Deep-Learned Features de Representações DeepDream e IA Explicável

Trabalho de Conclusão de Curso (TCC) apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, junto ao Conselho de Curso de Bacharelado em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CNPq - Proc. 04/2021

Orientador:

Prof. Dr. Leandro Alves Neves

**São José do Rio Preto
2023**

M385c

Martinez, João Manoel Cardoso

Classificação de imagens histológicas explorando Deep-Learned Features de representações DeepDream e IA explicável / João Manoel Cardoso

Martinez. -- São José do Rio Preto, 2023

77 p. : il., tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Leandro Alves Neves

1. Redes neurais (Computação). 2. Processamento de imagens. 3. Inteligência artificial. 4. Mamas Câncer. 5. Cólon (Anatomia) Câncer. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

João Manoel Cardoso Martinez

Classificação de Imagens Histológicas Explorando Deep-Learned Features de Representações DeepDream e IA Explicável

Trabalho de Conclusão de Curso (TCC) apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, junto ao Conselho de Curso de Bacharelado em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CNPq - Proc. 04/2021

Comissão Examinadora:

Prof. Dr. Leandro Alves Neves
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Rodrigo Capobianco Guido
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Lucas Correia Ribas
UNESP – Câmpus de São José do Rio Preto

**São José do Rio Preto
23 de Novembro de 2023**

Dedico este trabalho aos meus pais que sempre me deram todo o apoio que possibilitou minhas realizações. Dedico também ao meu irmão, minha cachorrinha Amora e todos os amigos que fiz em meio a essa jornada.

Agradecimentos

Primeiramente, agradeço aos meus pais, Carlos Alberto Martinez e Marinéia Madalena Cardoso Martinez, a quem devo todas as minhas conquistas e minha jornada pelo processo de graduação, impulsionadas por todo o carinho, apoio e confiança depositados em mim.

Direciono também meus agradecimentos aos professores que tive a oportunidade de conhecer e que me inspiraram por seus trabalhos e áreas de atuação, em especial, agradeço o Prof. Dr. Leandro Alves Neves por contribuir de forma significativa com meu crescimento acadêmico e profissional, sendo professor e também amigo durante a graduação, e por proporcionar, como orientador, o desenvolvimento de meus trabalhos científicos.

Estendo também meus agradecimentos para meu colega Leonardo Henrique da Costa Longo, pelo suporte e compartilhamento dos conhecimentos acerca dos principais conceitos que envolvem a área de aprendizado profundo, nos âmbitos teórico e prático.

Finalmente, agradeço ao apoio financeiro fornecido pelo CNPq e PROPe para o desenvolvimento deste trabalho.

"I am and always will be the optimist. The hoper of far-flung hopes,
and the dreamer of improbable dreams."

- 11^o Doutor - Doctor Who

Resumo

A aplicação de métodos de aprendizado de máquina para a análise de imagens histológicas tem proporcionado avanços significativos em diagnósticos auxiliados por computador. Neste trabalho, um estudo foi realizado para classificar padrões existentes em amostras histológicas coradas com hematoxilina-eosina, representativas de câncer de mama, câncer colorretal, tecido hepático e displasia oral. O método proposto explorou redes neurais convolucionais (VGG19, ResNet50, Inception V3 e Efficientnet B2), *gradient-weighted class activation mapping*, explicações por modelo agnóstico de interpretação local e representações baseadas no efeito inceptionism (DeepDream). As *deep features* foram analisadas via múltiplos classificadores (*rotation forest*, *multilayer perceptron*, *logistic*, *random forest*, *decorate*, *k-nearest neighbor*, *KStar* e *support vector machine*) com o intuito de indicar as melhores associações. A melhor solução foi obtida por meio da rede ResNet50 com representações extraídas da VGG19, explorando vetores de características com 40 e 100 atributos. Os resultados apresentaram taxas de AUC variando de 0.994 (UCSB) a 1 (CR). As *deep features* mais exploradas foram obtidas por meio das imagens originais (maior ocorrência), seguidas pelas representações geradas por explicações de modelo agnóstico de interpretação local, DeepDream e *gradient-weighted class activation mapping*. Estes resultados podem apoiar o desenvolvimento de técnicas voltadas ao reconhecimento de padrões de imagens H&E, com maior robustez em diagnósticos auxiliados por computador.

Palavras-chave: Redes neurais convolucionais, *deep features*, reconhecimento de padrões, representações DeepDream, XAI, imagens histológicas

Abstract

The application of machine learning methods to analyze histological images has led to significant advances in computer-aided diagnosis. In this work, a study was carried out to classify patterns present in histological samples stained with hematoxylin-eosin, representative of breast cancer, colorectal cancer, liver tissue and oral dysplasia. The proposed method explored convolutional neural networks (VGG19, ResNet50, InceptionV3 and EfficientNet B2), gradient-weighted class activation mapping, local interpretable model-agnostic explanations and representations based on the inceptionism effect (DeepDream). The deep features were analyzed via multiple classifiers (rotation forest, multilayer perceptron, logistic, random forest, decorate, k-nearest neighbor, KStar and support vector machine) in order to indicate the main associations. The best solution was obtained through the ResNet50 with the representations from the VGG19, exploring feature vectors with 40 and 100 attributes. The results were AUC rates ranging from 0.994 (UCSB) to 1 (CR). The most explored deep features were obtained via original images (highest occurrence), followed by local interpretable model-agnostic explanations, DeepDream and gradient-weighted class activation mapping representations. These results can support the development of techniques addressed to pattern recognition of H&E images, with more robust computer-aided diagnoses.

Keywords: *Convolutional neural networks, deep features, pattern recognition, DeepDream representations, XAI, histological images*

Lista de Figuras

1	Arquitetura típica de CNN com as saídas de cada camada.	22
2	Representação da arquitetura do modelo de rede VGG19.	26
3	Representação de um bloco residual na arquitetura <i>ResNet-50</i>	27
4	Representação da arquitetura do modelo de rede <i>ResNet-50</i>	28
5	Representação da arquitetura do modelo de rede <i>EfficientNet-b2</i>	30
6	Representação da arquitetura do modelo de rede <i>InceptionV3</i>	31
7	Visão geral da estratégia proposta para investigar imagens H&E.	47
8	Amostras H&E das bases: UCSB benigno (a) e maligno (b); CR benigno (c) e maligno (d); LG macho (e) e fêmea (f); DISP saudável (g) e severo (h).	49
9	Representações Grad-CAM das imagens H&E originais (a), explorando o modelo VGG19 (b), modelo ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).	52
10	Representações LIME das imagens H&E originais (a), considerando o modelo VGG19 (b), modelo ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).	53
11	Representações DeepDream via imagens H&E originais (a), considerando o modelo VGG19 (b), ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).	55

Lista de Tabelas

1	Representação esquemática de uma matriz de confusão	39
2	Valores de ACC (%) alcançados via modelos CNN aplicados diretamente nos conjuntos de imagens histológicas.	57
3	Valores de acurácia (%) via ResNet50 (CNN 2) com representações extraídas de VGG19 (CNN 1).	58
4	Valores de acurácia (%) da aplicação de Inception V3 (CNN 2), usando representações extraídas de Efficientnet B2 (CNN 1).	58
5	Valores de acurácia (%) com VGG19 (CNN 2) e representações extraídas de ResNet50 (CNN 1).	58
6	Valores de acurácia (%) obtidos por Efficientnet B2 (CNN 2), explorando as representações extraídas de Inception V3 (CNN 1).	58
7	Média de taxas AUC obtidas com múltiplos classificadores e <i>deep features</i> por meio da arquitetura ResNet50 (CNN 2), usando representações extraídas de VGG19 (CNN 1).	59
8	Média de taxas AUC obtidas com múltiplos classificadores e <i>deep features</i> via Inception V3 (CNN 2), processando as representações extraídas de Efficientnet B2 (CNN 1).	60
9	Médias de taxas AUC considerando múltiplos classificadores e <i>deep features</i> extraídas de VGG19 (CNN 2), usando as representações obtidas por ResNet50 (CNN 1).	60

10	Médias de taxas AUC obtidas com múltiplos classificadores e <i>deep features</i> via Efficientnet B2 (CNN 2), explorando as representações extraídas de Inception V3 (CNN 1).	60
11	Sumário das melhores associações resultantes do modelo proposto. . .	61
12	Porcentagem (%) da distribuição de atributos que compuseram as melhores soluções.	62
13	Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de imagens de câncer de mama.	63
14	Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de câncer colorretal.	64
15	Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de imagens de tecido hepático.	64
16	Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de displasia oral.	64

Lista de Abreviações

ACC Acurácia

AUC Área sob a Curva ROC

CAM Class Activation Mapping

CNN Convolutional Neural Network

DD DeepDream

Dec Decorate

FN Falso Negativo

FP Falso Positivo

Grad-CAM Gradient-weighted Class Activation Mapping

H&E Hematoxilina-Eosina

IBk Instance Based k

ILSVRC ImageNet Large Scale Visual Recognition Challenge

LIME Locally-Interpretable Model-agnostic Explanation

Log Logistic

MP Multilayer Perceptron

RandF Random Forest

ReLU Rectified Linear Activation Function

ResNet Residual Neural Network

ROC Receiver Operating Characteristic

RotF Rotation Forest

SVM Support Vector Machine

VGGNet Visual Geometry Group Network

VN Verdadeiro Negativo

VP Verdadeiro Positivo

XAI Explainable Artificial Intelligence

Sumário

1	Introdução	15
1.1	Justificativas e Motivação	16
1.2	Objetivos	18
1.3	Organização do Projeto	19
2	Fundamentação Teórica	20
2.1	Redes Neurais Convolucionais e Conceitos	20
2.1.1	Funções de Ativação	21
2.1.2	Problema do Gradiente de Fuga	23
2.1.3	O problema do <i>Overfitting</i>	23
2.1.4	Transferência de Aprendizado	24
2.2	Modelos de CNNs	25
2.2.1	VGGNet	25
2.2.2	ResNet	26
2.2.3	EficientNet	27
2.2.4	InceptionNet	29
2.3	Técnicas de Inteligência Artificial Explicável e DeepDream	30
2.3.1	Grad-CAM	31
2.3.2	LIME	32
2.4	DeepDream	32
2.5	Extração de Características	33

2.6	Seleção das Melhores Características Via Algoritmo ReliefF	34
2.7	Classificação de Características	35
2.7.1	Baseados em Função	35
2.7.2	Baseados em Árvores de Decisão	36
2.7.3	Baseados em <i>Meta-Learning</i>	37
2.7.4	Baseados em <i>Lazy Learning</i>	38
2.8	Métricas para Avaliação de Desempenho	38
2.8.1	Acurácia (ACC)	39
2.8.2	Medida-F	40
2.8.3	Área sob a Curvatura ROC (AUC)	41
3	Revisão Bibliográfica	42
4	Metodologia	46
4.1	Estruturação do Esquema	46
4.2	Contexto de aplicação: imagens histológicas	47
4.3	Etapa 1 - Aplicação da Rede Neural Convolucional de Extração de Representações XAI	50
4.4	Etapa 2 - Extração das representações com as técnicas DD, LIME e Grad-CAM	50
4.4.1	Técnica Grad-CAM	51
4.4.2	Técnica LIME	52
4.4.3	Técnica DeepDream	54
4.5	Etapa 3 - Extração das <i>Deep Features</i> das Representações Grad-CAM, LIME e DD	54
4.6	Etapa 4 - Seleção das <i>Features</i> Ranqueadas via Algoritmo <i>ReliefF</i>	56
4.7	Etapa 5 - Classificações Baseadas em Múltiplos Classificadores	56
5	Resultados e Discussão	57

6	Conclusões	65
6.1	Contribuições e Divulgação	66
	Referências	68

Introdução

Na área de análise de imagens, o uso de técnicas para extrair atributos requer condições específicas para o processamento de dados naturais em sua forma bruta. Com isso, por décadas, a construção de um sistema de aprendizado de máquina exigiu uma engenharia cuidadosa para definir os melhores atributos para processo de classificação de padrões. Parte das dificuldades encontradas nesse processo foi minimizada com técnicas baseadas no conceito de aprendizado profundo, especialmente a partir de redes neurais convolucionais (do inglês Convolutional Neural Network (CNN)) (LECUN; BENGIO; HINTON, 2015).

É importante destacar que variações na arquitetura de uma CNN ou a exploração de seu poder de abstração, por meio da extração de *deep features* com um classificador externo à rede, são estratégias que podem resultar em modelos eficientes e computacionalmente acessíveis (DABEER; KHAN; ISLAM, 2019). Estudos presentes na literatura evidenciam que modelos híbridos, uso de seleção de atributos a partir de CNN com classificadores externos à rede, podem apresentar resultados iguais ou até melhores do que os fornecidos por um modelo específico de CNN (COCCIA, 2020). Nas abordagens híbridas, principalmente quando os classificadores são utilizados para categorizar *deep features*, uma etapa é a escolha dos atributos mais relevantes. Assim, os modelos disponíveis na literatura com foco em transferência de aprendizado profundo utilizam métodos supervisionados para realizar ranqueamentos e, consequentemente, identificar com eficácia as características mais importantes para a etapa de classifica-

ção.

Apesar dos avanços indicados previamente, ainda é possível investigar o poder discriminativo de modelos híbridos e das *deep features* a partir de diferentes formas de representações de imagens (ADADI; BERRADA, 2018), tais como as desenvolvidas para explicar os padrões de ativação neural de uma CNN (MAHENDRAN; VEDALDI, 2016; VEDALDI; ZISSERMAN, 2013; YOSINSKI et al., 2015). Assim, é possível utilizar os valores presentes na camada "*average pooling*", aplicada como uma estratégia regularizadora estrutural de uma CNN, para conhecer as regiões das imagens que embasaram a classificação. Técnicas relevantes podem ser exploradas para auxiliar no processo de classificação de uma CNN, tais como *gradient-weighted class activation mapping* (Grad-CAM) (RAJARAMAN et al., 2018a; REYES et al., 2020), *locally-interpretable model-agnostic explanation* (LIME) (RAJARAMAN et al., 2018a; REYES et al., 2020; SOUSA; VELLASCO; SILVA, 2019) e *DeepDream* (DD) (TOĞAÇAR; CÖMERT; ERGEN, 2021; MORDVINTSEV; OLAH; TYKA, 2015; SUZUKI et al., 2017). Neste cenário, a investigação de imagens histológicas coradas com Hematoxilina-Eosina (H&E) foi pouco explorada a partir de um modelo híbrido considerando representações de imagens via LIME, Grad-CAM e DD. Essas associações são contribuições relevantes para a classificação e o reconhecimento de padrões de doenças usando imagens H&E, bem como para a área de aprendizado de máquina. Os conhecimentos fornecidos por meio deste trabalho podem embasar o desenvolvimento de sistemas computacionais mais completos.

1.1 Justificativas e Motivação

O câncer de mama e colorretal são extremamente comuns, com milhares de novos casos sendo descobertos a cada ano. Segundo o Instituto Nacional de Câncer (INCA), a estimativa para a descoberta de novos casos de câncer de mama no Brasil é por volta de 73610 casos, sendo 66,54 novos casos a cada 100 mil mulheres, de acordo com le-

vantamento feito no ano de 2021 (Instituto Nacional de Câncer José Alencar Gomes da Silva, 2023). O câncer colorretal ocupa a terceira posição entre os tumores mais incidentes, alcançando estimativas de 21,10 casos por 100 mil pessoas (Instituto Nacional de Câncer José Alencar Gomes da Silva, 2023). Somente no Brasil, as estimativas para novas incidências desses tumores, entre os anos de 2023 a 2025, são de 73610 casos de câncer de mama, e 45630 casos de câncer colorretal (Instituto Nacional de Câncer José Alencar Gomes da Silva, 2023). Sendo assim, a análise imagens histológicas dos cânceres de mama e colorretal apresentam grande potencial e relevância para contribuir significativamente no âmbito da saúde, ajudando no desenvolvimento de sistemas de diagnóstico e visão computacional mais precisos e eficientes.

A displasia também se destaca como uma condição emergente e potencialmente responsável pela formação de neoplasias (tumores) malignos. Essa condição é caracterizada pelo desenvolvimento celular anormal e origem de más formações de tecidos e órgãos em todas as partes do corpo, tanto em humanos como em animais.

É importante destacar que os diagnósticos das doenças referidas decorrem de especialistas do âmbito da saúde, a partir de amostras de tecidos das regiões afetadas. A análise acontece ao nível celular, sendo necessária a utilização de microscópios para visualizar tecidos que, a princípio, são incolores. Técnicas de coloração devem ser aplicadas a fim de facilitar o reconhecimento visual das estruturas observadas, sendo a Hematoxilina & Eosina (H&E) a mais conhecida dentre elas. As amostras submetidas a essa técnica são comumente denominadas de imagens H&E.

