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

Leonardo Henrique da Costa Longo

Exploração de *Deep features* via Transfer Learning com Técnicas Fractais para Classificar Imagens H&E

São José do Rio Preto
2022

Leonardo Henrique da Costa Longo

Exploração de *Deep features* via Transfer Learning com Técnicas Fractais para Classificar Imagens H&E

Dissertação apresentada ao Programa de Pós-Graduação 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, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Orientador:

Prof. Dr. Leandro Alves Neves

São José do Rio Preto
2022

L856e

Longo, Leonardo Henrique da Costa
Exploração de Deep features via Transfer Learning com Técnicas Fractais para Classificar Imagens H&E / Leonardo Henrique da Costa Longo. -- São José do Rio Preto, 2022
121 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Leandro Alves Neves

1. Processamento de imagens. 2. Aprendizado do computador. 3. Redes neurais (Computação). 4. Fractais. 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.

Leonardo Henrique da Costa Longo

Exploração de *Deep features* via Transfer Learning com Técnicas Fractais para Classificar Imagens H&E

Dissertação apresentada ao Programa de Pós-Graduação 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, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Comissão Examinadora:

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

Prof. Dr. Fabricio Aparecido Breve
UNESP – Câmpus de Rio Claro

Prof. Dr. Jefferson Rodrigo de Souza
Universidade Federal de Uberlândia

São José do Rio Preto
12 de dezembro de 2022

Aos meus irmãos Elio, Felipe e Laura.

Agradecimentos

Agradeço primeiramente ao meu orientador Prof. Dr. Leandro Alves Neves, que nunca exitou em me auxiliar desde o primeiro momento que trabalhamos juntos e que hoje considero um grande amigo. Sem ele não seria possível realizar esse trabalho.

Agradeço também aos meus amigos e familiares que me apoiaram durante toda minha jornada acadêmica, me motivando nos momentos desafiadores e celebrando comigo nos momentos de alegria. Foram essas pessoas que me permitiram continuar trabalhando com um sorriso no rosto.

O orientador deste projeto agradece o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), #313643/2021-0, e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), #APQ-01129-21 e #APQ-00578-18.

"O mundo é muito ruim e quanto mais você estuda pior ele fica."

- Laurinha Lero

Resumo

Neste trabalho é apresentada uma proposta para investigar combinações de descritores *handcrafted* e *deep learned*, bem como possíveis padrões de associações em diferentes tipos de imagens histológicas. Os descritores *handcrafted* foram definidos da categoria de técnicas fractais aplicadas às imagens originais, bem como às representações obtidas por meio de técnicas de inteligência artificial explicável. Os descritores *deep learned* foram obtidos de diferentes arquiteturas de redes neurais convolucionais. Os descritores mais relevantes de cada combinação, definidos a partir de um algoritmo de ranqueamento, foram analisados em um *ensemble* composto pelos classificadores SVM, *Naive Bayes*, *Random Forest* e K-*Nearest Neighbors*. A metodologia proposta foi aplicada em imagens H&E representativas do câncer de mama, colorretal, displasia oral e tecido hepático. Os melhores resultados em cada conjunto foram taxas de acurácias de 94,83% a 100%, além de: conhecer padrões de combinações de técnicas para diferentes tipos de imagens H&E; indicar que o *ensemble* de descritores permite aumentar as taxas conquistadas em diferentes contextos, sobretudo quando combinadas com *deep features* obtidas via *transfer learning*; definir que combinação de ensembles (descritores e classificadores) pode fornecer desempenhos competitivos em relação aos disponíveis na Literatura, explorando um número reduzido de características.

Palavras-chave: *Deep features*. Técnicas fractais. *Ensemble* de descritores. *Ensemble* de classificadores. Imagens histológicas H&E.

Abstract

In this project it is proposed a methodology to investigate the ensemble of handcrafted and deep learned features, as well as to find possible patterns in multiple associations on images from different histological origins. The handcrafted features were selected from fractal techniques, applied to the original images as well to their respective explanations obtained with explainable artificial inteligence approaches. On the other hand, the deep learned features were extracted from several convolutional neural networks. The most relevant features from each ensemble, according to a ranking algorithm, were fed to and ensemble composed by the classifiers SVM, Naive Bayes, Random Forest and K-Nearest Neighbors. Our proposed method was performed with H&E images representing breast cancer, colorrectal cancer, oral displasia and liver tissue. The best results from each context achieved accuracies ranging from 94.83% to 100%, besides that we also: found a pattern for ensemble techniques considering multiple types of H&E images; demonstrated that the ensemble of features allows to improve accuracy on several scenarios, specially with deep learned features acquired with transfer learning; defined which ensemble combination (features and classifiers) could provide competitive results agains other proposed methods on the literatures while using a small number os features.

Keywords: Deep features. Fractal techniques. Feature ensemble. Ensemble of classifiers. H&E histological images

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

Introdução

As técnicas de aprendizado de máquina são úteis para realizar tarefas de classificação e reconhecimento de padrões nos mais diferentes domínios. No contexto de imagens, por exemplo, as técnicas fundamentadas em Redes Neurais Artificiais (RNA) proporcionaram avanços importantes (LECUN; BENGIO; HINTON, 2015). Uma RNA é composta por camadas interconectadas, de forma sequencial ou não, capazes de processar dados em diferentes níveis de abstração. Os modelos obtidos a partir de uma RNA são resultados da técnica de *backpropagation*, que permite, por meio do método matemático do gradiente, reajustar os parâmetros das camadas.

Neste contexto, as técnicas Redes Neurais Convolucionais (CNN), variantes de RNA, utilizam a estratégia de fragmentar a informação em diversos níveis e escalas para verificar como estes se relacionam no espaço da imagem. Esse comportamento é obtido com a aplicação consecutiva de operações de convolução a partir de diferentes filtros. Assim, o algoritmo consegue verificar padrões locais e globais, independente da posição, da escala ou da rotação em que se encontram (WANG et al., 2019a).

Existem diferentes estratégias utilizadas para compor as arquiteturas de uma CNN. Por exemplo, uma abordagem comum é a de construir arquiteturas com muitas camadas, profundas, para obter um processamento mais detalhado em cada passo da análise. A arquitetura *Visual Geometry Group* (VGG) (SIMONYAN; ZISSERMAN, 2014) usa

desta estratégia ao expandir sua profundidade com a construção de camadas de convolução compostas de máscaras menores, mas em maior quantidade. Essas várias operações menores, quando em conjunto, tendem a convergir para modelos mais precisos. Já as arquiteturas *Residual Neural Network* (ResNet) (HE et al., 2016) aumentaram a profundidade de uma rede com a utilização de blocos residuais para a passagem de valores entre as camadas. Essa prática é empregada para evitar a degradação da informação, consequência do grande número de camadas presentes (SRIVASTAVA; GREFF; SCHMIDHUBER, 2015). Outra estratégia que pode ser utilizada é incrementar a largura da rede. Uma das arquiteturas que utiliza essa estratégia é a InceptionNet (SZEGEDY et al., 2015). Este tipo de CNN considera operações em escalas diferentes dentro de uma mesma camada e concatena os resultados obtidos para alimentar a próxima. Essa abordagem explora padrões importantes em imagens de uma mesma classe podem ser encontrados em diferentes escalas e posições, assim convoluções de diferentes tamanhos em um mesmo nível podem encontrá-los.

É importante destacar que apesar de os modelos CNN conseguirem realizar diretamente as classificações de imagens ao fim do processamento, os valores em suas camadas internas podem conter informações importantes para descrever a imagem de entrada, uma vez que podem ser interpretados como quantificações de possíveis padrões (LEI et al., 2016). Isso ocorre por meio das *deep learned features*, ou *deep features*, as quais foram responsáveis por resultados relevantes no contexto de imagens médicas (WANG et al., 2019b; MAHBOD et al., 2019; ALINSAIF; LANG, 2020). Esse uso foi viabilizado principalmente com a aplicação de *transfer learning*, que consiste em aproveitar um modelo CNN treinado em um conjunto de amostras diferente do explorado na investigação. Essa prática permite refinar o treinamento para um novo modelo a fim de melhorar as taxas de distinções (TORREY; SHAVLIK, 2010). Apesar disso, é possível aprimorar os resultados e melhorar o processo de reconhecimento de padrões ao combinar estes descritores com outras técnicas de quantificações de texturas em

imagens, tais como os descritores *handcrafted*.

Neste contexto, diferentes combinações de descritores proporcionaram resultados relevantes nos mais variados tipos de imagens (NANNI; GHIDONI; BRAHNAM, 2017; HAGERTY et al., 2019; WEI et al., 2019; HASAN et al., 2019; LI et al., 2019a; TRIPATHI; SINGH, 2020, 2020). Dentre as combinações possíveis, o uso de descritores fractais (*handcrafted*), tais como dimensão fractal e lacunaridade, merecem destaque por conseguirem mensurar apropriadamente formas complexas, geralmente encontradas na natureza (FORTIN et al., 1992). Esses descritores mensuram a uniformidade e preenchimento do espaço de imagem sob análise, informações que podem favorecer o processo de classificação e reconhecimento de padrões, tais como em imagens médicas (ARALICA et al., 2020; JAIN; DUIN; MAO, 2000; QIN; PUCKETT; QIAN, 2020; ROBERTO et al., 2021). Também fazem parte deste grupo os descritores derivados da percolação, técnica que permite exportar o conceito de porosidade para o processamento de imagens e também obteve resultados significativos na aplicação para classificação de imagens no contexto histológico (ROBERTO et al., 2017; ROBERTO et al., 2019; CANDELERO et al., 2020).

Adicionalmente, uma vez que o uso de redes neurais para a classificação ou obtenção de *deep features* possibilita resultados relevantes, torna-se necessário investigar as informações que embasaram as tomadas de decisões. De forma mais simples, é necessário observar se os modelos obtidos fornecem resultados precisos pelos motivos corretos (SAMEK; MÜLLER, 2019). Este questionamento motivou o desenvolvimento de técnicas concentradas na área de inteligência Artificial explicável (IAx) (GUNNING; AHA, 2019), tais como *Class Activation Mapping* (CAM) (ZHOU et al., 2016) e *Local Interpretable Model-Agnostic Explanations* (LIME) (RIBEIRO; SINGH; GUESTRIN, 2016). Essas abordagens podem ser aplicadas para produzir visualizações das regiões de imagens que fundamentaram o processo de classificação de uma CNN. Com isto, estas imagens permitem realçar os padrões dominantes para a

identificação de uma determinada categoria e, por consequência, a quantificação destas pode resultar em dados significativos para um novo processo de classificação. Apesar destas observações, nota-se que esta hipótese ainda não foi verificada na Literatura.

Mesmo que as estratégias apresentadas previamente tenham suas vantagens, como são explorados vetores de descritores de diferentes origens, é importante considerar o número de descritores empregados para a realização da classificação. Esta prática deve ser realizada para minimizar a possibilidade da ocorrência de *overfitting* (NANNI; GHIDONI; BRAHNAM, 2017; HAGERTY et al., 2019; TRIPATHI; SINGH, 2020). Adicionalmente, um conjunto de descritores também pode ser altamente dependente da heurística do classificador utilizado para avaliar o modelo (DIETTERICH, 2000). Estes problemas são minimizados com o uso de classificadores construídos a partir de diferentes heurísticas, os quais são combinados por uma técnica nomeada *ensemble learning classification* (*ensemble* de classificadores) (SAGI; ROKACH, 2018). Técnicas que fazem parte desse conjunto exploram a chamada sabedoria de multidão, de forma que uma decisão tomada a partir de diferentes perspectivas, se combinadas, pode ser mais precisa. A justificativa é que possíveis erros individuais são compensados pelos acertos dos outros componentes (SAGI; ROKACH, 2018). A classificação a partir destes métodos indicou resultados relevantes quando aplicados em imagens de diferentes contextos (NANNI; LUMINI; ZAFFONATO, 2018; CHEN et al., 2019; DAS; BISWAS, 2019; BRUNESE et al., 2020) e despertam o interesse da comunidade para a sua implementação aliada aos métodos de aprendizado profundo. Como consequência, é possível investigar as combinações de descritores e técnicas mais relevantes para a análise, classificação e reconhecimento de padrões nas imagens complexas e relevantes cientificamente, como amostras histológicas coradas com Hematoxilina Eosina (H&E).

1.1 Motivação e Justificativas

O processo de quantificação de imagens pode ser realizado por meio de diferentes técnicas (KUMAR; BHATIA, 2014). Por outro lado, o uso de múltiplas técnicas pode implicar em questões adicionais envolvendo um sistema de classificação, seja pela situação de *overfitting* (NANNI; GHIDONI; BRAHNAM, 2017; HAGERTY et al., 2019; TRIPATHI; SINGH, 2020) ou por uma heurística tornar-se dependente e/ou ajustar-se melhor a um tipo específico de descritor (DIETTERICH, 2000). Assim, a indicação da relevância de um determinado descritor está diretamente relacionada com o tipo de imagem, diversidade de técnicas e processo de classificação (NIXON; AGUADO, 2019).