A atuação humana na análise de imagens H&E é caracterizada por ser um potencial motivo de problemáticas no processo de diagnóstico. Questões como a subjetividade de interpretação e fadiga são exemplos que interferem nos estudos clínicos. Nesse contexto, os avanços das ferramentas computacionais no armazenamento de dados para investigação com altas resoluções e processamento de imagens por meio de abordagens de aprendizado de máquina, são fatores motivadores do uso de sistemas computacio-

nais no auxílio aos processos de diagnóstico e para a histopatologia em geral. Através da análise de imagens via *machine learning*, tarefas como a identificação de padrões, muitos imperceptíveis ao olho humano, são facilitados (GREENSPAN; GINNEKEN; SUMMERS, 2016; LITJENS et al., 2017). Contudo, por mais que o uso de metodologias *deep learning* em sistemas de inteligência artificial consigam desempenhar tarefas complexas nos mais variados campos, inclusive na classificação de imagens histológicas, ainda há uma carência na transparência das informações sobre o modo como os modelos chegam a uma predição durante as fases de teste, o que se caracteriza como um obstáculo nas aplicações médicas (SAMEK; WIEGAND; MÜLLER, 2017). A fim de superar esse desafio, métodos para visualizar, explicar e interpretar esses modelos de aprendizado profundo têm atraído a atenção para garantir a interpretabilidade em soluções pautadas em inteligência artificial.

Dado o contexto, o uso de CNNs em associação com técnicas para garantir a transparência de IAs e de extração de características abre uma ampla frente de possibilidades de resultados mais promissores nas mais diversas aplicações. Contudo, uma abordagem que explore a combinação de tais métodos e técnicas com a análise de *features* de aprendizado profundo via diversos classificadores ainda foi pouco testada na literatura. A contribuição deste tipo de investigação pode ser substancial para a classificação e reconhecimento de padrões de doenças, como o câncer de mama, câncer colorretal e displasia, assim como para lançar uma luz no modo como os modelos realizam suas escolhas.

1.2 Objetivos

Neste projeto, investigamos a capacidade discriminativa de um modelo híbrido considerando *deep features* geradas a partir de representações visuais obtidas com as técnicas LIME, Grad-CAM e DD. Para tanto, os objetivos específicos foram:

- Gerar diferentes representações visuais de imagens histológicas H&E com as técnicas LIME, Grad-CAM e DD, aplicadas a partir de uma arquitetura CNN;
- Executar o modelo CNN considerando como entradas as representações obtidas na etapa anterior a fim de coletar as *deep features* correspondentes;
- Avaliar o poder discriminativo das *deep features*, explorando diferentes composições e classificadores da área de inteligência artificial;
- Indicar as associações mais apropriadas para classificar e reconhecer padrões nas imagens histológicas exploradas neste trabalho.

1.3 Organização do Projeto

Os principais conceitos para o entendimento do projeto são abordados no Capítulo 2, com informações sobre Redes Neurais Convolucionais, técnicas de IA Explicável e DeepDream, extração e seleção de características, assim como detalhes sobre as métricas de desempenho analisadas no desenvolvimento do trabalho. O Capítulo 3 expõe os trabalhos que serviram de base para os conceitos sobre redes neurais, técnicas de IA explicável (XAI) e DeepDream e suas associações com CNNs, além de considerações sobre trabalhos correlatos. A metodologia e o banco de imagens utilizados para o desenvolvimento do projeto estão descritos no Capítulo 4. Os resultados obtidos pelo modelo proposto são apresentados no Capítulo 5. Por fim, as conclusões sobre o modelo proposto, assim como suas contribuições e meios de divulgação são elencados no Capítulo 6.

Fundamentação Teórica

Neste capítulo são abordados os principais conceitos e conhecimentos necessários para o entendimento do projeto proposto, dos quais podem ser citados as Redes Neurais Convolucionais utilizadas, técnicas de Inteligência Artificial Explicável usadas no processo de extração, além de ranqueamento e classificação de *features*, e métricas para avaliação dos modelos de aprendizado de máquina.

2.1 Redes Neurais Convolucionais e Conceitos

As Redes Neurais Convolucionais, ou Convolutional Neural Networks (CNNs), foram desenvolvidas para processar dados estruturados como, por exemplo, matrizes que representam imagens coloridas com vetores 2D com intensidades de pixels (LECUN; BENGIO; HINTON, 2015). A topologia de uma CNN, em suma, é baseada em três princípios-chave: campos receptivos locais, pesos compartilhados e amostragem espacial ou temporal (LECUN et al., 1998). Tal definição está relacionada às várias camadas convolucionais que uma CNN possui, o que possibilita a extração de características na rede.

Na Figura 1 é ilustrado o esquema de arquitetura convencional de uma CNN, a qual possui, inicialmente, algumas camadas convolucionais intercaladas com camadas *pooling* e, por fim, apresenta camadas totalmente conectadas (ZHOU; GREENSPAN; SHEN, 2017). Para fim de entendimento, camadas de convolução desempenham o

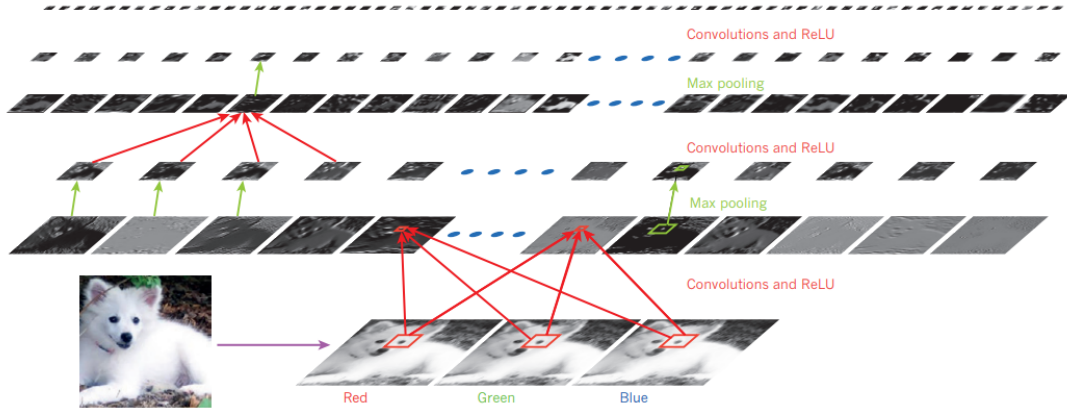
papel de extrair características locais da matriz de entrada. Tal ação ocorre por meio de *Kernels*, que são núcleos de aprendizado presentes na camada que calculam novos mapas de características. Segue-se que o nível de complexidade das características extraídas nas camadas é diretamente proporcional ao número de camadas, ou seja, as primeiras camadas convolucionais extraem características mais simples, como cantos, bordas e texturas, enquanto as características de maior complexidade são extraídas nas próximas camadas convolucionais, com destaque para a última, onde as características de mais alto nível são extraídas (SANTOS; PONTI, 2019). As camadas de *pooling* buscam reduzir a complexidade computacional da rede, a partir do gradual encolhimento das dimensões espaciais, reduzindo assim a resolução de mapas de características anteriores (AFFONSO et al., 2017), o que possibilita que padrões mais relevantes sejam o foco da rede (NAMATÊVS, 2017). Há duas abordagens envolvidas na aplicação de *pooling*. ***max pooling*** uma estratégia que seleciona o máximo valor de cada camada *pooling* para reter as características mais proeminentes do mapa; e ***average pooling*** que visa reter os valores médios de características do mapa, suavizando a imagem sem perder a essência das características na imagem. Tais métodos combinados possibilitam a localização do problema de *overfitting* que tende a ocorrer em CNNs (YU et al., 2014). O conceito de *overfitting* é abordado mais adiante na seção 2.1.3

2.1.1 Funções de Ativação

As funções de ativação são caracterizadas por funções matemáticas importantes para o controle da amplitude do sinal de saída de cada neurônio em um intervalo finito, calculando a relevância da informação dada de entrada em cada neurônio e camada nas redes neurais (KARLIK; OLGAC, 2011). A seguir, alguns exemplos de funções de ativação comumente utilizadas em CNNs.

- ReLU

Figura 1: Arquitetura típica de CNN com as saídas de cada camada.



Fonte: (LECUN; BENGIO; HINTON, 2015)

A função ReLU (unidade linear retificada), presente na Equação 2.1, necessita de baixo custo computacional a medida que ativa apenas neurônios com entradas positivas.

$$ReLU(x) = \max(0, x) \quad (2.1)$$

- *Sigmoid*

A Equação 2.2 denota a função Sigmoid, recorrente em problemas que envolvem classificação binária, já que assume apenas os valores 0 e 1.

$$Sigmoid(x) = \frac{1}{1 + \exp(-x)} \quad (2.2)$$

- *Softmax*

Esta função de ativação, presente na Equação 2.3, trata-se de uma função sigmoidal, que atua transformando as saídas do classificador em valores entre 0 e 1

e divide pela soma das saídas, obtendo como resultado a probabilidade de uma entrada pertencer a uma determinada classe.

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (2.3)$$

2.1.2 Problema do Gradiente de Fuga

Para alcançar melhores taxas de precisão de classificação em CNNs, o aumento da quantidade de camadas convolucionais é uma das estratégias mais comuns. Porém, por mais que haja certa melhora na precisão de classificação da rede, uma série de problemas que impactam a fase de treinamento da rede pode ocorrer. Um desses problemas é denominado como gradiente de fuga (HOCHREITER, 1998).

O gradiente de fuga é caracterizado pela ocorrência de valores muito pequenos para a atualização de pesos em conexões situadas ao início de uma CNN, o que acaba por prejudicar o reforço de conexões relevantes, o que consequentemente leva a características pouco distintas que são extraídas e desfavorecem o poder de classificação da rede.

2.1.3 O problema do *Overfitting*

O *Overfitting*, traduzido para português como sobreajuste, indica o caso de um modelo que se ajusta excessivamente bem a um único conjunto de dados já observado. Em outras palavras, a condição de sobreajuste acontece quando o modelo não generaliza de modo eficiente dados já observados (dados de treinamento) com dados não observados (dados de teste) (YING, 2019), promovendo altíssimas taxas de validação que não representam a capacidade de aplicação do modelo ao conjunto de dados de teste. Casos que são passíveis de ocorrência de tal problema podem estar relacionados

ao tamanho do banco de imagens escolhido, no qual o baixo número de representações no banco acaba por afetar as fases de treinamento e teste do modelo de aprendizado de máquina, sendo que, geralmente, para evitar a ocorrência de *overfitting* são necessárias milhares de imagens.

Existem alguns métodos que auxiliam na redução de *overfitting* em CNNs, dos quais podem se destacar, por exemplo, *Batch Normalization* e *Dropout*.

- *Batch Normalization*

Método baseado na normalização dos pesos obtidos através da saída de uma camada de rede, realizando o processo de retro propagação, *back-propagation* em inglês, dos gradientes por meio dos parâmetros normalizados consequentemente.

- *Dropout*

A técnica *dropout* (SRIVASTAVA et al., 2014) consiste na desativação de neurônios de forma aleatória baseada em uma probabilidade pré-definida. A técnica busca reduzir a complexidade de ligações e dependências entre neurônios, para assim poder mitigar a ocorrência de *overfitting* e melhorar, por consequência, a eficiência dos treinos em conjuntos menores de amostras.

2.1.4 Transferência de Aprendizado

A técnica de Transferência de Aprendizado, em inglês *Transfer Learning*, se pauta no princípio de "reutilizar" o aprendizado de um modelo para auxiliar o aprendizado de um segundo modelo. As vantagens conferidas pela abordagem se destacam na facilitação de treinamento de novos modelos, por meio de um novo conjunto de dados para ajudar no aprendizado, além de reforçar conexões relevantes para o início de processamento, o que possibilita, por exemplo, mitigar o problema de *overfitting*. Com isso, torna-se possível treinar, através do novo conjunto de dados, treinar uma rede com

um número menor de dados rotulados, o que acelera substancialmente todo o processo de treinamento.

Um modelo amplamente usado para *transfer learning* é o ImageNet, um banco de dados hierárquico de nós com centenas de milhares de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). Não obstante, foi criado o ImageNet Large Scale Visual Recognition Challenge (ILSVRC), o qual busca avaliar algoritmos para detecção de objetos e classificação de imagens, possibilitando a comparação de progresso em reconhecimento em um âmbito vasto de variedades (IMAGENET, 2021).

2.2 Modelos de CNNs

Há diferentes modelos de arquitetura para Redes Neurais Convolucionais, cada qual com sua abordagem específica para realizar o processo de classificação e reconhecimento de padrões em imagens. Dentre os exemplos mais utilizados por sua relevância e eficiência para área, pode-se destacar os modelos *VGGNet* (Seção 2.2.1), *ResNet* (Seção 2.2.2), *EfficientNet* (Seção 2.2.3) e *InceptionNet* (Seção 2.2.4).

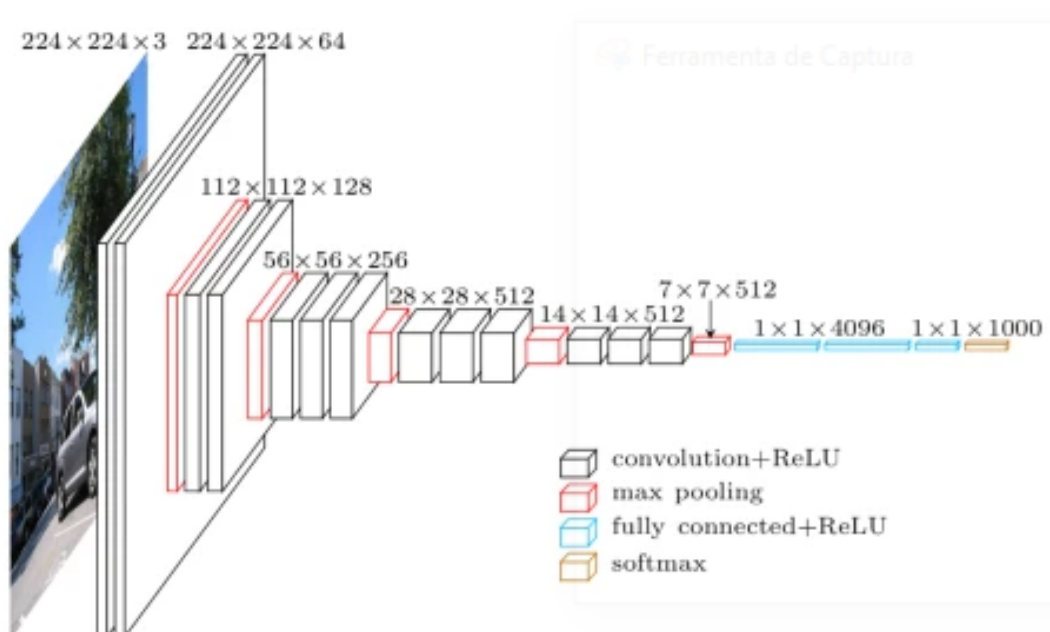
2.2.1 VGGNet

VGGNet, ou *Visual Geometry Group Network*, proposta por (SIMONYAN; ZISSEMAN, 2014), desenvolvida como uma rede neural profunda com aproximadamente 138 milhões de parâmetros, se destacou como um das arquiteturas mais populares para o reconhecimento de objetos e classificação de imagens devido suas altas performances durante o ILSVRC 2014 (RUSSAKOVSKY et al., 2015). A arquitetura de uma *VGGNet* segue o padrão de uma arquitetura CNN com múltiplas camadas, as quais podem variar em quantidade conforme seus submodelos, como, por exemplo, *VGG16* que possui 16 camadas, e *VGG19* com 19 camadas. As camadas de cada modelo são convolucionais, com exceção das 3 últimas camadas sendo totalmente co-

nectadas.

A *VGG19* é composta por 16 camadas convolucionais e, compondo o último bloco da rede, 3 camadas totalmente conectadas equipadas com a função ReLU e, ao final, uma camada *softmax*. Na estrutura também há 5 camadas de *pooling* entre os blocos que compõem as camadas convolucionais. Para exemplificar, a arquitetura de uma *VGG19* está ilustrada na Figura 2.

Figura 2: Representação da arquitetura do modelo de rede VGG19.