De maneira complementar, os modelos CNN buscam estabelecer possíveis padrões presentes nas imagens a partir de quantificações automáticas em diferentes escalas. Isso gera, para uma mesma entrada, múltiplas quantificações de padrões globais e locais de uma textura sob investigação. Assim, essas quantificações podem ser utilizadas como descritores e, consequentemente, podem evitar a indicação explícita de quais técnicas devem ser empregadas. Trabalhos recentes testaram essa hipótese em diferentes contextos e obtiveram bons resultados (NANNI; LUMINI; ZAFFONATO, 2018; HAGERTY et al., 2019; ALMARAZ-DAMIAN et al., 2020; OHATA et al., 2021; KUMAR et al., 2022). Mesmo assim, é possível constatar que os autores não descartaram os descritores *handcrafted*, uma vez que estes podem conter informações que não são contabilizadas durante as operações de convolução dos modelos CNN. O uso de *ensemble* de descritores *handcrafted* e *deep learned* pode resultar também em classificações mais precisas, conforme destacado em (NANNI; GHIDONI; BRAHNAM, 2017; HAGERTY et al., 2019; WEI et al., 2019; HASAN et al., 2019; LI et al., 2019a; TRIPATHI; SINGH, 2020; TURKOGLU, 2021; KOC et al., 2022). Outro destaque é que propostas pautadas em *ensemble learning* permitem melhorar os desempenhos de classificações (NANNI et al., 2020), sendo indicadas como uma das

principais frentes para o desenvolvimento de novos modelos (KHAN et al., 2020). Também, o campo de IAx tem contribuído significativamente com o aprimoramento dos modelos *ensemble learning*, especialmente na validação e interpretação dos resultados (KHEDKAR et al., 2019; WELLS; BEDNARZ, 2021), a fim de garantir que as classificações corretas foram determinadas pelos motivos corretos.

É importante considerar que a obtenção de descritores a partir de uma CNN resulta em um número expressivo de dados para o vetor de características, fato que pode inviabilizar investigações de contextos com um número reduzido de amostras (BISHOP, 2006; WATANABE, 1985), tais como os de imagens histológicas H&E para os estudos de câncer colorretal, câncer de mama, displasia oral e tecidos hepáticos. Esta situação pode ser contornada com a combinação de técnica para selecionar os descritores mais relevantes, tais como o algoritmo *ReliefF* (KILICARSLAN; ADEM; CELIK, 2020; TURKOGLU, 2021). Mais ainda, esta estratégia motiva a investigação de combinações de diferentes associações (KUNCHEVA, 2014; DAS; BISWAS, 2019; BRUNESE et al., 2020; SUŁOT et al., 2021; ISLAM; NAHIDUZZAMAN, 2022).

Neste contexto, mesmo com algumas iniciativas observadas (LUMINI; NANNI, 2018; KAUSAR et al., 2019; ROBERTO et al., 2021; TOĞAÇAR; CÖMERT; ERGEN, 2021), modelos de *ensemble learning* explorando descritores fractais multiescalas e multidimensionais, como os aqui investigados, ainda não foram totalmente explorados na literatura, especialmente por meio de quantificações de imagens representativas de CAM e LIME. Como diferenciais dessas estratégias, é possível investigar, por exemplo: 1. se é possível definir padrões de combinações de técnicas para diferentes tipos de imagens H&E; 2. se descritores fractais multiescalas e multidimensionais permitem aprimorar os resultados conquistados via *deep features* com *transfer learning* em imagens H&E; 3. se descritores fractais obtidos a partir de representações CAM e LIME podem contribuir com os desempenhos de um *ensemble learning*; 4. se a combinação de ensembles (descritores e classificadores) pode indicar desempenhos

competitivos em relação aos disponíveis na literatura.

Por fim, a investigação proposta para classificação e o reconhecimento de padrões em amostras H&E representativas do câncer colorretal, câncer de mama, displasia oral e tecido hepático pode contribuir significativamente com informações relevantes para a área de aprendizado de máquina e especialistas interessados nos temas aqui explorados. Isso pode ser explicado, por exemplo, a partir dos dados contidos em (SIEGEL et al., 2022). Os autores apontam que em 2022 estão previstos pelo menos 290 mil novos casos de câncer de mama para os Estados Unidos, o que representará 31% de todos os diagnósticos de câncer mais comuns em mulheres. Outro exemplo é o câncer colorretal, apontado como o terceiro mais comum em ambos os gêneros, com pelo menos 150 mil novos casos somente neste ano. Outro exemplo são os casos de displasia oral, lesões que afetam a estrutura das células e podem potencialmente se transformar em câncer (FONSECA-SILVA et al., 2016). O diagnóstico deste tipo de lesão é baseado no tamanho e severidade das alterações e apresenta diferentes níveis de intensidade (KUMAR; ABBAS; FAUSTO, 2005), no entanto, o diagnóstico em seus estágios iniciais é de suma importância para a designação do tratamento correto e impedir que a evolução da lesão resulte na formação de um câncer (KUMAR et al., 2008). Para 2022 Siegel et al. (2022) preveem pelo menos 54 mil novos casos de câncer nesta região, um dos 10 tipos mais comuns em pessoas do sexo masculino.

1.2 Objetivos

O objetivo proposto neste trabalho é uma investigação envolvendo *ensembles* de descritores e classificadores para analisar imagens histológicas H&E representativas do câncer de mama, colorretal, displasia oral e tecido hepático. Desta forma, os objetivos específicos para isto são:

1. Investigar o poder discriminativo de *deep features* obtidas de diferentes arquite-

turas com o uso de um *ensemble* de classificadores;

2. Verificar os desempenhos de descritores fractais multiescalas e multidimensionais (técnicas *handcrafted*), obtidos de imagens originais e representações IAx, em um *ensemble* de classificadores;

3. Realizar *ensembles* de descritores e comparar os resultados em relação aos obtidos a partir dos itens 1 e 2;

4. Definir as melhores soluções para cada conjunto sob investigação e identificar possíveis padrões de associações.

Capítulo 2

Fundamentação Teórica

Neste capitulo são apresentados os principais conceitos para a compreensão e elaboração do trabalho. Nas seções a seguir estão detalhes das técnicas aplicadas para obter os descritores *handcrafted*, dos métodos *deep-learning* utilizados para definir os descritores *deep learned*, das técnicas de IAx, da abordagem para a redução de dimensionalidade, dos algoritmos que podem ser utilizados para compor um *ensemble* de classificadores e uma visão geral sobre a categoria de imagens histológicas com coloração H&E.

2.1 Descritores *Handcrafted*: Técnicas Fractais

A geometria fractal é um conceito definido por (MANDELBROT; PIGNONI, 1983) para descrever estruturas complexas por sua irregularidade e/ou fragmentação, como são comumente encontrados em elementos naturais. Uma de suas principais vantagens é a independência da escala na análise, consequência da auto-similaridade das estruturas geométricas desta classe. Isto significa que fractais apresentam as mesmas características quando analisados estatisticamente em diferentes escalas de observação. Exemplos comuns encontrados na natureza de formas que respeitam estas regras são nuvens e litorais (FORTIN et al., 1992).