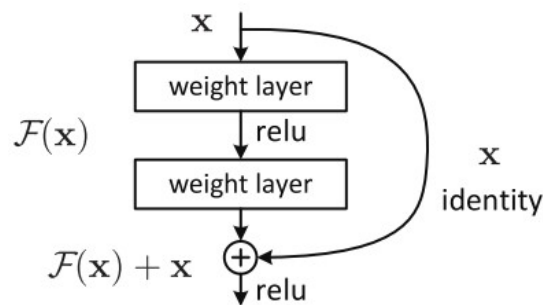
Fonte: (SIMONYAN; ZISSERMAN, 2014)

2.2.2 ResNet

Residual Neural Network, mais conhecida como *ResNet*, alcançou um desempenho considerado estado-da-arte nas classificações do ILSVRC 2015 (RUSSAKOVSKY et al., 2015; HE et al., 2016). Além disso, a *ResNet* se dispõe de vários sub-modelos que recebem seus nomes a partir do número de camadas que a constitui, como, por exemplo, *ResNet-18*, *ResNet-34*, *ResNet-101*, entre outros. O modelo de 50 camadas é ilustrado pelo diagrama esquemático presente na Figura 4 (HE et al., 2016).

A *ResNet* é composta por blocos residuais, os quais definem a abordagem específica dessa arquitetura. A partir desses blocos, tornou-se possível minimizar o problema de Gradiente de Fuga, a medida que redes com centenas ou até milhares de camadas são treinadas, o que discorre pelo motivo de haver uma retro propagação de gradientes diretamente para as camadas anteriores. O processo que faz uso dos blocos residuais é ilustrado na Figura 3. Em linhas gerais, são aplicadas as chamadas *skip connections* que somam uma saída x de uma camada anterior para a saída $f(x)$ da próxima camada, exatamente antes da função *ReLU*. Desse modo, a saída do bloco residual resultante é $\text{ReLU}(f(x) + x)$. Além do exposto, é importante destacar que as *skip connections* garantem o desempenho entre uma camada anterior para a próxima, a medida que permite que o modelo consiga aprender uma função de identidade.

Figura 3: Representação de um bloco residual na arquitetura *ResNet-50*.

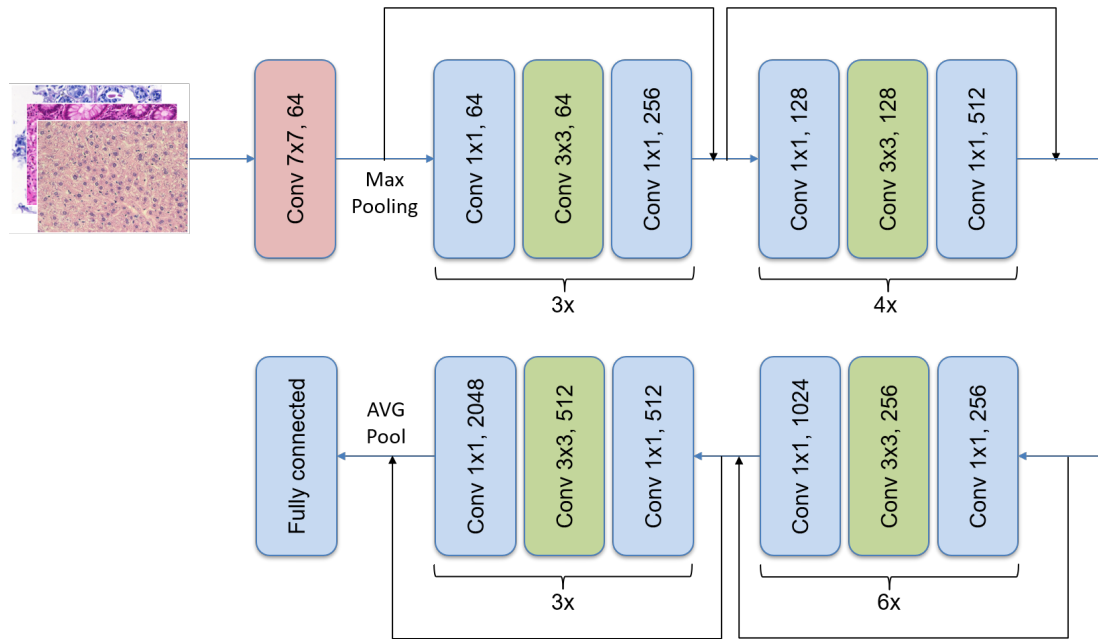


Fonte: (HE et al., 2016)

2.2.3 EfficentNet

A EfficentNet foi projetada com o intuito de utilizar uma abordagem mais sistemática para fins de otimização de uma CNN, incorporando as quatro principais formas existentes para melhorar os resultados da rede, segundo (TAN; LE, 2020):

- Ampliar a largura dos mapas de características;
- Ampliar o número de camadas (profundidade) da rede;
- Ampliar a resolução das imagens de entrada;

Figura 4: Representação da arquitetura do modelo de rede *ResNet-50*.

Fonte: Elaborado pelo autor

- Combinar as três abordagens.

A utilização arbitrária de tais abordagens torna o processo de otimização custoso em termos de complexidade de implementação, além de fornecer resultados não satisfatórios em termos de acurácia e eficiência (TAN; LE, 2020). Sendo assim, a *EfficientNet* faz uso da quarta abordagem ao partir do princípio de definição de uma forma sistemática para o modo que essas estratégias são utilizadas ao priorizar o objetivo de garantir a obtenção de bons resultados de classificação. Portanto, foi proposto o método *compound scaling method* que busca analisar fatores de escala constantes (representados por α , β e γ) ao determinar de forma equilibrada as abordagens de otimização de rede já descritas (TAN; LE, 2020). A Equação 2.4 denota como esses fatores são levados em conta.

$$\begin{aligned}
d &= \alpha^\phi, w = \beta^\phi, r = \gamma^\phi, \\
\alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\
\alpha \geq 1, \beta \geq 1, \gamma \geq 1
\end{aligned} \tag{2.4}$$

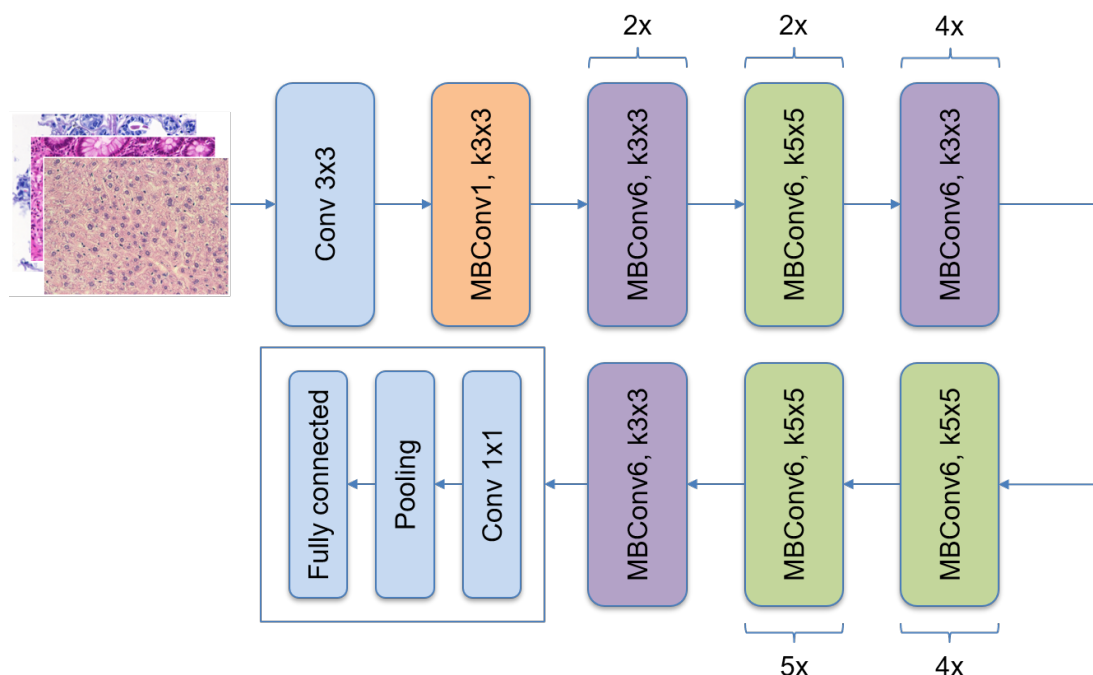
Os resultados para d referem-se a escala de profundidade (*depth*), enquanto w representa a de comprimento (*width*), e escala resolução (*resolution*) em r . A letra ϕ denota os recursos computacionais disponíveis para as escalas do modelo. Os valores α , β e γ , especificam como esses recursos são utilizados para as escalas de profundidade, comprimento e resolução, respectivamente (TAN; LE, 2020). Através do *grid search*, esses valores são úteis na definição da melhor combinação entre os fatores de escala. Com essa metodologia, o modelo *EfficientNet* tem tido desempenho muito satisfatório de forma geral.

Para efeito de exemplo, o diagrama da Figura 5 representa o diagrama do sub-modelo *EfficientNet-b2* (TAN; LE, 2020).

2.2.4 InceptionNet

A *InceptionNet* (SZEGEDY et al., 2015) utiliza da proposta *Network-in-Network* (LIN; CHEN; YAN, 2014), chamada de *Inception model* no caso, que busca aumentar o poder representativo das redes neurais, através da introdução de micro redes neurais, construídas com estruturas mais complexas que a convencional, em meio a cada camada convolucional. Outras versões dessa rede foram desenvolvidas com o intuito de aprimorar as anteriores, tais como exemplo a *GoogLeNet* (*InceptionV1*), *InceptionV2* e *InceptionV3*.

A rede *InceptionV3* foi apresentada em (SZEGEDY et al., 2016), junto da *InceptionV2*. Algumas mudanças foram implementadas para aumentar a acurácia e reduzir a complexidade computacional, explorando recursos como: *Label Smoothing*, que

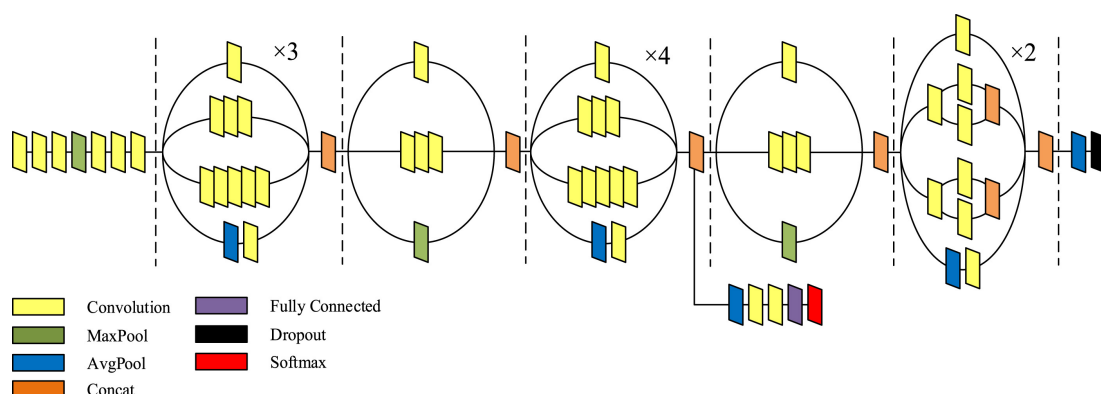
Figura 5: Representação da arquitetura do modelo de rede *EficientNet-b2*.

Fonte: Elaborado pelo autor

consiste em uma técnica de regularização que introduz ruído às classes; convoluções 7×7 ; e, por fim, a utilização de classificadores auxiliares para propagar informações da classe para camadas mais afundo na rede, com o uso de *batch normalization*. Sua arquitetura é composta por camadas de convolução, *average pooling*, *max pooling*, concatenações de micro redes, *dropouts* e, por fim, camadas totalmente conectadas, com *softmax* ao final. Um diagrama esquematizado da arquitetura da rede *InceptionV3* pode ser visto na Figura 6.

2.3 Técnicas de Inteligência Artificial Explicável e DeepDream

Conhecida pela sigla XAI (*Explainable Artificial Intelligence*), o conjunto de técnicas e processos de Inteligência Artificial Explicável têm importância reconhecida no contexto de interpretação humana dos resultados gerados via aprendizado de máquina.

Figura 6: Representação da arquitetura do modelo de rede *InceptionV3*.

Fonte: (DONG et al., 2020)

Surgiu como alternativa para a resolução de uma condição presente em aplicações *Deep learning*, conhecido como "problema da caixa-preta", visto que modelos de conhecimento profundo possuem muitas camadas ocultas, o que dificulta a compreensão de como eles alcançam suas conclusões (REYES et al., 2020).

Em suma, com relação às contribuições de tais técnicas, os argumentos favoráveis que mais se destacam assumem que é uma ferramenta poderosa para a inteligibilidade do sistema de predições com aprendizado profundo. Isso proporciona a compreensão dos resultados, por meio da interpretabilidade, com informações envolvendo os acertos e erros, bem como sobre a detecção de possíveis dados enviesados (SAMEK; WIEGAND; MÜLLER, 2017).

Dentre as técnicas XAI mais conhecidas e relevantes recentemente, destacam-se o Gradient_weighted Class Activation Mapping (Grad-CAM), abordado na seção 2.3.1, e também a técnica conhecida como Local Interpretable Model-Agnostic Explanations (LIME), abordada na seção 2.3.2.

2.3.1 Grad-CAM

O modelo Grad-CAM é uma generalização da abordagem CAM (*Class Activation Mapping*), a qual permite conhecer as regiões mais relevantes da imagem que fundamentam o processo de predição da rede convolucional (ZHOU et al., 2016). Isso

ocorre por meio de um mapa de ativação neural que denota as classes de ativação de uma imagem, representadas em cores que servem de índice para o grau de ativação, de forma análoga a um mapa de calor (cores do espectro azul ao vermelho). Diferentemente de CAM, o Grad-CAM permite realizar a geração de mapas de ativação das informações presentes em distintas camadas. Além disso, a função ReLU é utilizada na abordagem para que pesos negativos das camadas não interfiram na classificação.

2.3.2 LIME

A técnica LIME (*Local Interpretable Model-agnostic Explanations*) tem como função aproximar modelos "caixa preta" de aprendizado de máquina, definidos por serem complexos e não-lineares, para modelos interpretáveis simples e lineares. Essa estratégia embasa explicações sobre as predições. Para isso, cada explicação é obtida dividindo a imagem em *superpixels* contíguos que constituem componentes interpretáveis utilizados na explicação da classificação da amostra analisada. Cada explicação é alcançada considerando as predições de um modelo CNN ao analisar um determinado número de perturbações na imagem original. Tais perturbações geram máscaras representativas que permitem a obtenção de predições da rede. Um modelo simples é gerado e ajustado com base nas versões perturbadas, ponderadas por suas similaridades com as amostras e previsões correspondentes. Por fim, a explicação na forma de *superpixels* é gerada com base nos maiores valores positivos de pesos para classificar a imagem (RIBEIRO; SINGH; GUESTRIN, 2016).

2.4 DeepDream

O DeepDream é uma técnica de simulação baseada na dimensão imaginária do cérebro humano (MORDVINTSEV; OLAH; TYKA, 2015). A técnica busca amplificar padrões e características na imagem ao considerar a pareidolia algorítmica das infor-

mações presentes nas camadas de uma CNN (SUZUKI et al., 2017). Os dados de saída incluem os padrões observados na imagem, processados com os outros dados presentes na etapa de treinamento. Ademais, é possível aplicar o algoritmo de forma iterativa em seus próprios dados de saída para obter múltiplas impressões de representação.

Ainda segundo (MORDVINTSEV; OLAH; TYKA, 2015), a técnica permite entender visualmente a capacidade, em termos de desempenho, das redes neurais em tarefas de classificação. Além disso, aprimorar arquiteturas de rede, bem como observar o que a rede aprendeu durante o treinamento, são mais exemplos da capacidade do método DeepDream.

2.5 Extração de Características

Em processos de análise de dados, especialmente para o processamento de imagens, a extração de características é de suma importância, pois por meio de um conjunto de características há um embasamento resultante que impacta a busca por valores que representam informações visuais em uma imagem. Este fato contribui de maneira substancial no processo de classificação e uma imagem sob investigação (PEDRINI; SCHWARTZ, 2007). As características podem representar cores, níveis de borda, texturas e outros aspectos presentes em uma imagem. Os atributos podem ser armazenados e organizados na estrutura de um vetor de características, o que garante maior facilidade para análise em questões de acesso, classificação e visualização (GUYON et al., 2006). Para o caso de modelos CNN, as características obtidas são denominadas *Deep Features*.

Para fins de compreensão dos dados mais relevantes na imagem para os processos de classificação e reconhecimento de padrões, várias técnicas podem ser aplicadas para a análise das *features* a partir dos vetores de características.