Computacionalmente, os descritores fractais podem ser mensurados a partir de diferentes técnicas, tais como o *box-counting* (NIKOLAIDIS; NIKOLAIDIS; TSOUROS, 2011) e *gliding-box* (IVANOVICI; RICHARD, 2010). Estes cálculos geralmente empregam a estratégia de contagem de caixas. A ideia por trás desta contagem está em utilizar caixas para delimitar subdivisões ao longo do espaço da imagem e verificar como é o conteúdo de cada um destes recortes. Em seguida, de acordo com suas composições, são contabilizados os tipos de caixas presentes na imagem. Variações nos tamanhos das caixas utilizadas para as divisões proporcionam as características multiescalas destas propriedades.

Um algoritmo comumente utilizado para a contagem de caixas é o do *Gliding-Box* (IVANOVICI; RICHARD, 2010), em que uma caixa desliza pela imagem verificando se os pixels de cada região estão contidos no espaço da caixa. Esta verificação é dependente do cálculo da distância entre o centro da caixa e os pixels da região, com isso é possível obter formas distintas de se interpretar o valor e posição do pixel. Ao considerar um pixel na linha x, coluna y e intensidade de cor z, este pode ser interpretado como um ponto no espaço com as coordenadas (x,y,z) num espaço tridimensional. No entanto, imagens coloridas possuem diferentes canais de cores com intensidades distintas. A partir disso, Ivanovici e Richard (2010) apresentou uma abordagem multidimensional, em que um pixel é definido por sua posição espacial e suas intensidades nos canais de cores. Assim, um pixel em uma imagem que utiliza o padrão de cores RGB é representado por (x,y,r,g,b), um vetor n-dimensional no qual a cor da imagem contribui na contabilização da informação. A partir desses conceitos, é possível mensurar o conteúdo de imagens via dimensão fractal e a lacunaridade, possibilitando conhecer quanto do espaço está preenchido e quão organizado é o preenchimento do espaço, respectivamente (IVANOVICI; RICHARD, 2010).

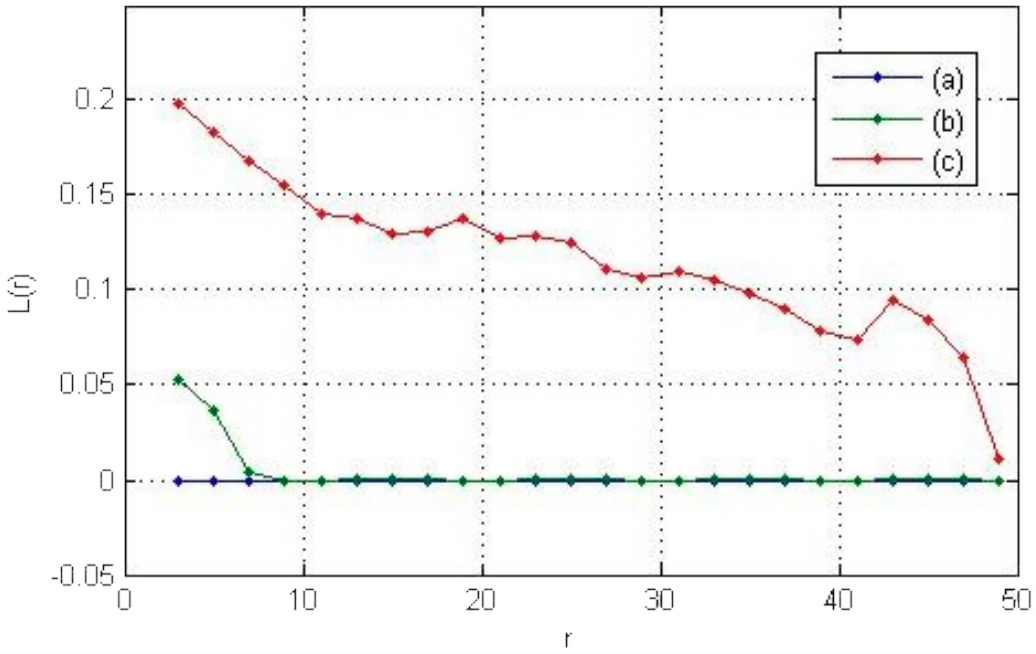
A dimensão fractal é baseada na ideia de expansão dos conceitos Euclidianos, onde as medidas estão contidas em um espaço n-dimensional, sendo n um número inteiro

maior que zero. Conforme o proposto por (MANDELBROT; PIGNONI, 1983), a dimensão fractal cria subdivisões no espaço observado a fim de generalizar as medidas ao incluir valores irracionais, resultando em um atributo livre de escala de observação. Nesse caso, o valor depende somente do número de subdivisões aplicadas ao espaço para mensurar um elemento. Uma das formas de se compreender este valor é interpretá-lo como a medida da rugosidade da estrutura analisada (PENTLAND, 1984)

No que lhe concerne, a lacunaridade é uma medida complementar à dimensão fractal, pois quantifica a distribuição e a organização de pixels contidos em uma imagem (MANDELBROT; PIGNONI, 1983). Os valores de lacunaridade permitem conhecer como os padrões estão organizados em diferentes escalas de observação. Para melhor ilustrar o que esta propriedade representa, na Figura 2.1 estão exemplos das curvas de lacunaridade das imagens contidas na Figura 2.2. Assim, para imagens com conteúdo homogêneo, como o que está na Figura 2.2a, o valor da lacunaridade é 0 (2.1a), visto que a distribuição do valor dos pixels é uniforme. Para imagens com padrões, como a ilustrada na Figura 2.2b, é possível observar que os valores da lacunaridade são mais elevados no início da curva (2.1b), visto que as caixas ainda não cobrem totalmente o padrão presente. Porém, assim que r se aproxima de uma medida que consegue englobar um padrão, o valor de lacunaridade diminui consideravelmente. Por fim, dada a imagem indicada na Figura 2.2c, com distribuição aleatória de valores, nota-se um aumento no valor da lacunaridade (2.1c), com alterações em função da variação de r, resultado da ausência de padrões no preenchimento do espaço.