2.6 Seleção das Melhores Características Via Algoritmo

ReliefF

A seleção de características é um processo que visa a obtenção de subconjuntos de *features* consideradas mais relevantes para a classificação. Para isso, há algoritmos baseados em filtragem, os quais realizam o ranqueamento de atributos a partir de propriedades intrínsecas, sem dependência da classificação (DANASINGH; BALAMURUGAN; EPIPHANY, 2016).

O ReliefF é um dos algoritmos baseados em filtragem, sendo uma expansão do algoritmo Relief (KIRA; RENDELL, 1992). Sua função é ranquear as características mais importantes para a classificação, por meio de um critério pré-definido ao cruzar instâncias com as classes de classificação. Além disso, também analisa uma comparação média de características vizinhas e suas contribuições, o que garante maior confiabilidade e robustez ao método (KONONENKO; ŠIMEC; ROBNIK-ŠIKONJA, 1997). Utilizando o critério da distância de Manhattan, a distância absoluta entre instâncias e as classes das imagens embasam o ranqueamento.

A saída do algoritmo ReliefF é um vetor de características organizado conforme a ordem de relevância com base nos pesos calculados de cada característica. Os pesos W das características A são obtidos através do cálculo de aproximação das diferenças de probabilidade P_r , considerando a instância mais próxima da classe diferente (I_d) e a instância mais próxima à mesma classe I_s . A Equação 2.5 ilustra o cálculo explicado.

$$W[A] = P(I_d) - P(I_s) \quad (2.5)$$

2.7 Classificação de Características

A classificação de características visa a obtenção de métricas para a avaliação da relevância de um conjunto de características, por exemplo, o conjunto filtrado a partir do ranqueamento via ReliefF. Com isso, a classificação pode ser otimizada, tendo em vista que será obtido um subconjunto de *features* capaz de indicar os melhores desempenhos de classificação.

Os algoritmos de classificação se dividem em múltiplas categorias, indicando a estratégia que embasa o processo de classificação, por exemplo, algoritmos baseados em função, árvores de decisão, *meta learning* e *lazy learning*. As subseções a seguir exploram, respectivamente, cada categoria e os classificadores de cada uma usados neste trabalho.

2.7.1 Baseados em Função

Algoritmos baseados em função identificam a classe a qual um objeto pertence a partir da combinação linear de suas características. Alguns exemplos são *Multilayer Perceptron*, *Logistic* e *Support Vector Machine*.

- ***Multilayer Perceptron* (MP)**

O MP é um algoritmo de rede neural supervisionada baseado no modelo *Perceptron* que busca aprender a relação entre dados lineares e não lineares. De modo geral, a rede neural possui camadas de entrada e de saída, além de uma ou mais camadas ocultas com muitos neurônios. As entradas têm seus pesos iniciais combinados por meio de uma soma e submetidos a uma função de ativação arbitrariamente escolhida, para então propagar as combinações lineares para a próxima camada até as camadas de saída, enquanto realiza o processo de retro propagação entre elas, o que garante que a rede "aprenda" com os dados.

- ***Logistic (Log)***

Logistic (CESSIE; HOUWELINGEN, 1992) é um algoritmo de classificação para construir e utilizar modelos de regressão logística com um estimador de cadeia. O objetivo se resume em utilizar funções lineares para cálculo de probabilidade de classes, as quais servem de entrada para outra função. Permite estimar uma relação entre múltiplas variáveis de previsão e variáveis (classes) alvo.

- ***Support Vector Machine (SVM)***

Proposto por (VAPNIK, 1963), o algoritmo SVM faz uso de uma função *kernel* a fim de mapear os dados de entrada, por meio de pontos em um determinado espaço, para um espaço de características de alta dimensão, denominado hiperplano, o qual define categorias que contêm espacialmente os dados de validação, sendo esse o parâmetro pelo qual esses dados são classificados.

2.7.2 Baseados em Árvores de Decisão

Classificadores baseados em árvores de decisão utilizam uma estrutura de árvore, com nós representando características, a conexão entre os nós (galho) representa uma decisão, e as classes são as folhas. Para este trabalho, *Random Forest* foi o classificador baseado em decisão de árvore escolhido.

- ***Random Forest (RandF)***

O algoritmo (BREIMAN, 2001) constrói e combina várias árvores de decisão que realizam a classificação, visando obter resultados de treinamento melhores do que os de uma única árvore, além de mitigar o problema de *overfitting*. A princípio, são escolhidas aleatoriamente K características do conjunto de treinamento, utilizadas para encontrar o nó raiz, ou seja, a característica mais relevante

para a tomada de decisão. Consequentemente, outras características são relacionadas aos nós subsequentes, fazendo com que o resultado da classificação seja apresentado nas folhas das árvores. As diversas árvores de decisão diferentes são geradas por meio da repetição desse processo, oferecendo um número de predições importantes para a análise de precisão das classificações.

2.7.3 Baseados em *Meta-Learning*

Classificadores baseados em *meta-learning*, também chamados de *meta-classifiers* ou métodos *ensemble*, têm o intuito de aprender algoritmos que aprendem pelos dados de saída de outros algoritmos de aprendizado, ou seja, usa algoritmos que aprendem a melhor forma de combinar as predições de outros algoritmos, criando um ambiente de aprendizado combinado (*ensemble learning*). Um sistema *meta-learning* é composto por de um subsistema de aprendizado que se adapta com a acumulação de experiência adquirida através do conhecimento obtido em aprendizados prévios (KHAN et al., 2020). No caso, os classificadores atuam baseados nas saídas da CNN e aprendem a utilizar as melhores predições feitas para realizar as suas próprias.

- ***Rotation Forest (RotF)***

Rotation Forest é um do método de classificação *ensemble* que se assemelha ao classificador *Random Forest* a medida que procura mitigar sua maior limitação (RODRIGUEZ; KUNCHEVA; ALONSO, 2006). O algoritmo "rotaciona" aleatoriamente cada árvore de decisão, dividindo as características do modelo em um determinado número de partições para obter uma matriz de rotação. Com isso, multiplicando a matriz obtida com as partições de atributos originais, as *features* são rotacionadas e, assim, as árvores de decisão são obtidas.

- ***Decorate (Dec)***

O classificador *Decorate* (MELVILLE; MOONEY, 2003) constrói diversas com-

combinações de classificadores de modo iterativo, aprendendo ao utilizar treinamentos de instâncias adicionais gerados artificialmente. Ao produzir uma diversidade de combinações eficientes, *Decorate* demonstra ser altamente efetivo em dados de treinamento limitados.

2.7.4 Baseados em *Lazy Learning*

Classificadores guiados por *Lazy Learning* utilizam um "atraso" no processo de generalização do conjunto de dados de treinamento. Esse atraso ocorre, pois os classificadores desta categoria armazenam todo o conjunto de dados de treinamento para realizar o processo de classificação somente quando um conjunto de testes é fornecido (AHA, 1997). Os algoritmos K-estrela e *K-nearest Neighbour* são alguns exemplos de classificadores representativos dessa categoria.

- ***K-estrela***

Também conhecido como KStar ou K, é um classificador que considera a medida de entropia, assim como a distância entre as instâncias, para determinar a probabilidade de uma instância se tornar outra.

- ***K-nearest Neighbour (IBk)***

Conhecido também como IBk (*Instance Based k*) (AHA; KIBLER; ALBERT, 1991), recorre à estratégia *K-nearest neighbours*, onde *K* representa o número de classes que estão em análise. Em linhas gerais, é formado um grupo com as *K* características mais próximas, para que então seu desempenho seja analisado.

2.8 Métricas para Avaliação de Desempenho

Para fins de avaliação após a aplicação dos classificadores, adotam-se métricas que denotam, de forma aprofundada, as previsões realizadas. Exemplos de métricas que

podem ser analisadas e foram utilizadas para a avaliação dos resultados neste trabalho, são: área sob a Curvatura ROC (AUC), acurácia (ACC) e medida-F (F-score) (DUDA; HART; STORK, 2012), abordadas nas subseções 2.8.1, 2.8.3 e 2.8.2 respectivamente.

Para o cálculo das métricas utilizam-se, como parâmetros, as categorias encontradas em uma estrutura chamada matriz de confusão. Essas categorias são dadas a partir da classificação binária do modelo, conforme as classes em análise. As categorias, de forma generalizada, são:

- **Verdadeiros Positivos (VP):** valores positivos classificados de forma correta pelo modelo como positivos;
- **Verdadeiros Negativos (VN):** valores negativos classificados de forma correta pelo modelo como negativos;
- **Falsos Positivos (FP):** valores negativos classificados de forma incorreta pelo modelo como positivos;
- **Falsos Negativos (FN):** valores positivos classificados de forma incorreta pelo modelo como negativos.

A matriz de confusão pode ser esquematizada conforme a Tabela 1.

Tabela 1: Representação esquemática de uma matriz de confusão

	Classificado como Positivo	Classificado como Negativo
Positivo	VP	VN
Negativo	FP	FN

2.8.1 Acurácia (ACC)

A métrica ACC é muito comum no âmbito da classificação de imagens. Seu objetivo é estipular a efetividade e exatidão dos resultados extraídos do classificador, via valores presentes na matriz de confusão. O cálculo é dado por:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.6)$$

O resultado obtido é analisado em forma de porcentagem entre 0% e 100%.

2.8.2 Medida-F

Também conhecida com *F-score*, essa métrica visa indicar qualidade geral da classificação, com base em duas outras métricas denominadas de Precisão(*P*) e *Recall* (*R*). *P* indica a razão entre as amostras classificadas como uma determinada classe, com o total de amostras identificadas como essa mesma classe correta ou incorretamente ($VP + FP$) (Equação 2.7).

$$P = \frac{VP}{VP + FP} \quad (2.7)$$

De forma similar, *R* é calculado a fim de obter a proporção das classes identificadas corretamente, Equação 2.8. O resultado pode ser interpretado como a sensibilidade de classificação.

$$R = \frac{VP}{VP + FN} \quad (2.8)$$

A partir do resultado dessas duas medidas, é possível obter sua média harmônica, também conhecida como *F-score*, dada pela Equação 2.9.

$$Fscore = \frac{2 \cdot P \cdot R}{P + R} \quad (2.9)$$

Os resultados se dão por valores entre 0 e 1, sendo que, as taxas quanto mais próximas de 1, maior a qualidade de desempenho classificativo.

2.8.3 Área sob a Curvatura ROC (AUC)

A métrica AUC está relacionada ao poder discriminativo do classificador. A curva ROC (*Receiver Operating Characteristic*) tem como função facilitar a visualização e organização na análise de desempenho de uma classificação. O cálculo é dado através dos valores de R (Equação 2.8) e Especificidade (E , Equação 2.10), que, por sua vez, denota a taxa de amostras negativas classificadas corretamente como negativas.

$$E = \frac{VN}{VN + FP} \quad (2.10)$$

Com isso, é possível obter a efetividade do modelo ao evitar classificações falsas, demonstrando seu poder discriminativo através da métrica AUC (Equação 2.11).

$$AUC = \frac{R + E}{2} \quad (2.11)$$

Os valores obtidos se encontram no intervalo entre 0 e 1, sendo que, taxas próximas a 1 constituem alto poder discriminativo, enquanto taxas próximas e menores que 0,5 indicam alto grau de aleatoriedade (FAWCETT, 2006).

Revisão Bibliográfica

Neste capítulo, são apresentados trabalhos relacionados ao contexto de aplicação de CNNs para a classificação e reconhecimento de padrões em imagens histológicas, especialmente de câncer de mama, câncer colorretal, amostras de tecido hepático e displasia oral. Nesse âmbito, há também a exposição de trabalhos relacionados à utilização de XAI, bem como DeepDream, em prol do reconhecimento das decisões tomadas através das redes neurais.

Em Dabass et al. (DABASS; VIG; VASHISTH, 2018), os autores propuseram uma técnica automática para classificar câncer colorretal via imagens H&E, usando um modelo CNN com 31 camadas. A abordagem proposta utilizou um conjunto de dados com 165 imagens (775 x 522), que foi dividido em treinamento (80%) e teste (20%). As amostras foram organizadas em cinco classes, em adição a um grupo binário (benigno *versus* maligno). Na classificação binária, o método forneceu um valor de acurácia de 96,97%. A classificação envolvendo as cinco classes indicou um valor de acurácia de 93,94%.

Em (SHALLU; MEHRA, 2018), Shallu e Mehra realizaram experimentos para classificar imagens histológicas de câncer de mama com modelos CNN, explorando *transfer learning versus* o treinamento completo das arquiteturas. Os autores utilizaram os modelos VGG16, VGG19 e ResNet50, com classificador baseado em regressão logística. A proposta explorou 7909 imagens, divididas nos grupos benigno e maligno, com 2480 e 5429 amostras, respectivamente. A estratégia de treinamento completo

utilizou *data augmentation*, com rotações de imagem em três ângulos específicos. Os autores concluíram que a abordagem *transfer learning* foi capaz de fornecer os melhores desempenhos, indicando uma acurácia máxima de 92,20%.

A utilização de abordagens que exploram *deep features* também foram observadas nos modelos de (MAHENDRAN; VEDALDI, 2016; VEDALDI; ZISSERMAN, 2013; YOSINSKI et al., 2015). Neste contexto, os métodos de (NANNI et al., 2020; NANNI; GHIDONI; BRAHNAM, 2018; NANNI et al., 2019) foram desenvolvidos para classificar amostras H&E. As estratégias exploraram *deep features* com *handcrafted features*. A proposta de (NANNI et al., 2019) considerou um conjunto de 6 arquiteturas CNN (AlexNet, GoogleNet, VGG16, VGG19, ResNet50 e DenseNet), pré-treinadas em ImageNet. As *handcrafted features* foram usadas na etapa de treinamento do classificador SVM para constituir um conjunto de atributos. Além disso, a proposta considerou *data augmentation*. Os resultados mostraram valores significativos na discriminação de tecidos de fígado e câncer colorretal.

Mesmo com os avanços vistos na Literatura especializada, ainda é possível investigar *deep features* baseadas em representações XAI (ADADI; BERRADA, 2018; MAHENDRAN; VEDALDI, 2016; VEDALDI; ZISSERMAN, 2013; YOSINSKI et al., 2015). Abordagens baseadas em Grad-CAM(RAJARAMAN et al., 2018a; REYES et al., 2020), LIME(RAJARAMAN et al., 2018a; REYES et al., 2020; SOUSA; VELLASCO; SILVA, 2019) e DD(TOĞAÇAR; CÖMERT; ERGEN, 2021; MORDVINTSEV; OLAH; TYKA, 2015; SUZUKI et al., 2017) também podem contribuir para essa tarefa com novas combinações.

Neste contexto, técnicas XAI (Grad-CAM e LIME) foram consideradas fundamentalmente para sustentar explicações de classificações (RAJARAMAN et al., 2018a; REYES et al., 2020; SOUSA; VELLASCO; SILVA, 2019). Por exemplo, o modelo de (XIE et al., 2021) utilizou a técnica Grad-CAM para destacar as características aprendidas através da arquitetura ResNet50, explorando 841 imagens histológicas WSI

(*whole-slide image*): classes melanoma (392) e nevo (449). O modelo alcançou uma taxa de acurácia de 93,30%. Não obstante, a proposta de (LI et al., 2023) considerou classificações de tipos de câncer cervical, via arquiteturas CNN e técnicas XAI (*Guided* Grad-CAM e Grad-CAM) para interpretabilidade diagnóstica. Os modelos AlexNet, VGG19, Xception e ResNet50 foram escolhidos para classificar câncer cervical maligno e não-maligno. O conjunto de dados foi definido com 8496 imagens histológicas. A rede Xception expressou o maior desempenho, com AUC de 0,974.

Em (HAM et al., 2022), os autores propuseram um sistema baseado na CNN Xception capaz de classificar imagens gastroscópicas. A técnica Grad-CAM foi aplicada como uma estratégia de *data augmentation*. O estudo usou um conjunto de dados com amostras representativas de diferentes classes, como câncer de intestino, gastrite, tumor submucoso, úlcera e pólipos. A técnica Grad-CAM indicou as áreas observadas via Xception, também explorando uma abordagem de aumento de dados. O sistema alcançou uma taxa AUC de 0,885, sem *data augmentation*, contra um valor de AUC de 0,903 com *data augmentation*.

Abir et al. (ABIR et al., 2022) aplicaram a técnica LIME para garantir a validade e confiabilidade de classificações explorando imagens de leucemia linfoblástica. O conjunto de imagens considerou 15114 imagens microscópicas de 118 paciente. A arquitetura Inception V3 foi escolhida para a tarefa de classificação, atingindo 98,38% de acurácia. A proposta de (FERRER-SÁNCHEZ et al., 2022) também explorou LIME para explicar classificações de lesões de leucoplasia oral via estratégias CNN. O estudo considerou um *dataset* com 261 amostras, as quais foram segmentadas por meio da arquitetura U-Net. Os resultados foram relevantes na predição de imagens representativas de risco de displasia.

Por outro lado, a técnica DeepDream (TOĞAÇAR; CÖMERT; ERGEN, 2021) ainda não foi completamente explorada no contexto de imagens H&E. O modelo de (TOĞAÇAR; CÖMERT; ERGEN, 2021) foi desenvolvido para distinguir estágios dis-

tintos da doença de Alzheimer. A proposta combinou DeepDream, hipercoluna e aprimoramento de imagem em cores difusas para melhorar a qualidade do *dataset*: 6400 imagens divididas em treinamento (5121) e teste (1279). Os autores utilizaram a arquitetura VGG16, explorando as *deep features* mais relevantes com regressão linear. A classificação foi realizada utilizando o algoritmo SVM, proporcionando um valor geral de 99,94%.

Em (JIANG et al., 2021), os autores investigaram a classificação de imagens histológicas com modelos CNN (Inception V3 e AlexNet), utilizando Grad-CAM e DeepDream para explicar as classificações. O método explorou imagens representativas de nódulos, organizadas como lesões benignas e pré-invasivas, adenocarcinoma minimamente invasivo, e adenocarcinoma invasivo. O modelo foi capaz de expressar um valor de acurácia de 93%.

A partir dos métodos apresentados anteriormente, é possível observar múltiplas estratégias voltadas ao estudo de imagens histológicas. As aplicações foram conduzidas com combinações de arquiteturas CNN, *transfer learning*, conjunto de características, múltiplos classificadores e técnicas XAI. Neste contexto, aplicações das técnicas Grad-CAM e LIME foram amplamente utilizadas para produzir soluções compreensíveis (LI et al., 2023; ABIR et al., 2022; FERRER-SÁNCHEZ et al., 2022; TOĞAÇAR; CÖMERT; ERGEN, 2021; XIE et al., 2021). As abordagens proporcionaram resultados relevantes quando aplicadas como aprimoramento de dados (HAM et al., 2022; TOĞAÇAR; CÖMERT; ERGEN, 2021). Entretanto, os modelos não foram investigados como no estudo proposto neste trabalho, nem mesmo na utilização combinada de modelos CNN ou em *deep features* extraídas de representações XAI e DeepDream, em múltiplos conjuntos de imagem H&E (câncer de mama, câncer colorretal, tecido hepático e displasia oral).

Metodologia

Neste capítulo são apresentadas informações sobre a metodologia e estratégias seguidas que fundamentaram este projeto de forma detalhada, dentre eles os modelos de CNN utilizados, a extração de representações Grad-CAM, LIME e DD, composição dos vetores de *deep features*, processo de ranqueamento e seleção das características mais relevantes, aplicação de múltiplos classificadores e métricas utilizadas na análise a avaliação dos resultados obtidos.

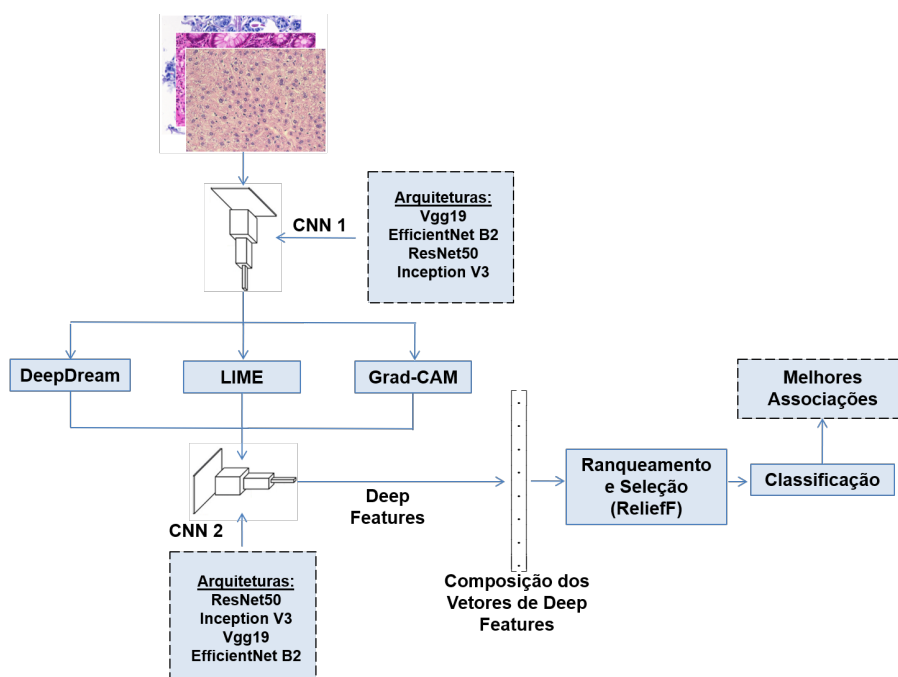
4.1 Estruturação do Esquema

O esquema proposto foi dividido em etapas. A primeira etapa consistiu na aplicação de uma CNN, com a abordagem *transfer learning* via ImageNet, utilizando cada conjunto de imagens H&E para obter um modelo de aprendizado profundo através de *fine-tuning*. Consequentemente, na segunda etapa, as técnicas DeepDream, LIME e Grad-CAM foram aplicadas para extrair representações das imagens H&E, utilizando cada modelo treinado na etapa anterior. A terceira etapa considerou diferentes combinações de CNN, também utilizando *transfer learning*, para extrair *deep features* de cada representação obtida na segunda etapa. Na quarta etapa, o algoritmo ReliefF (URBANOWICZ et al., 2018) foi aplicado para classificar as características mais relevantes, utilizando diferentes composições. Por fim, a quinta etapa considerou múltiplos classificadores para identificar os atributos mais relevantes e indicar as principais

combinações.

O modelo proposto foi desenvolvido e testado na plataforma Google Colab, utilizando a linguagem de programação Python. Os *scripts* foram executados em um computador com um processador Intel Core i3-6006U 2.0GHz, 4 GB de RAM e recursos de computação em nuvem. As técnicas DD, LIME e Grad-CAM foram implementadas utilizando o *framework* PyTorch, considerando as bibliotecas *Archvision* e *Torch*. Um diagrama esquemático do modelo proposto é mostrado na Figura 7, com CNN 1 representando a abordagem utilizada para extrair representações XAI e CNN 2 (diferente de CNN 1) para extrações de *deep features*.

Figura 7: Visão geral da estratégia proposta para investigar imagens H&E.



Fonte: Elaborado pelo autor.

4.2 Contexto de aplicação: imagens histológicas

A abordagem proposta neste projeto foi testada em imagens com importância científica reconhecida, tais como imagens H&E representativas de tumores colorretais,

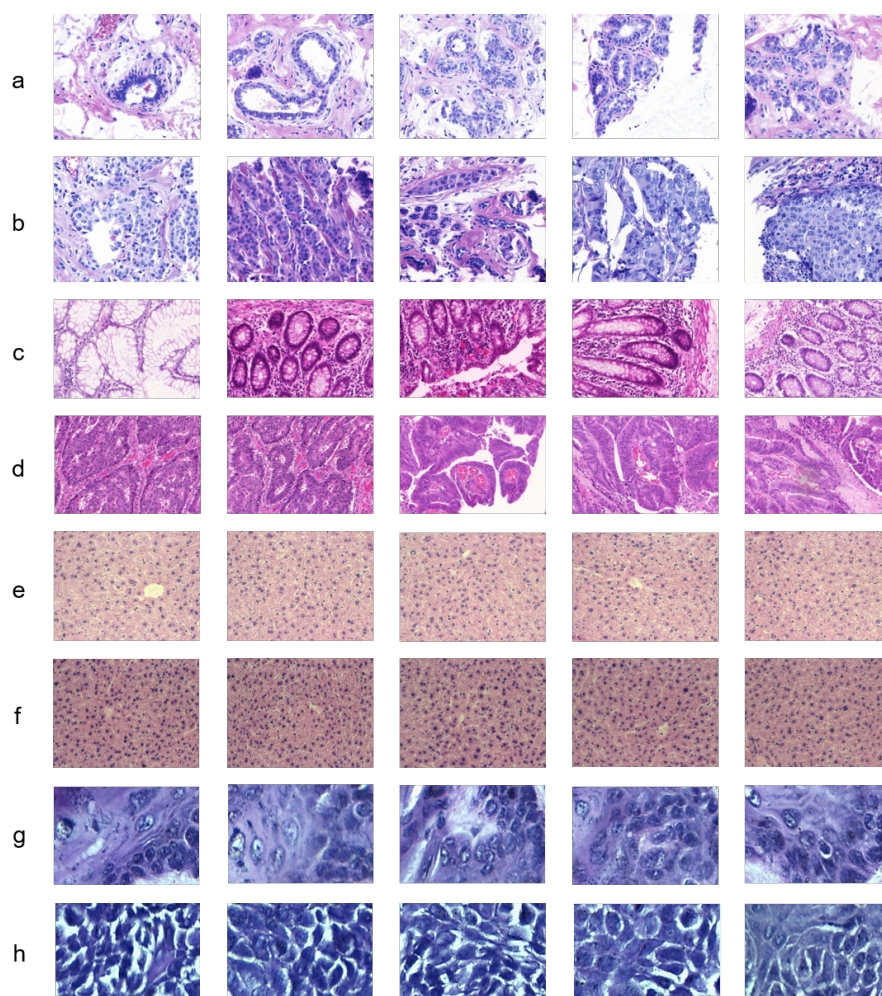
câncer de mama, tecido hepático e displasia oral.

- **Câncer de mama (UCSB).** Este banco de imagens é composto por 58 imagens histológicas obtidos a partir de biópsias coradas com Hematoxilina e Eosina (H&E). Todas as imagens foram disponibilizadas pela University of California Santa Barbara (GELASCA et al., 2008). As imagens possuem dimensões de 768x896 *pixels*, padrão de cor RGB e taxa de quantização de 24 bits.
- **Câncer Colorretal (CR).** O banco é constituído por imagens histológicas derivadas de 16 seções coradas em H&E do estágio T3 ou T4 de câncer colorretal (SIRINUKUNWATTANA et al., 2017). A digitalização das seções histológicas em *whole-slide images* (WSI) foi realizada com um digitalizador Zeiss MIRAX MIDI com resolução de *pixel* 0.465 μ m. As imagens foram categorizadas nos grupos benigno ou maligno. Neste trabalho foram utilizadas 151 imagens com dimensões de 775 x 522 *pixels*, divididas em 67 casos benignos e 84 casos malignos.
- **Tecido hepático (LG).** Esta base é nomeada como *liver gender* (LG) e disponibilizada pelo Atlas of Gene Expression in Mouse Ageing Project (AGEMAP) (AGEMAP, 2020). A base é constituída por imagens com dimensões de 417 x 312 *pixels* representativas de tecido hepático de camundongos separados como machos e fêmeas. Assim, essas duas classes representam o gênero da amostra coletada, totalizando 265 exemplos: macho com 150 imagens e fêmea com 115 amostras.
- **Displasia Oral (DISP).** O conjunto de dados de imagem foi construído a partir de 30 cortes de tecido de língua de camundongos corados com H&E previamente submetidos a um carcinógeno durante dois experimentos realizados em 2009 e 2010, devidamente aprovados pelo Comitê de Ética no Uso de Animais, sob protocolo número 038/39 no Federal Universidade de Uberlândia. Um total de 66

imagens histológicas foram obtidas usando o microscópio óptico LeicaDM500 com aumento de 400, usando o modelo de cores RGB com resolução de 2048×1536 pixels. O conjunto de dados de imagem considerado neste trabalho é composto por amostras saudáveis e displasia grave com 74 e 222, respectivamente (SILVA et al., 2022). A resolução de cada imagem é de 452×250 pixels.

Na Figura 8 encontram-se exemplos das imagens de cada base e suas respectivas classes.

Figura 8: Amostras H&E das bases: UCSB benigno (a) e maligno (b); CR benigno (c) e maligno (d); LG macho (e) e fêmea (f); DISP saudável (g) e severo (h).



Fonte: Elaborado pelo autor.

4.3 Etapa 1 - Aplicação da Rede Neural Convolucional de Extração de Representações XAI

Essa etapa consistiu na aplicação de uma arquitetura CNN para obter as extrações com as técnicas Grad-CAM, LIME e DD, utilizando a abordagem *transfer learning* para obter os atributos mais relevantes da camada de ativação. As representações foram obtidas pela camada de *average pooling*, reconhecida por conter as principais *features* necessárias para o processo de classificação. A abordagem *transfer learning* foi definida com o uso do conjunto de dados *ImageNet*, o que possibilitou o reconhecimento de padrões e a classificação em pequenas quantidades de amostras (OLIVAS JOSE DAVID MARTIN GUERRERO, 2009). Além disso, vale notar que o processo de *fine-tuning* foi implementado para diferenciar cada conjunto de dados baseado em suas respectivas categorias: dados benignos e malignos para câncer de mama, tumor colorretal, e displasia oral; masculino e feminino para as amostras de tecido hepático. Com uma taxa de aprendizagem de 0,01, treinamento utilizando o método *k-fold cross-validation* ($k = 5$), e considerando 10 *epochs*, as condições exploradas aqui podem ser comumente observadas na literatura especializada. Após o processo de treinamento, os modelos CNN foram aplicados para discriminar cada tipo de imagem.

4.4 Etapa 2 - Extração das representações com as técnicas DD, LIME e Grad-CAM

Esta etapa visou aplicar as técnicas Grad-CAM, LIME e DD para gerar suas respectivas representações de cada uma das imagens histológicas H&E. Os detalhes estão nas próximas subseções.

4.4.1 Técnica Grad-CAM

A extração das representações via *class activation mapping* teve como função o reconhecimento das regiões da imagem que influenciaram as classificações de cada CNN (ZHOU et al., 2016). Neste trabalho, a técnica Grad-CAM foi escolhida devido a sua robustez em gerar mapas de diferentes camadas. Além disso, o Grad-CAM, diferentemente de CAM, faz uso da função ReLU para mitigar a influência de pesos negativos incluídos na camada sob investigação, principalmente porque tais pesos não fazem parte das regiões comumente usadas para determinar a classificação final. As condições aplicadas neste trabalho foram baseadas nas descrições de Rajamaran et al. (RAJARAMAN et al., 2018b) e Reyes et al. (REYES et al., 2020), nas quais o resultado do método Grad-CAM para uma classe c foi interpretado como uma soma ponderada dos mapas de *deep features*, e resumida nas Equações 4.1 e 4.2 (RAJARAMAN et al., 2018a; REYES et al., 2020).

$$GradM_c(x,y) = ReLU(\sum_k \alpha_k^c f_k(x,y)), \quad (4.1)$$

em que $f_k(x,y)$ denota a ativação de um elemento espacial (x,y) no k -ésimo mapa de característica; α_k^c foi o peso obtido pelo cálculo do gradiente de uma pontuação de predição; S_c é referente ao k -ésimo mapa de características.