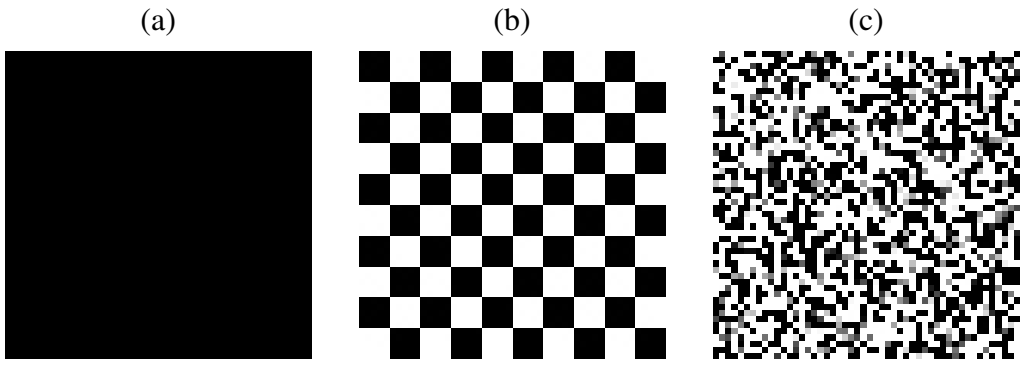
Outro descritor pertencente a esta categoria é a percolação. Esta propriedade foi definida por Broadbent e Hammersley (1957) e compreendida tradicionalmente como o movimento de filtragem de um líquido por materiais porosos. Esta teoria é útil enquanto explorada no contexto de conjuntos desordenados, dado que este fenômeno ocorre de forma aleatória. Com isto, o preenchimento de espaços analisados também torna-se aleatório e resulta na construção de aglomerados (ou *clusters*) (STRELNI-

Figura 2.1: Curvas de lacunaridade para as imagens na Figura 2.2



Fonte: Elaborado pelo Autor

Figura 2.2: Imagens geradas computacionalmente (50x50 pixels) a fim de exemplificar diferentes valores de lacunaridade: (a) Imagem uniforme com a cor preto; (b) Tabuleiro de xadrez em preto e branco; (c) Ruído gerado aleatoriamente em níveis de cinza.



Fonte: Elaborado pelo Autor

KER; HAVLIN; BUNDE, 2009). A percolação então pode ser observada a partir destes *clusters*, de forma que ocorre percolação no sistema quando um *cluster* conecta duas de suas extremidades. Este descritor pode ser categorizado como parte das técnicas fractais uma vez que a topologia destas estruturas apresentam características

relacionas às outras propriedades deste grupo. Além disto, é possível expandir estes conceitos para o contexto de imagens ao considerar a conectividade da vizinhança de um pixel para obter os agrupamentos (YAMAGUCHI; HASHIMOTO, 2010).

2.2 Redes Neurais Convolucionais (CNNs)

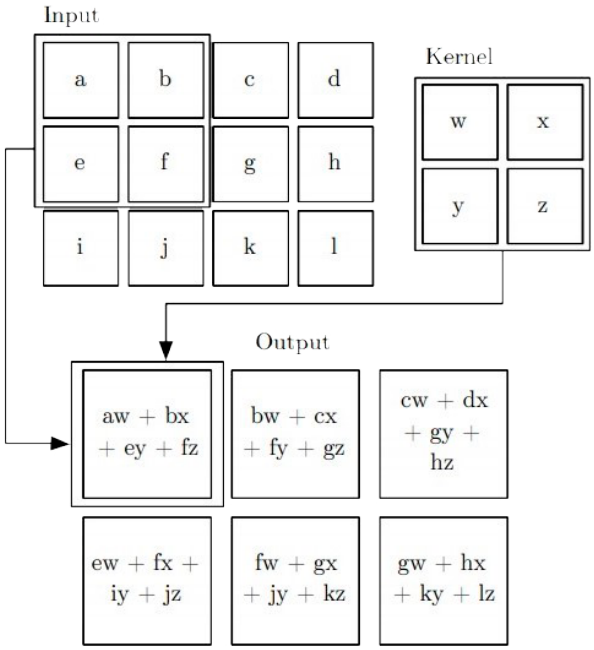
Redes Neurais Convolucionais (do inglês *Convolutional Neural Networks*, ou CNNs) são redes neurais que possuem camadas nas quais o processo de convolução é aplicado ao sinal de entrada. Esta aplicação é interessante para dados topológicos como imagens, visto que a posição dos valores na composição do sinal está correlacionada ao seu conteúdo e significado. Assim, o processo de convolução permite uma forma de pré-processamento dos dados, analisando todo o conjunto em busca de pequenos padrões e como eles se relacionam (LECUN; BENGIO et al., 1995; LECUN et al., 1998).

2.2.1 Convolução

Para entender a composição de uma camada convolucional, torna-se necessário compreender a operação de convolução. Esta operação se dá por meio da aplicação de uma máscara com pesos, ou *kernel*, que percorre todo sinal de interesse. Para cada posição que a máscara é aplicada, a soma do produto dos pesos pelos valores da região é calculada (GONZALEZ; WOODS, 2009). Um exemplo ilustrativo deste processo está na Figura 2.3. Assim, a convolução pode ser interpretada como um mapeamento de como cada região interage com a máscara.

Neste contexto, um *perceptron* aplicado a uma imagem, região a região, até percorrer toda a sua área, teria o mesmo efeito que a aplicação de uma convolução. Como consequência, é possível implementar algumas das práticas das redes neurais convencionais, como ajustes no *kernel* realizados pelo *backtracking* para refinar o modelo.

Figura 2.3: Ilustração de convolução em duas dimensões.



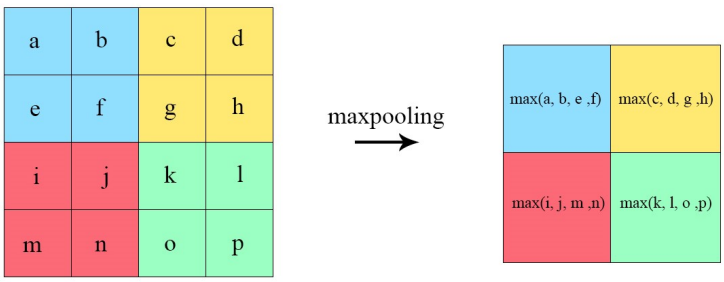
Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

Além disso, o fato da busca por algum padrão percorrendo toda a imagem implica que a posição não influencia o processamento. Isto porque esta varredura elimina a necessidade da repetição de pesos em diferentes pontos, em que é possível ou esperado encontrar o padrão (LECUN; BENGIO et al., 1995; LECUN et al., 1998). Portanto, uma camada convolucional é composta por diversos *kernels* de mesmo tamanho, resultantes em diversos mapeamentos por convolução da imagem. Estes alimentam as camadas seguintes até serem convertidas em um vetor unidimensional de valores. Este vetor contém as entradas para camadas nomeadas *fully-connected*. Por conta disso, as camadas convolucionais representam um processo de extração de descritores, diminuindo o pré-processamento necessário para a classificação dos dados sob investigação.

2.2.2 *Pooling*

Geralmente, a composição de uma camada convolucional possui três estágios: inicial, em que são aplicadas as convoluções; ativação, em que uma função de ativação agrupa os resultados em um único mapa de características; e, o *pooling*. Existem formas diferentes de se aplicar o *pooling*, mas que, em geral, têm o mesmo objetivo, o de aplicar funções para sintetizar as regiões do mapeamento e reduzir o volume de dados sob análise (GOODFELLOW; BENGIO; COURVILLE, 2016). O *max-pooling*, ilustrado na Figura 2.4, por exemplo, consiste em extrair o maior valor de uma vizinhança para compor o resultado. Outras abordagens utilizadas são calcular a média da vizinhança ou selecionar o valor mais próximo ao pixel central da vizinhança selecionada. Independente da metodologia aplicada, o resultado não deve perder muita da informação extraída e deve destacar as características mais marcantes daquela região, além de seu tamanho menor permitir que o processamento nas próximas camadas seja otimizado.