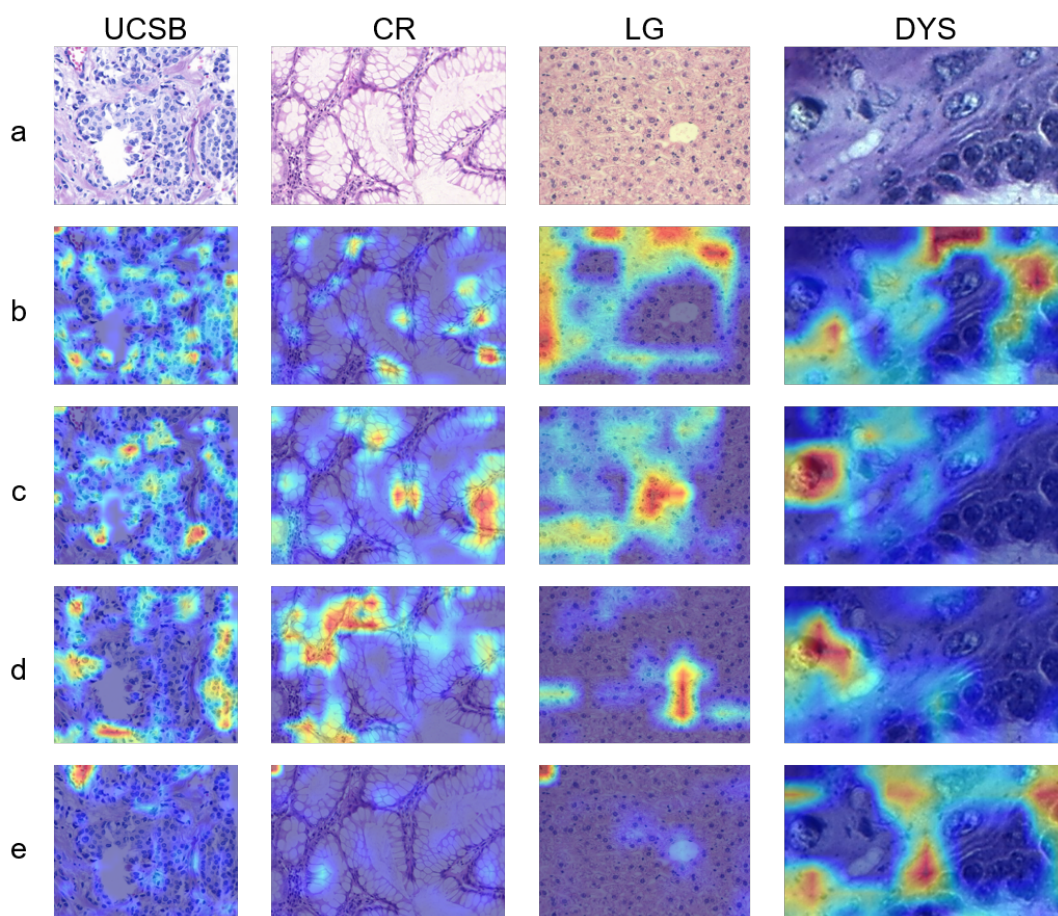
Como complemento, α_k^c foi obtido de:

$$\alpha_k^c = \sum_{x,y} \frac{\partial S_c}{\partial f_k(x,y)}. \quad (4.2)$$

Para a implementação desta técnica, foi utilizado o pacote Grad-Cam para PyTorch. O mapa de calor obtido (*colormap*) foi definido baseado no grau de ativação de

cada região analisada: menor ativação com tons de azul; maior ativação com tons de vermelho; e intermediárias com tonalidades entre azul e vermelho. As imagens foram obtidas da camada *average pool* de cada CNN. Alguns exemplos das representações Grad-CAM são mostrados na Figura 9.

Figura 9: Representações Grad-CAM das imagens H&E originais (a), explorando o modelo VGG19 (b), modelo ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).



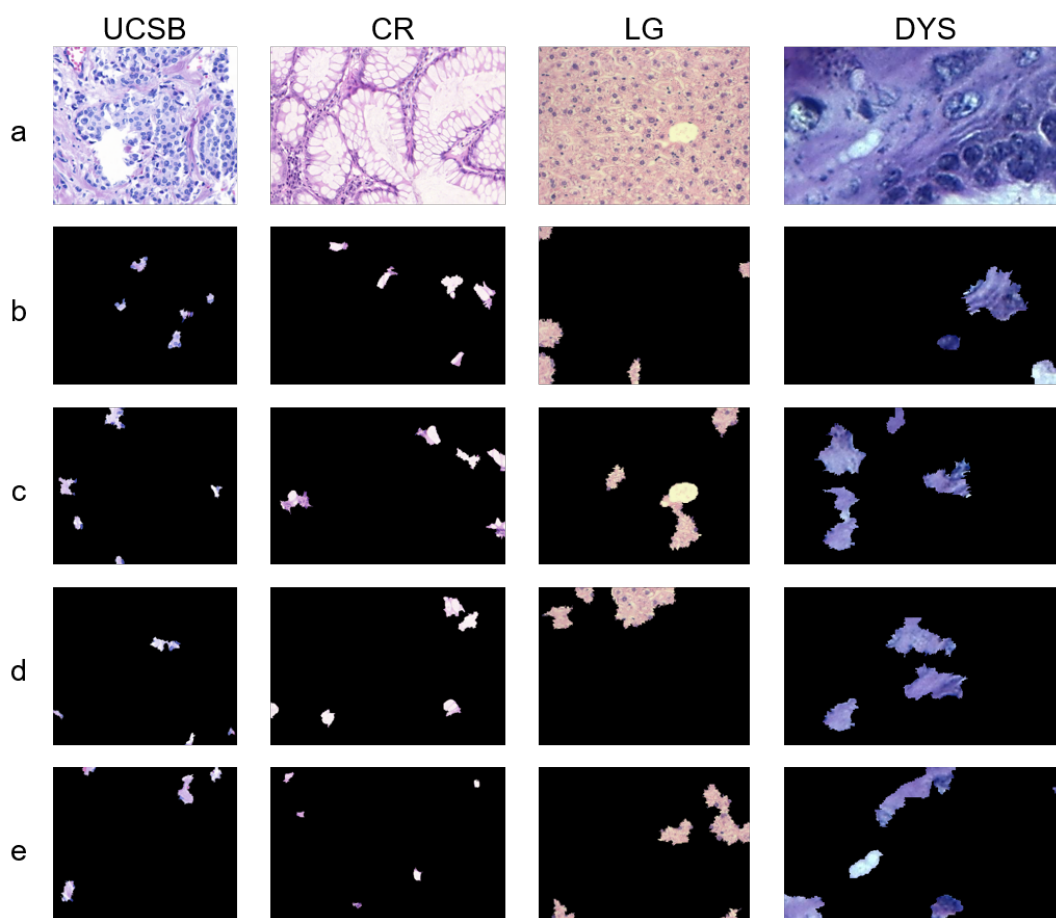
Fonte: Elaborado pelo autor.

4.4.2 Técnica LIME

A técnica LIME foi explorada como uma estratégia para produzir representações baseadas em interpretabilidade local. A técnica realizou automaticamente uma aproximação do modelo complexo local para um modelo simples, como um modelo linear, em torno da amostra de entrada a ser interpretada (RIBEIRO; SINGH; GUESTRIN,

2016). A técnica segmentou a imagem de entrada em regiões (*superpixels*) para selecionar aqueles com maior impacto na saída. O melhor resultado representou a explicação da classificação (RIBEIRO; SINGH; GUESTRIN, 2016). Cada explicação foi derivada da predição gerada pelo modelo CNN, levando em consideração um conjunto de perturbações aplicadas à entrada original. Neste trabalho, foram usadas 1000 perturbações, cada uma delas criadas por meio da remoção aleatória de segmentos da imagem, visando explorar as regiões mais relacionadas à saída original. As explicações consistiram em representações dos cinco *superpixels* que mais impactaram a classificação. Alguns exemplos das representações LIME obtidas estão ilustrados na Figura 10

Figura 10: Representações LIME das imagens H&E originais (a), considerando o modelo VGG19 (b), modelo ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).



Fonte: Elaborado pelo autor.

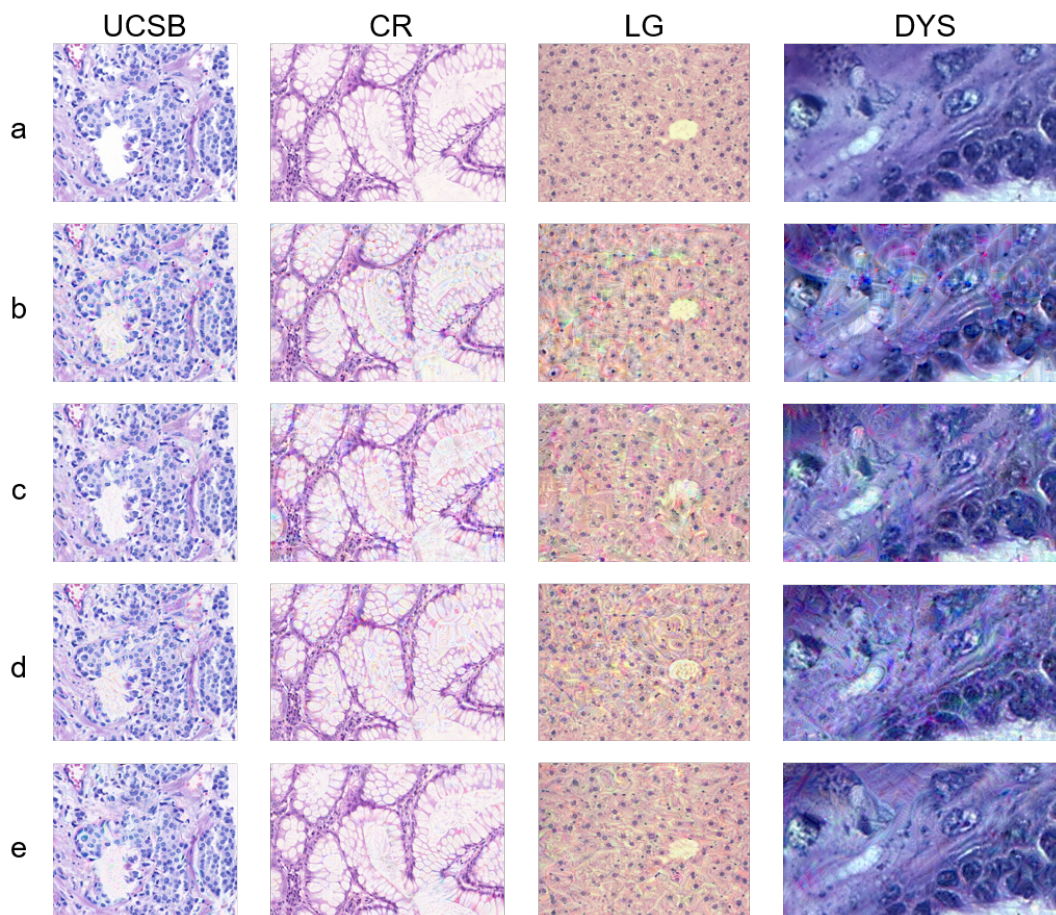
4.4.3 Técnica DeepDream

As representações DD foram baseadas no efeito conhecido como *inceptionism*, fundamentada nas dimensões imaginárias do cérebro humano (MORDVINTSEV; OLAH; TYKA, 2015). A técnica consistiu em um processo de simulação para indicar padrões e/ou características disponíveis nas imagens histológicas, utilizando pareidolia artificial através da CNN sob investigação (SUZUKI et al., 2017). O algoritmo DD alimenta repetidamente cada imagem para a CNN para encontrar os padrões mais significantes, esse processo ocorreu na etapa de treinamento da arquitetura, via algoritmo de *back-propagation* (retropropagação). Além disso, a estratégia mudou os pesos de neurônios específicos para identificar e exagerar padrões na imagem histológica original. Esse processo foi repetido para obter a imagem desejada por meio de um valor suficientemente pequeno de perda. Os padrões foram então exagerados, fornecendo uma nova imagem (TOĞAÇAR; CÖMERT; ERGEN, 2021; MORDVINTSEV; OLAH; TYKA, 2015; SUZUKI et al., 2017). Nesse trabalho, as novas imagens foram obtidas com 40 iterações, 5 oitavas (também conhecidas como escalas de análise), um fator escalar de 1,4 e representações através das camadas totalmente conectadas (TOĞAÇAR; CÖMERT; ERGEN, 2021; MORDVINTSEV; OLAH; TYKA, 2015; SUZUKI et al., 2017). Essas condições permitiram explorar os padrões disponíveis do menor ao maior nível de abstração, evidenciando os detalhes nas imagens resultantes. A Figura 11 mostra alguns exemplos das representações DD no contexto das amostras H&E.

4.5 Etapa 3 - Extração das *Deep Features* das Representações Grad-CAM, LIME e DD

Esta etapa consistiu em fornecer as imagens extraídas pelas técnicas Grad-CAM, LIME e DD para um modelo CNN 2, com o intuito de obter as *deep features* da camada

Figura 11: Representações DeepDream via imagens H&E originais (a), considerando o modelo VGG19 (b), ResNet50 (c), Inception V3 (d) e EfficientNet B2 (e).



Fonte: Elaborado pelo autor.

de *average pooling*, considerando o processo de *k-fold cross-validation* (TOĞAÇAR; CÖMERT; ERGEN, 2021; LIN; CHEN; YAN, 2013). Os vetores de características foram definidos pelas imagens H&E originais e as representações Grad-CAM, LIME e DD correspondentes. Algumas combinações testadas aqui foram: VGG19 (CNN 1) com ResNet50 (CNN 2); EfficientNet B2 (CNN 1) com Inception V3 (CNN 2); ResNet50 (CNN 1) com VGG19 (CNN 2); Inception V3 (CNN 1) com EfficientNet B2 (CNN 2). Os resultados foram utilizados como entrada para a próxima etapa.

4.6 Etapa 4 - Seleção das *Features* Ranqueadas via Algoritmo *ReliefF*

O algoritmo ReliefF foi aplicado para analisar cada vetor de características, identificando os elementos mais generalizáveis e relevantes. Este algoritmo depende de um método estatístico inspirado por treinamento baseado em instância (DUDA; HART; STORK, 2012), utilizando cálculos comparativos entre os dados contidos em cada instância durante o processo de ranqueamento. Deste modo, ReliefF estima a qualidade e relevância de cada característica, atribuindo pesos que determinam o melhor ranqueamento (URBANOWICZ et al., 2018). Neste contexto, foi aplicado o algoritmo ReliefF utilizando o pacote Weka (WAIKATO, 2019). Para a análise de desempenho, diferentes composições dos valores com melhor ranqueamento foram utilizadas, com 10, 20, 30, 40, 50 e 100 *features*, baseado na abordagem adotada em (RIBEIRO et al., 2019; OLIVEIRA et al., 2023).

4.7 Etapa 5 - Classificações Baseadas em Múltiplos Classificadores

Para analisar a capacidade de classificação das características, diferentes métodos foram explorados, como: RotF e Dec - baseados em *meta-learning*; MP, Log e SVM - baseados em função; RandF - baseado em árvore de decisão; KStar e IBk - baseados em *lazy learning* (DUDA; HART; STORK, 2012). Os múltiplos classificadores indicaram uma estratégia útil em identificar a combinação capaz de fornecer a melhor associação, analisando a acurácia, e especialmente as métricas área sob a curva ROC (AUC) e F1-Score.

Resultados e Discussão

Os modelos CNN foram aplicados diretamente nas imagens histológicas, conforme a metodologia (etapa 1). Os valores de acurácia fornecidos pelas arquiteturas são apresentados na Tabela 2.

Tabela 2: Valores de ACC (%) alcançados via modelos CNN aplicados diretamente nos conjuntos de imagens histológicas.

	UCSB	CR	LG	DISP
VGG19	59,88	85,60	71,50	89,66
EfficientNet B2	61,00	64,52	50,80	55,63
ResNet50	60,50	81,00	68,80	78,20
Inception V3	49,40	61,22	56,26	54,55

Ao aplicar as arquiteturas diretamente, desempenhos pouco significativos são notados, com os melhores valores de acurácia de 61% a 89,66%. O modelo VGG19 foi responsável pelos melhores desempenhos em três dos quatro *datasets*, CR (85,60%), LG (71,50%), e DISP (89,66%). EfficientNet B2 indicou o melhor desempenho para os dados UCSB (61%).

Em seguida, as combinações de CNN 1 e CNN 2 foram testadas com o intuito de verificar os desempenhos discriminativos dos modelos CNN 2 via representações Grad-CAM, LIME e DD. As combinações testadas aqui foram indicadas como VGG19 (CNN 1) com ResNet50 (CNN 2); EfficientNet B2 (CNN 1) com Inception V3 (CNN 2); ResNet50 (CNN 1) com VGG19 (CNN 2); Inception V3 (CNN 1) com EfficientNet B2 (CNN 2). As Tabelas 3, 4, 5 e 6 apresentam os valores de acurácia em cada

combinação, considerando os conjuntos UCSB, CR, LG e DISP.

Tabela 3: Valores de acurácia (%) via ResNet50 (CNN 2) com representações extraídas de VGG19 (CNN 1).

	CAM	LIME	DD
UCSB	56,60	50,00	56,40
CR	55,10	49,80	71,50
LG	54,20	59,30	62,70
DISP	63,50	63,20	58,60

Tabela 4: Valores de acurácia (%) da aplicação de Inception V3 (CNN 2), usando representações extraídas de Efficientnet B2 (CNN 1).

	CAM	LIME	DD
UCSB	59,77	49,54	48
CR	50,75	55	55,30
LG	56,60	53,53	58,77
DISP	49,82	51,32	56,77

Tabela 5: Valores de acurácia (%) com VGG19 (CNN 2) e representações extraídas de ResNet50 (CNN 1).

	CAM	LIME	DD
UCSB	51,85	52,00	56,59
CR	53,48	60,75	86,82
LG	53,02	60,94	80,66
DISP	51,81	76,21	77,21

Tabela 6: Valores de acurácia (%) obtidos por Efficientnet B2 (CNN 2), explorando as representações extraídas de Inception V3 (CNN 1).