Figura 2.4: Exemplo de *maxpooling* aplicado em regiões 2x2 com passo de tamanho 2.



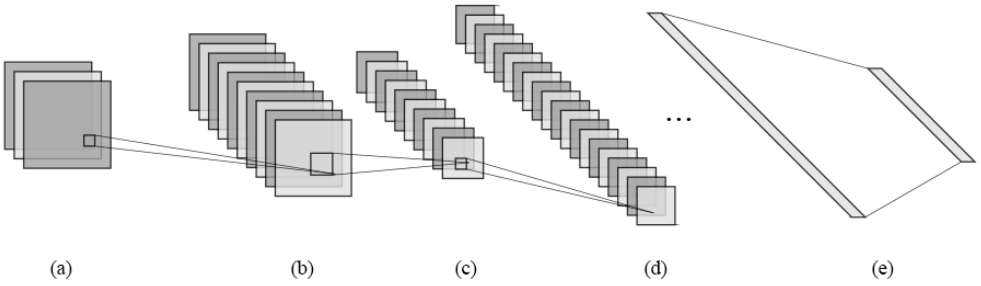
Fonte: Elaborado pelo Autor

2.2.3 Arquitetura de uma CNN

Com os métodos mais comuns na composição da arquitetura de CNNs descritos nas subseções 2.2.1 e 2.2.2, é possível delinear um padrão básico de arquitetura. De forma geral, as camadas aplicam operações topológicas (convolução ou *pooling*),

sucessivamente, até que se obtenha um número reduzido de descritores, e a rede se comporte como uma rede neural convencional, com as camadas *fully-connected*. Um exemplo simples que segue este padrão está ilustrado na Figura 2.5. Outro detalhe é como as operações impactam os dados, visto que a convolução resulta em múltiplos mapeamentos e o *pooling* reduz o volume de descritores.

Figura 2.5: Exemplo de arquitetura de uma CNN, em que: (a) imagem de entrada; (b) resultados da aplicação das máscaras de convolução em (a); (c) filtros obtidos em (b) após o *pooling*; (d) resultado da aplicação de uma nova convolução em (c); e, (e) camadas *fully-connected* obtidas após múltiplas operações de convolução e *pooling*.



Fonte: Elaborado pelo Autor com ferramenta online (LENAIL, 2019)

A partir destas e outras operações é possível construir redes com características que podem proporcionar desempenhos distintos, dependendo dos objetivos e contextos sob investigação. Alguns exemplos de modelos são VGG Net(SIMONYAN; ZISSERMAN, 2014), ResNet(HE et al., 2016), DenseNet (HUANG et al., 2017), Inception Net (SZEGEDY et al., 2015) e EfficientNet (TAN; LE, 2019).

VGG Net

As redes do tipo VGG tiveram sua proposta inicial por Simonyan e Zisserman (2014) e os autores buscaram aprimorar a precisão da classificação por meio do aumento de sua profundidade. Essa arquitetura explora pequenos filtros de convolução com o intuito de buscar padrões em múltiplas escalas. Redes que seguem esta arquitetura são divididas em cinco blocos de camadas convolucionais, todos conectados sequencialmente, além de uma camada de *max-pooling* na parte final. As operações de

convolução são aplicadas com máscaras de tamanho 3x3 e com um volume crescente conforme a profundida do bloco. Ao fim do bloco convolucional estão três camadas *fully-connected* com 4096, 4096 e 1000 neurônios, respectivamente, seguidas de uma função de ativação *softmax*.

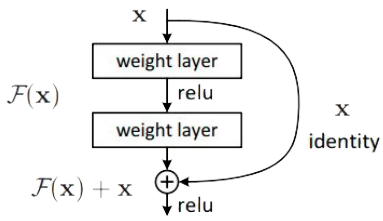
Esta proposta permitiu obter redes com até 19 camadas e que conseguem fornecer valores de acurácias relevantes. No entanto, esta abordagem multiescala também impactou o tempo para o treinamento do modelo (SIMONYAN; ZISSERMAN, 2014). Assim, é necessário mais treinamento para obter uma precisão refinada.

ResNet

Um dos desafios para redes muito profundas está em sua etapa de treinamento, pois o grande volume de operações resulta na explosão/desaparecimento dos gradientes de aprendizado. Para solucionar este problema, o modelo ResNet foi proposto explorando o princípio da passagem residual de valores entre camadas (HE et al., 2016). Esta arquitetura facilita o aprendizado, pois, enquanto uma rede não residual visa otimizar uma função $F(x)$ para de classificar uma entidade x, a passagem residual faz com que a função otimizada passe a ser $F(x)+x$. Esta pequena variação na função permite que a otimização dependa da parte residual, mesmo que $F(x)$ não seja eficiente. Para colocar este conceito em prática, a arquitetura ResNet recorre a blocos residuais, como o ilustrado na Figura 2.6.

Os blocos residuais são compostos por ao menos duas camadas com pesos que, além de estarem conectadas consecutivamente, possuem um atalho que transfere os valores do início ao fim do bloco, adicionando-os ao valor calculado pelas camadas internas. Por meio desta estratégia, o modelo ResNet forneceu resultados compatíveis ou superiores em relação aos modelos gerados a partir de outras arquiteturas, mesmo com redes de profundidades entre 34 e 152 camadas(HE et al., 2016).

Figura 2.6: Ilustração de um bloco residual da ResNet.



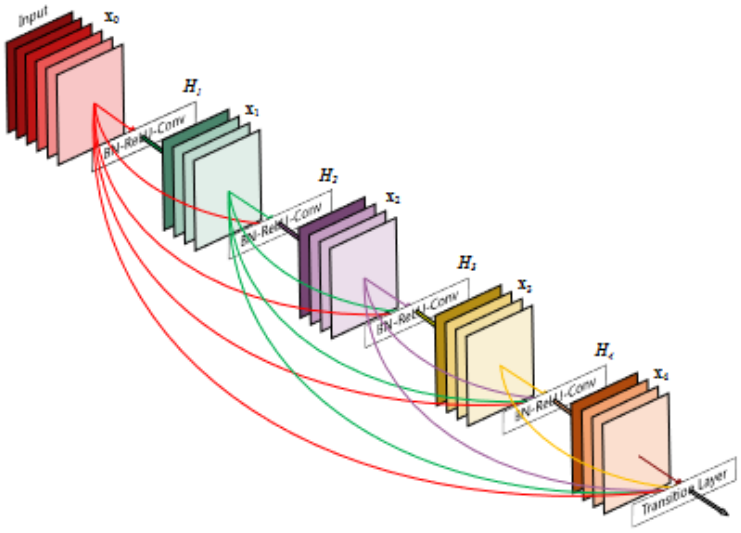
Fonte: (HE et al., 2016)