	CAM	LIME	DD
UCSB	57,84	62,91	55,30
CR	47,42	53,18	65,30
LG	54,43	55,85	57,64
DISP	55,24	61,52	51,62

A partir desses resultados, é possível verificar que as representações extraídas diretamente dos modelos CNN 2 indicaram desempenhos de 49% a 86%, aproximadamente. Os desempenhos mais interessantes foram obtidos com VGG19 (CNN 2)

com representações extraídas de ResNet50 (CNN 1), especialmente através das imagens DD (Tabela 5). Os destaques foram CR, LG e DISP com valores de acurácia de 86,82%, 80,66% e 77,21%, respectivamente. Os valores de acurácia alcançados com os conjuntos CR (86,82%) e LG (80,66%) com representações DD ultrapassaram aqueles obtidos via imagens originais (Tabela 5). Quando consideradas as representações Grad-CAM e LIME, os resultados foram menos significativos independentemente da combinação de CNN 1 e CNN 2.

Após essas associações, o modelo proposto foi completamente aplicado para cada conjunto de dados H&E, considerando as diferentes combinações apresentadas na metodologia, especialmente a inclusão das outras etapas (seções 4.5 a 4.7). As composições dos valores de melhor ranqueamento, com 10, 20, 30, 40, 50 e 100 *deep features* (RIBEIRO et al., 2019; OLIVEIRA et al., 2023) são apresentadas nas Tabelas 7, 8, 9 e 10. Ao considerar os classificadores explorados nesse trabalho e o número de atributos, as taxas de AUC alcançadas por cada classificador foram usadas para calcular a média de desempenho em cada cenário de características. Com isso, os valores mais expressivos de AUC com o menor número de características foram destacados em negrito.

Tabela 7: Média de taxas AUC obtidas com múltiplos classificadores e *deep features* por meio da arquitetura ResNet50 (CNN 2), usando representações extraídas de VGG19 (CNN 1).

	Número de características					
	10	20	30	40	50	100
UCSB	0,963	0,971	0,967	0,965	0,963	0,974
CR	0,990	0,988	0,993	0,997	0,996	0,994
LG	0,982	0,991	0,992	0,990	0,990	0,993
DISP	0,975	0,972	0,983	0,987	0,987	0,982

Ao considerar a Tabela 7 (arquitetura ResNet50 (CNN 2) com representações extraídas de VGG19 (CNN 1)), as maiores médias com o menor número de características foram taxas de AUC que variaram de 0,974 (UCSB) a 0,997 (CR). Esses valores foram definidos por 100 descritores para os conjuntos UCSB (0,974) e LG (0,993), e

Tabela 8: Média de taxas AUC obtidas com múltiplos classificadores e *deep features* via Inception V3 (CNN 2), processando as representações extraídas de Efficientnet B2 (CNN 1).

	Número de características					
	10	20	30	40	50	100
UCSB	0,926	0,930	0,944	0,933	0,937	0,938
CR	0,956	0,976	0,976	0,988	0,987	0,986
LG	0,942	0,966	0,973	0,976	0,977	0,987
DISP	0,958	0,953	0,964	0,974	0,974	0,979

Tabela 9: Médias de taxas AUC considerando múltiplos classificadores e *deep features* extraídas de VGG19 (CNN 2), usando as representações obtidas por ResNet50 (CNN 1).

	Número de características					
	10	20	30	40	50	100
UCSB	0,922	0,916	0,914	0,912	0,923	0,931
CR	0,980	0,991	0,991	0,992	0,993	0,996
LG	0,960	0,957	0,951	0,958	0,961	0,968
DISP	0,955	0,969	0,973	0,968	0,974	0,974

Tabela 10: Médias de taxas AUC obtidas com múltiplos classificadores e *deep features* via Efficientnet B2 (CNN 2), explorando as representações extraídas de Inception V3 (CNN 1).

	Número de características					
	10	20	30	40	50	100
UCSB	0,930	0,946	0,955	0,965	0,965	0,967
CR	0,988	0,994	0,992	0,991	0,993	0,995
LG	0,943	0,961	0,972	0,986	0,987	0,987
DISP	0,927	0,953	0,946	0,953	0,954	0,965

40 para CR (0,997) e DISP (0,987). Quando é observada a associação entre Efficientnet B2 (CNN 1) e Inception V3 (CNN 2) (Tabela 8), é notado que as maiores médias com o menor número de atributos foram valores AUC de 0,944 (UCSB) a 0,988 (CR), com 30 atributos para UCSB (0,944), 40 para CR (0,988), e 100 para LG (0,987) e DISP (0,979). Pela Tabela 9 (modelo VGG19 (CNN 2) usando representações extraídas por ResNet50 (CNN 1)), os resultados mais expressivos indicaram taxas AUC de 0,931 (UCSB) a 0,996 (CR). Os conjuntos UCSB, CR e LG (0,968) foram classificados com 100 características. Nessa combinação, o conjunto DISP (0,974 de AUC) foi classificado com 50 *deep features*. Levando em consideração a associação Inception V3 (CNN 1) e Efficientnet B2 (CNN 2), Tabela 10, os valores médios de AUC também representaram desempenhos importantes, de 0,965 (DISP) a 0,995 (CR), explorando 100 (UCSB, CR e DISP) e 50 (LG) atributos. Sendo assim, após aplicar o modelo completo, é possível identificar as principais combinações para cada *dataset* H&E, incluindo o classificador mais relevante em cada contexto. As melhores associações, segundo as maiores médias de AUC vistas nas tabelas anteriores, estão sumarizadas na Tabela 11, com os valores F1-Score e ACC inclusos.

Tabela 11: Sumário das melhores associações resultantes do modelo proposto.

	CNN 1 e CNN 2	Características	Classificador	AUC	ACC	F1-Score
UCSB	VGG19 com ResNet50	100	MP	0,994	96,50	0,965
CR	VGG19 com ResNet50	40	IBk	1	100	1
LG	VGG19 com ResNet50	100	RotF	0,999	99,20	0,992
DISP	VGG19 com ResNet50	40	Dec	0,998	97,90	0,980

Esse sumário indica que os valores de ACC mais significativos são maiores que 96% e as taxas de AUC maiores que 0,99, com uma classificação ideal observada no conjunto CR. É possível notar que a combinação de VGG19 com ResNet50 forneceu as melhores soluções para todos os conjuntos de dados investigados aqui, indicando um padrão importante na classificação de múltiplos tipos de imagens H&E. Também, conjuntos com 40 e 100 características foram os mais usados. Apesar disso, os classificadores foram diferentes. Ao considera a métrica F1-Score, a qualidade alcançada por

cada modelo também merece ser destacada, com valores maiores que 0,96 em diferentes tipos de amostras H&E. Ademais, distinções significativas foram alcançadas em comparação com os desempenhos dos modelos aplicados diretamente às imagens originais (Tabela 2), uma abordagem comumente vista nesse campo específico de estudo. Por exemplo, as melhores soluções indicaram valores de ACC que variam de 96,50% (UCSB) a 100% (CR) em comparação a 60,50% (UCSB) a 81% (CR) via ResNet50. A diferença entre esses resultados indica os ganhos que foram de 19% (CR) a 36% (UCSB) com o modelo proposto. Outros ganhos significantes podem ser calculados dos testes discutidos previamente, mas os valores indicados aqui são suficientes para destacar as contribuições deste estudo.

Para destacar as origens dos atributos, na Tabela 12 é exibido a distribuição percentual dos atributos que definiram as melhores associações (Tabela 11). É observado que as características resultantes das imagens originais foram aquelas com o maior ocorrência (maior que 85%), seguidas das representações LIME, DD e Grad-CAM. Estas composições são vistas nos conjuntos UCSB, LG e DISP, indicando modelos próximos ao ideal (taxas AUC de 0,994 a 0,999). Em adição, apesar da solução para o conjunto CR não considerar atributos das representações, o esquema proposto identificou uma solução com a capacidade de proporcionar a melhor distinção.

Tabela 12: Porcentagem (%) da distribuição de atributos que compuseram as melhores soluções.

Dataset	Features	Porcentagem (%)			
		DD	Grad-CAM	LIME	Originais
UCSB	100	3	2	9	86
CR	40	0	0	0	100
LG	100	1	0	0	99
DISP	40	0	0	5	95

Por fim, as Tabelas 13, 14, 15 e 16 mostram uma visão geral dos resultados alcançados com este trabalho em relação a importantes abordagens disponíveis na Literatura. É importante notar que estas propostas levaram a modelos quase ideais, a

medida que utilizam diferentes tipos de características, classificadores, e ajustes a fim de analisar *datasets* específicos. Neste contexto, é observado que as soluções obtidas neste estudo foram capazes de indicar valores de acurácia entre aqueles com melhor ranqueamento, com ênfase nos contextos de câncer colorretal e displasia oral como principais abordagens. A respeito do conjunto LG, o modelo obtido expressou um desempenho sutilmente menor (0,42%) quando comparado ao descrito por Roberto et al. em (ROBERTO et al., 2021). Considerando as amostras de câncer de mama, a solução apresentada é a segunda melhor ranqueada, com um valor de acurácia 1,5% menor. Outras comparações poderiam ser consideradas visando garantir estes resultados em mais condições e experimentos, ou até mesmo fazer ajustes para definir os limites de cada modelo. No entanto, é válido notar também que combinações de estratégias CNN com representações Grad-CAM, LIME e DD e múltiplos classificadores não foram usadas nestes trabalhos relacionados. Então, é possível evidenciar outra contribuição deste estudo, uma estratégia viável para melhorar desempenhos de métodos tradicionais de *deep learning*.

Tabela 13: Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de imagens de câncer de mama.

Referência	Método	Acurácia
(OLIVEIRA et al., 2023)	ResNet50, <i>avg_pool layer</i> , RelieF e <i>Ensemble</i> com 35 <i>deep features</i>	98,00
Proposto	VGG19+ResNet50 com DeepDream, Grad-CAM e LIME + classificador MP	96,50
(NANNI et al., 2020)	9 modelos CNN, <i>handcrafted features</i>	96,33
(NANNI; GHIDONI; BRAHNAM, 2018)	8 modelos CNN, <i>handcrafted features</i>	95,00
(KAUSAR et al., 2019)	<i>Color normalization</i> , Haar wavelet e CNN proposta	91,00
(CANDELERO et al., 2020)	Le-Net, <i>handcrafted features</i> multiescala e multidimensionais, Haralick, LBP	90,52
(ROBERTO et al., 2021)	ResNet50 com <i>fine-tuning</i> , <i>handcrafted features</i> multiescala e multidimensionais	89,66
(SETHY; BEHERA, 2022)	3 modelos CNN e <i>handcrafted features</i> (GLCM, HOG, LBP)	84,20

Tabela 14: Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de câncer colorretal.

Referência	Método	Acurácia
Proposto	VGG19+ResNet50 com DeepDream, Grad-CAM and LIME + classificador IBk	100
(ROBERTO et al., 2021)	ResNet50 com <i>fine-tuning</i> , <i>handcrafted features</i> multiescala e multidimensionais	99,39
(OLIVEIRA et al., 2023)	ResNet50, <i>avg_pool layer</i> , RelieF e <i>Ensemble</i> com 35 <i>deep features</i>	98,00
(NANNI; GHIDONI; BRAHNAM, 2018)	8 modelos CNN, <i>handcrafted features</i>	97,60
(NANNI et al., 2020)	9 modelos CNN, <i>handcrafted features</i>	97,50
(NANNI et al., 2019)	6 modelos CNN, <i>handcrafted features</i>	97,00
(CANDELERO et al., 2020)	Le-Net, <i>handcrafted features</i> multiescala e multidimensionais, Haralick, LBP	91,06

Tabela 15: Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de imagens de tecido hepático.

Referência	Método	Acurácia
(ROBERTO et al., 2021)	ResNet50 com <i>fine-tuning</i> , <i>handcrafted features</i> multiescala e multidimensionais	99,62
(OLIVEIRA et al., 2023)	ResNet50, <i>avg_pool layer</i> , RelieF e <i>Ensemble</i> com 5 <i>deep features</i>	99,32
Proposto	VGG19+ResNet50 com DeepDream, Grad-CAM e LIME + classificador RotF	99,20
(ANDREARCZYK; WHELAN, 2017)	<i>Collective</i> T-CNN	98,20
(HUANG et al., 2013)	Novo conjunto de características de imagem e <i>Ensemble</i> de classificador SVM	97,01
(WATANABE; KOBAYASHI; WADA, 2016)	Descritores GIST, PCA e LDA	93,70

Tabela 16: Visão geral dos valores de acurácia (%) obtidos por diferentes abordagens para a classificação de displasia oral.

Referência	Método	Acurácia
Proposto	VGG19+ResNet50 com DeepDream, Grad-CAM e LIME + classificador Dec	97,90
(AZARMEHR et al., 2022)	Busca em arquitetura neural e <i>handcrafted features</i> (morfológico e não-morfológico)	95,20
(DABASS; VIG; VASHISTH, 2018)	Fatiamento <i>Bit-plane</i> , Algoritmo <i>active contour snake</i> e regiões de queratina	95,08
(ADEL et al., 2018)	<i>Handcrafted features</i> (SIFT, SURF, ORB)	92,80
(SILVA et al., 2022)	<i>Handcrafted features</i> (morfológico e não-morfológico)	92,40
(BAIK et al., 2014)	Morfologia, quantidade de DNA, distribuição de cromatina e <i>Random Forests</i>	80,60

Conclusões

Neste trabalho, foi desenvolvida uma estratégia capaz de fornecer soluções baseadas em modelos CNN com representações Grad-CAM, LIME e DD, incluindo múltiplos classificadores, com o objetivo de classificar imagens H&E. Os contextos de investigação foram câncer de mama, câncer colorretal, tecido hepático e displasia oral. Os melhores resultados foram obtidos por meio da associação de redes VGG19 (CNN 1) com ResNet50 (CNN 2), explorando vetores de características com 40 e 100 atributos. Os resultados foram valores de acurácia que variaram entre 96,50% (UCSB) a 100% (CR). A métrica F1-Score revelou a qualidade alcançada por cada modelo, com valores significantes (maiores que 0,96) em diferentes tipos de amostras H&E. As *deep features* mais usadas foram baseadas nas imagens originais (maior ocorrência), seguidas das representações LIME, DD e Grad-CAM. Ao considerar os conjuntos UCSB, LG e DISP, as combinações contribuíram para resultados próximos aos ideais (taxas AUC de 0,994 a 0,999).

Os principais resultados foram comparados com técnicas consolidadas de aprendizado de máquina, modelos CNN aplicados diretamente para classificar os conjuntos H&E. Foi visto que os modelos principais expressaram importantes desempenhos em ganhos, com os maiores valores de acurácia: as diferenças variaram de 19% (CR) a 36% (UCSB). Portanto, as informações e combinações descritas neste trabalho podem auxiliar significativamente a expansão de técnicas consolidadas voltadas ao reconhecimento de padrões de imagens histológicas, fornecendo mais robustez aos diagnósticos

auxiliados por computador.

Por fim, considerando uma visão geral ilustrativa das soluções sobre as abordagens significativas disponíveis na Literatura, os desempenhos alcançados aqui foram competitivos, com valores entre os melhor classificados, com ênfase nos contextos de câncer colorretal e displasia oral (melhores soluções). Para os outros conjuntos de dados, os modelos obtidos indicaram taxas sutilmente menores, fazendo com que sejam as segundas melhores soluções.

Em trabalhos futuros, pretende-se que: mais modelos CNN sejam combinados com as representações Grad-CAM, LIME e DD correspondentes para definir outras vantagens e limites; testar as influências dos diferentes parâmetros de cada técnica para a obtenção de novas ideias e perspectivas para a análise de imagens H&E.

6.1 Contribuições e Divulgação

As principais contribuições alcançadas com esse trabalho foram as seguintes:

- Uma estratégia baseada em representações LIME, Grad-CAM e DeepDream ainda não explorada para reconhecimento de padrões de imagens H&E relevantes;
- Definição das melhores soluções por meio das principais representações (LIME, Grad-CAM e DD), suas *deep features* correspondentes, e a utilização de diferentes classificadores;
- Um esquema computacional capaz de associar quatro arquiteturas CNN distintas e consolidadas para fornecer os melhores modelos;
- Resultados que ultrapassaram aqueles indicados em abordagens importantes disponíveis na Literatura especializada, em classificação de câncer colorretal e displasia oral (melhores soluções).