DenseNet

Apesar de eficiente, a passagem residual de valores não é a única estratégia empregada para incrementar a profundidade das redes sem haver degradação na informação passada entre camadas. Assim, o modelo DenseNet, que recebe esse nome por suas camadas serem densamente conectadas, foi proposto a partir do princípio da passagem de valores entre camadas, mas distintamente ao observado na ResNet (HUANG et al., 2017). Na DenseNet, a transmissão de valores é realizada para todas as camadas subsequentes. Logo, essa arquitetura indicou melhor distribuição e reaproveitamento da informação a ser processada, fato que permitiu uma redução no número de parâmetros treináveis e sem gerar impactos negativos no processo de classificação. Logo, redes de até 250 camadas, com resultados relevantes em relação aos demais modelos, foram observadas na literatura (HUANG et al., 2017).

A construção de redes neurais do tipo DenseNet é realizada por meio de blocos densos, ilustrados na Figura 2.7. Como já descrito, nesses blocos cada camada fornece seus resultados para as camadas seguintes. Nesta passagem, os mapeamentos são concatenados aos da camada subsequente e, desta forma, cada camada recebe cada vez mais valores. Como consequência, é necessário a aplicação de camadas de *pooling* entre os blocos densos, assim é possível reduzir a dimensão dos mapeamentos propagados e eliminar a presença redundâncias na informação.

Figura 2.7: Ilustração de um bloco denso da arquitetura DenseNet.



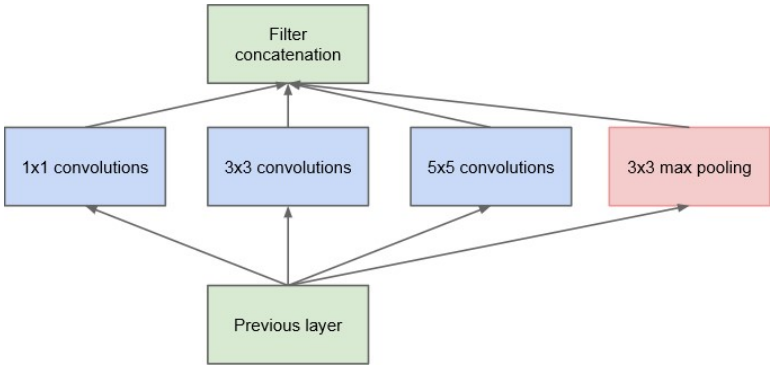
Fonte: (HUANG et al., 2017)

Inception Net

Enquanto algumas arquiteturas buscam aumentar a profundidade das redes, existem outras que investigam diferentes formas de expansão do processamento para melhorar seus resultados. Como redes muito profundas costumam resultar em *overfitting*, incrementar a largura das camadas de uma rede foi outro princípio explorado (SZEGEDY et al., 2015). Este incremento é resultado da aplicação de filtros de diferentes tamanhos na mesma camada, por meio de blocos de *Inception*. Em um bloco simples (*naive*), ilustrado em 2.8, cada um dos filtros deve contribuir com informação a partir da observação em diferentes escalas. Em sequência, as análises são concatenadas e alimentam a próxima camada.

Contudo, mesmo filtros pequenos, com dimensão de $5x5$, acabam aumentando o custo computacional da arquitetura, principalmente pelo aumento do volume de dados obtidos após muitas convoluções. Para solucionar este problema, uma variação no bloco *Inception* foi proposto (SZEGEDY et al., 2015), a fim de reduzir a saída obtida ao aplicar a convolução com filtros $1x1$. Esta variação é ilustrada na Figura 2.9. O uso desta estratégia permitiu construir uma rede com 22 camadas e conseguiu

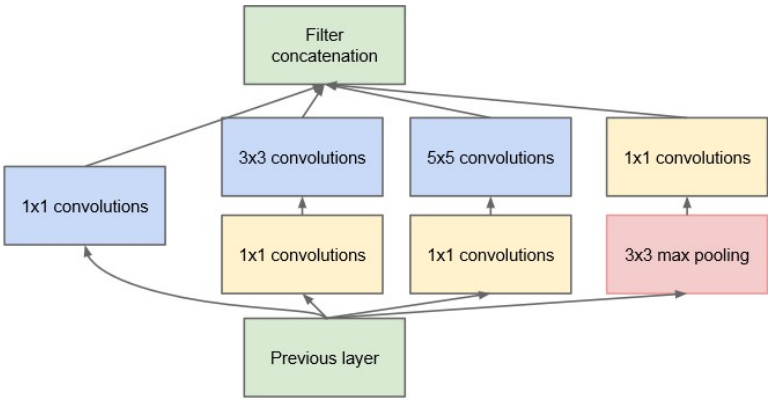
Figura 2.8: Ilustração de um bloco Inception *Naive*.



Fonte: (SZEGEDY et al., 2015)

fornecer taxas relevantes em diferentes tipos de imagens. Além disso, este modelo foi aprimorado para usar transferência residual entre as camadas e melhorar ainda mais o seu desempenho (SZEGEDY et al., 2017).

Figura 2.9: Ilustração de um bloco Inception com redução de dimensão.



Fonte: (SZEGEDY et al., 2015)

EfficientNet

Além do incremento na profundidade (d) e largura(w) das redes, existem outras técnicas que buscam incrementar a resolução (r) das imagens de entrada para aprimorar o processamento. No entanto, Tan e Le (2019) identificaram que existe um limite na melhora obtida ao escalar as arquiteturas em apenas um destes três fatores. Ainda, neste estudo foi proposto um fator ϕ para ajustar proporcionalmente cada um destes

parâmetros, segundo a regra descrita na Equação 2.1, em que d, w e r são os valores que multiplicam, respectivamente, as dimensões originais da profundidade, largura e resolução da arquitetura. A partir destas observações, os autores propuseram uma arquitetura, denominada EfficientNet, para realizar eficientemente as modificações de escala, conforme as proporções encontradas. Em sua proposta os autores aplicaram sete escalas diferentes à arquitetura de base, todas resultaram em redes com expressivamente menos parâmetros, desta forma seu treinamento e execução se tornam menos custosos. Além disto, todas as escalas forneceram resultados competitivos em relação a outras arquiteturas amplamente utilizadas na literatura.