Além do exposto, o trabalho foi divulgado nos seguintes veículos:

- Martinez, J. M. C.; Neves, L. A. Ensemble com técnicas de aprendizado de máquina e IA explicável para classificar imagens histológicas, XXXIV Congresso de Iniciação Científica da Unesp (CIC), 2022.
- O trabalho foi selecionado para a 2ª fase da conferência, realizada em novembro de 2022.
- Neves L., Martinez J., Longo L., Freire Roberto G., Tosta T., de Faria P., Loyola A., Cardoso S., Silva A., Zanchetta do Nascimento M. and Rozendo G. (2023). Classification of H&E Images via CNN Models with XAI Approaches, DeepDream Representations and Multiple Classifiers. In Proceedings of the 25th International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-648-4, SciTePress, pages 354-364.
- Artigo selecionado para publicação no journal "SN Computer Science", "Recent Trends on Enterprise Information Systems", como parte da 25th International Conference on Enterprise Information Systems: Martinez J., Neves L., Longo L., Freire Roberto G., Tosta T., de Faria P., Loyola A., Cardoso S., Silva A., Zanchetta do Nascimento M. and Rozendo G., "Association of Deep Features with DeepDream Representations, XAI Techniques and Multiple Classifiers for the Classification of Histological Images".

Referências

ABIR, W.; UDDIN, M.; KHANAM, F.; TAZIN, T.; KHAN, M.; MASUD, M.; AL-JAHDALI, S. Explainable ai in diagnosing and anticipating leukemia using transfer learning method. *Computational Intelligence and Neuroscience*, Hindawi, v. 2022, p. 1–14, 04 2022. Disponível em: <https://doi.org/10.1155/2022/5140148>.

ADADI, A.; BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, PP, p. 1–1, 09 2018.

ADEL, D.; MOUNIR, J.; EL-SHAFFEY, M.; ELDIN, Y. A.; MASRY, N. E.; ABDEL-RAOUF, A.; ELHAMID, I. S. A. Oral epithelial dysplasia computer aided diagnostic approach. In: IEEE. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. [S.l.], 2018. p. 313–318.

AFFONSO, C.; ROSSI, A. L. D.; VIEIRA, F. H. A.; FERREIRA, A. C. P. de L. et al. Deep learning for biological image classification. *Expert Systems with Applications*, Elsevier, v. 85, p. 114–122, 2017.

AGEMAP, N. I. o. A. *The Atlas of Gene Expression in Mouse Aging Project (AGEMAP)*. 2020. <https://ome.grc.nia.nih.gov/iicbu2008/agemap/index.html>. Acesso em: 04/05/2020.

AHA, D. W. Lazy learning. In: *Lazy learning*. [S.l.]: Springer, 1997. p. 7–10.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine learning*, Springer, v. 6, p. 37–66, 1991.

ANDREARCZYK, V.; WHELAN, P. F. Deep learning for biomedical texture image analysis. In: IRISH PATTERN RECOGNITION & CLASSIFICATION SOCIETY (IPRCS). [S.l.], 2017.

AZARMEHR, N.; SHEPHARD, A.; MAHMOOD, H.; RAJPOOT, N.; KHURRAM, S. A. Automated oral epithelial dysplasia grading using neural networks and feature analysis. In: *Medical Imaging with Deep Learning*. [s.n.], 2022. Disponível em: <https://openreview.net/forum?id=ABl-dIO4g74>.

BAIK, J.; YE, Q.; ZHANG, L.; POH, C.; ROSIN, M.; MACAULAY, C.; GUILLAUD, M. Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. *Cellular Oncology*, Springer, v. 37, p. 193–202, 2014.

- BREIMAN, L. Random forests. *Machine Learning*, Springer, v. 45, p. 5–32, 10 2001.
- CANDELERO, D.; ROBERTO, G. F.; NASCIMENTO, M. Z. D.; ROZENDO, G. B.; NEVES, L. A. Selection of cnn, haralick and fractal features based on evolutionary algorithms for classification of histological images. In: IEEE. *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.], 2020. p. 2709–2716.
- CESSIE, S. L.; HOUWELINGEN, J. C. V. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, [Wiley, Royal Statistical Society], v. 41, n. 1, p. 191–201, 1992. ISSN 00359254, 14679876. Disponível em: <http://www.jstor.org/stable/2347628>.
- COCCIA, M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. *Technology in Society*, Elsevier, v. 60, p. 101198, 2020.
- DABASS, M.; VIG, R.; VASHISTH, S. Five-grade cancer classification of colon histology images via deep learning. In: *CRC Press*. [S.l.: s.n.], 2018. p. 18.
- DABEER, S.; KHAN, M. M.; ISLAM, S. Cancer diagnosis in histopathological image: Cnn based approach. *Informatics in Medicine Unlocked*, Elsevier, p. 100231, 2019.
- DANASINGH, A. A.; BALAMURUGAN, S.; EPIPHANY, J. L. Literature review on feature selection methods for high-dimensional data. *International Journal of Computer Applications*, v. 136, 02 2016.
- DONG, N.; ZHAO, L.; WU, C.; CHANG, J. Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, v. 93, p. 106311, 2020. ISSN 1568-4946. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494620302519>.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.
- FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <https://www.sciencedirect.com/science/article/pii/S016786550500303X>.
- FERRER-SÁNCHEZ, A.; BAGAN, J.; VILA-FRANCÉS, J.; MAGDALENA-BENEDITO, R.; BAGAN-DEBON, L. Prediction of the risk of cancer and the grade of dysplasia in leukoplakia lesions using deep learning. *Oral Oncology*, v. 132, p. 105967, 2022. ISSN 1368-8375. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1368837522002561>.
- GELASCA, E. D.; BYUN, J.; OBARA, B.; MANJUNATH, B. Evaluation and benchmark for biological image segmentation. In: *IEEE International Conference on Image Processing*. [s.n.], 2008. Disponível em: http://vision.ece.ucsb.edu/publications/elisa_ICIP08.pdf.

GREENSPAN, H.; GINNEKEN, B. V.; SUMMERS, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, IEEE, v. 35, n. 5, p. 1153–1159, 2016.

GUYON, I.; GUNN, S.; NIKRAVESH, M.; ZADEH, L. *Feature Extraction: Foundations and Applications*. Springer Berlin Heidelberg, 2006. (Studies in Fuzziness and Soft Computing). ISBN 9783540354871. Disponível em: <https://books.google.com.br/books?id=x5hdbK8bIG0C>.

HAM, H.-S.; LEE, H.-S.; CHAE, J.-W.; CHO, H. C.; CHO, H.-C. Improvement of gastroscopy classification performance through image augmentation using a gradient-weighted class activation map. *IEEE Access*, IEEE, v. 10, p. 99361–99369, 2022.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

HOCHREITER, S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, World Scientific, v. 6, n. 02, p. 107–116, 1998.

HUANG, H.-L.; HSU, M.-H.; LEE, H.-C.; CHAROENKWAN, P.; HO, S.-J.; HO, S.-Y. Prediction of mouse senescence from he-stain liver images using an ensemble svm classifier. In: SPRINGER. *Intelligent Information and Database Systems: 5th Asian Conference, ACIIDS 2013, Kuala Lumpur, Malaysia, March 18-20, 2013, Proceedings, Part II* 5. [S.l.], 2013. p. 325–334.

IMAGENET. *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*. 2021. Disponível em: <http://www.image-net.org/challenges/LSVRC/>.

Instituto Nacional de Câncer José Alencar Gomes da Silva. *Síntese de estimativas para casos de câncer - INCA*. 2023. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/sintese-de-resultados-e-comentarios>.

JIANG, B.; ZHANG, Y.; ZHANG, L.; BOCK, G.; VLIEGENTHART, R.; XIE, X. Human-recognizable ct image features of subsolid lung nodules associated with diagnosis and classification by convolutional neural networks. *European Radiology*, v. 31, 04 2021.

KARLIK, B.; OLGAC, A. V. Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, v. 1, n. 4, p. 111–122, 2011.

KAUSAR, T.; WANG, M.; IDREES, M.; LU, Y. Hwdcnn: Multi-class recognition in breast histopathology with haar wavelet decomposed image based convolution neural network. *Biocybernetics and Biomedical Engineering*, Elsevier, v. 39, n. 4, p. 967–982, 2019.

- KHAN, I.; ZHANG, X.; REHMAN, M.; ALI, R. A literature survey and empirical study of meta-learning for classifier selection. *IEEE Access*, IEEE, v. 8, p. 10262–10281, 2020.
- KIRA, K.; RENDELL, L. A. A practical approach to feature selection. In: *Machine Learning Proceedings 1992*. [S.l.]: Elsevier, 1992. p. 249–256.
- KONONENKO, I.; ŠIMEC, E.; ROBNIK-ŠIKONJA, M. Overcoming the myopia of inductive learning algorithms with relief. *Applied Intelligence*, Springer, v. 7, n. 1, p. 39–55, 1997.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, may 2017. ISSN 0001-0782. Disponível em: <https://doi.org/10.1145/3065386>.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Taipei, Taiwan, v. 86, n. 11, p. 2278–2324, 1998.
- LI, Y.-x.; CHEN, F.; SHI, J.-j.; HUANG, Y.-l.; WANG, M. Convolutional neural networks for classifying cervical cancer types using histological images. *Journal of Digital Imaging*, Springer, v. 36, n. 2, p. 441–449, 2023.
- LIN, M.; CHEN, Q.; YAN, S. *Network In Network*. arXiv, 2013. Disponível em: <https://arxiv.org/abs/1312.4400>.
- LIN, M.; CHEN, Q.; YAN, S. *Network In Network*. 2014.
- LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.
- MAHENDRAN, A.; VEDALDI, A. Visualizing deep convolutional neural networks using natural pre-images. *International Journal of Computer Vision*, v. 120, 12 2016.
- MELVILLE, P.; MOONEY, R. J. Constructing diverse classifier ensembles using artificial training examples. In: *Ijcai*. [S.l.: s.n.], 2003. v. 3, p. 505–510.
- MORDVINTSEV, A.; OLAH, C.; TYKA, M. Inceptionism: Going deeper into neural networks. 2015.
- NAMATĚVS, I. Deep convolutional neural networks: Structure, feature extraction and training. *Information Technology and Management Science*, Sciendo, v. 20, n. 1, p. 40–47, 2017.

NANNI, L.; BRAHNAM, S.; GHIDONI, S.; MAGUOLO, G. General purpose (genp) bioimage ensemble of handcrafted and learned features with data augmentation. *CoRR*, abs/1904.08084, 2019. Disponível em: <http://arxiv.org/abs/1904.08084>.

NANNI, L.; GHIDONI, S.; BRAHNAM, S. Ensemble of convolutional neural networks for bioimage classification. *Applied Computing and Informatics*, Elsevier, 2018.

NANNI, L.; GHIDONI, S.; BRAHNAM, S.; LIU, S.; ZHANG, L. *Ensemble of Hand-crafted and Deep Learned Features for Cervical Cell Classification*. Cham: Springer, 2020. v. 186. 117-135 p.

OLIVAS JOSE DAVID MARTIN GUERRERO, M. M. S. J. R. M. B. A. J. S. L. E. S. *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques*. [S.l.]: Information Science Reference-Imprint of: IGI Publishing, 2009.

OLIVEIRA, C. I. de; NASCIMENTO, M. Z. do; ROBERTO, G. F.; TOSTA, T. A.; MARTINS, A. S.; NEVES, L. A. Hybrid models for classifying histological images: An association of deep features by transfer learning with ensemble classifier. *Multimedia Tools and Applications*, Springer, p. 1–24, 2023. Disponível em: <https://doi.org/10.1007/s11042-023-16351-4>.

PEDRINI, H.; SCHWARTZ, W. *Análise De Imagens Digitais: Princípios, Algoritmos E Aplicações*. Cengage Learning, 2007. ISBN 9788522105953. Disponível em: <https://books.google.com.br/books?id=13KAPgAACAAJ>.

RAJARAMAN, S.; CANDEMIR, S.; KIM, I.; THOMA, G.; ANTANI, S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 8, n. 10, p. 1715, 2018.

RAJARAMAN, S.; CANDEMIR, S.; KIM, I. C.; THOMA, G.; ANTANI, S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences*, v. 8, p. 1715, 09 2018.

REYES, M.; MEIER, R.; PEREIRA, S.; SILVA, C. A.; DAHLWEID, F.-M.; TENGG-KOBLIGK, H. v.; SUMMERS, R. M.; WIEST, R. On the interpretability of artificial intelligence in radiology: challenges and opportunities. *Radiology: artificial intelligence*, Radiological Society of North America, v. 2, n. 3, p. e190043, 2020.

RIBEIRO, M. G.; NEVES, L. A.; NASCIMENTO, M. Z. do; ROBERTO, G. F.; MARTINS, A. S.; TOSTA, T. A. A. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. *Expert Systems with Applications*, Elsevier, v. 120, p. 262–278, 2019.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY,

USA: Association for Computing Machinery, 2016. (KDD '16), p. 1135–1144. ISBN 9781450342322. Disponível em: <https://doi.org/10.1145/2939672.2939778>.

ROBERTO, G. F.; LUMINI, A.; NEVES, L. A.; NASCIMENTO, M. Z. do. Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. *Expert Systems with Applications*, Elsevier, v. 166, p. 114103, 2021.

RODRIGUEZ, J.; KUNCHEVA, L.; ALONSO, C. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 10, p. 1619–1630, 2006.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.

SAMEK, W.; WIEGAND, T.; MÜLLER, K.-R. *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*. 2017.

SANTOS, F. P. dos; PONTI, M. A. Alignment of local and global features from multiple layers of convolutional neural network for image classification. In: IEEE. *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2019. p. 241–248.

SETHY, P. K.; BEHERA, S. K. Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis. *Multi-media Tools and Applications*, Springer, v. 81, n. 7, p. 9631–9643, 2022.

SHALLU; MEHRA, R. Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express*, Korean Institute of Communications and Information Sciences (KICS), v. 4, n. 1, p. 248, 2018.

SILVA, A. B.; MARTINS, A. S.; TOSTA, T. A. A.; NEVES, L. A.; SERVATO, J. P. S.; ARAÚJO, M. S. de; FARIA, P. R. de; NASCIMENTO, M. Z. do. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Systems with Applications*, Elsevier, v. 193, p. 116456, 2022.

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

SIRINUKUNWATTANA, K.; PLUIM, J. P.; CHEN, H.; QI, X.; HENG, P.-A.; GUO, Y. B.; WANG, L. Y.; MATUSZEWSKI, B. J.; BRUNI, E.; SANCHEZ, U. et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, Elsevier, v. 35, p. 489–502, 2017.

SOUSA, I. P. D.; VELLASCO, M. M. B. R.; SILVA, E. C. D. Local interpretable model-agnostic explanations for classification of lymph node metastases. *Sensors (Basel, Switzerland)*, Multidisciplinary Digital Publishing Institute (MDPI), v. 19, n. 13, 2019.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

SUZUKI, K.; ROSEBOOM, W.; SCHWARTZMAN, D. J.; SETH, A. K. A deep-dream virtual reality platform for studying altered perceptual phenomenology. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–11, 2017.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2818–2826.

TAN, M.; LE, Q. V. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. 2020.

TOĞAÇAR, M.; CÖMERT, Z.; ERGEN, B. Enhancing of dataset using deepdream, fuzzy color image enhancement and hypercolumn techniques to detection of the alzheimer's disease stages by deep learning model. *Neural Computing and Applications*, Springer, p. 1–13, 2021.

URBANOWICZ, R. J.; MEEKER, M.; CAVA, W. L.; OLSON, R. S.; MOORE, J. H. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, Elsevier, v. 85, p. 189–203, 2018.

VAPNIK, V. N. Pattern recognition using generalized portrait method. *Automation and remote control*, v. 24, n. 6, p. 774–780, 1963.

VEDALDI, A.; ZISSERMAN, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *preprint*, 12 2013.

WAIKATO, T. U. of. *weka Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. 2019. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>.

WATANABE, K.; KOBAYASHI, T.; WADA, T. Semi-supervised feature transformation for tissue image classification. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 12, p. e0166413, 2016.

- XIE, P.; ZUO, K.; LIU, J.; CHEN, M.; ZHAO, S.; KANG, W.; LI, F. et al. Interpretable diagnosis for whole-slide melanoma histology images using convolutional neural network. *Journal of healthcare engineering*, Hindawi, v. 2021, 2021.
- YING, X. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, IOP Publishing, v. 1168, n. 2, p. 022022, feb 2019. Disponível em: <https://dx.doi.org/10.1088/1742-6596/1168/2/022022>.
- YOSINSKI, J.; CLUNE, J.; NGUYEN, A.; FUCHS, T.; LIPSON, H. Understanding neural networks through deep visualization. 06 2015.
- YU, D.; WANG, H.; CHEN, P.; WEI, Z. Mixed pooling for convolutional neural networks. *Rough Sets and Knowledge Technology*, p. 364–375, 2014.
- ZHOU, B.; KHOSLA, A.; LAPEDRIZA, A.; OLIVA, A.; TORRALBA, A. Learning deep features for discriminative localization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2921–2929.
- ZHOU, S. K.; GREENSPAN, H.; SHEN, D. *Deep learning for medical image analysis*. [S.l.]: Academic Press, 2017.