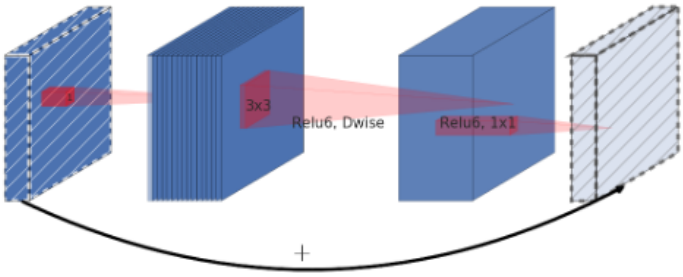
$$\begin{cases} d = \alpha^{\phi} \\ w = \beta^{\phi} \\ r = \gamma^{\phi} \\ \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{cases} \tag{2.1}$$

Para a arquitetura de base, os autores utilizaram o bloco residual invertido, conforme ilustrado na Figura 2.10. Esse tipo de bloco foi desenvolvido primeiramente para a construção da rede MobileNetV2 (SANDLER et al., 2018), arquitetura que obteve bons desempenhos em ambientes com poucos recursos computacionais. Estes blocos exploram a convolução em profundidade para filtrar os valores mapeados e remover as dependências lineares. Adicionalmente, esse tipo de bloco também conta com a estratégia de passagem residual de informação do início ao fim do bloco a fim de reduzir a degradação dos valores processados, assim como na arquitetura ResNet.

2.3 Transferência de Aprendizado

Transferência de aprendizado, ou *transfer learning*, é uma estratégia proposta inicialmente por (PRATT, 1993). A transferência de aprendizado visa aproveitar o co-

Figura 2.10: Ilustração de um bloco residual invertido.



Fonte: (SANDLER et al., 2018)

nhecimento obtido a partir de um modelo para acelerar o aprendizado em outro. Este aproveitamento pode melhorar os resultados obtidos no segundo modelo. Isso ocorre porque a transferência de conhecimento, mesmo sem refinamento, permite aumentar a precisão no início do treinamento por iniciar os pesos das camadas com referências não aleatórias. Além disso, o treinamento para o novo modelo tende a ser mais rápido e pode atingir resultados melhores do que os obtidos sem o auxílio de um conhecimento prévio (TORREY; SHAVLIK, 2010). Além disso, outro benefício da transferência de aprendizado está em permitir aplicações de modelos CNN em contextos com um número reduzido de amostras para o treinamento. No caso, torna-se possível obter um modelo treinado em um contexto que se tenha dados suficientes para uma boa generalização e, então, empregar o conhecimento adquirido no contexto de interesse. A transferência de aprendizado pode falhar caso o modelo tenha sido treinado em um contexto de imagens com pouca generalização ou o contexto final seja totalmente distinto do modelo alvo.

2.4 Inteligencia Artificial Explicável

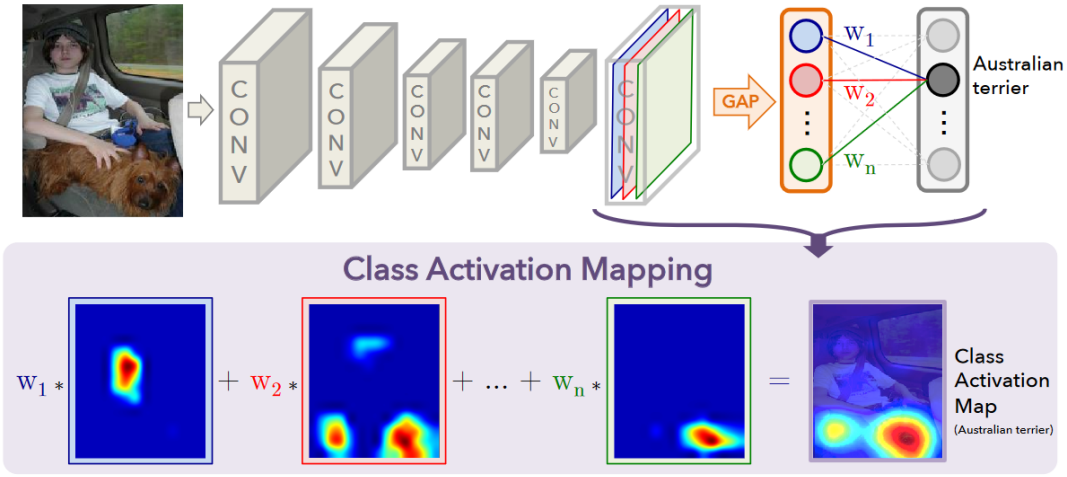
Redes neurais artificias conseguem obter uma precisão notável em diversos contextos. Por outro lado, o processo comumente utilizado na tomada de decisão e/ou classificação não é trivial, fato que torna sua compreensão uma tarefa altamente complexa, especialmente para entender como o resultado foi obtido (ALICIOGLU; SUN,

2022; HOLZINGER et al., 2022). Esta situação é um importante desafio no campo de reconhecimento de padrões via redes neurais artificiais, com destaque para definir se as tomadas de decisões foram realizadas pelos motivos corretos (SAMEK; MÜLLER, 2019). Para solucionar parte desta dificuldade, estudos relevantes foram direcionados para o campo de pesquisa nomeado como inteligencia artificial explicável (IAx) (ZHOU et al., 2016; RIBEIRO; SINGH; GUESTRIN, 2016; ALMARAZ-DAMIAN et al., 2020; ALICIOGLU; SUN, 2022; HOLZINGER et al., 2022). Neste, busca-se, por exemplo, elaborar estratégias de aprendizado de máquina para tornar o conhecimento compreensível aos especialistas e ainda garantir modelos mais confiáveis (GUNNING; AHA, 2019), aplicando técnicas fundamentadas em mapas de ativação de classe (CAMs) (ZHOU et al., 2016) e *Local Interpretable Model-Agnostic Explanations* (LIMEs) (RIBEIRO; SINGH; GUESTRIN, 2016).

2.4.1 Mapa de Ativação de Classe

O CAM foi elaborado para auxiliar a detecção do posicionamento de objetos classificados em imagens (ZHOU et al., 2016). Este mapeamento realiza uma regressão a partir da camada final de uma CNN e se baseia nos pesos das conexões com as camadas anteriores. A ideia é identificar quais características da imagem contribuíram efetivamente para a classificação. Assim, é possível atribuir pesos para a contribuição de cada pixel e criar um mapa de intensidades, ou calor, que indique as zonas mais relevantes para o resultado obtido. Este processo ocorre a partir da saída da CNN, considerando uma classe de interesse e os pesos W que conectam a camada final até a saída desta classe. Em seguida, os mapeamentos resultantes destes pesos são realizados conforme a saída da última camada convolucional da arquitetura. Por fim, as máscaras são combinadas e redimensionadas para corresponder ao tamanho da imagem de entrada e indicar as regiões de interesse correspondentes. Um exemplo deste processo está ilustrado na Figura 2.11.

Figura 2.11: Ilustração da regressão realizada para a saída de uma das classes de uma CNN que permite a obtenção de um CAM.



Fonte: (ZHOU et al., 2016)

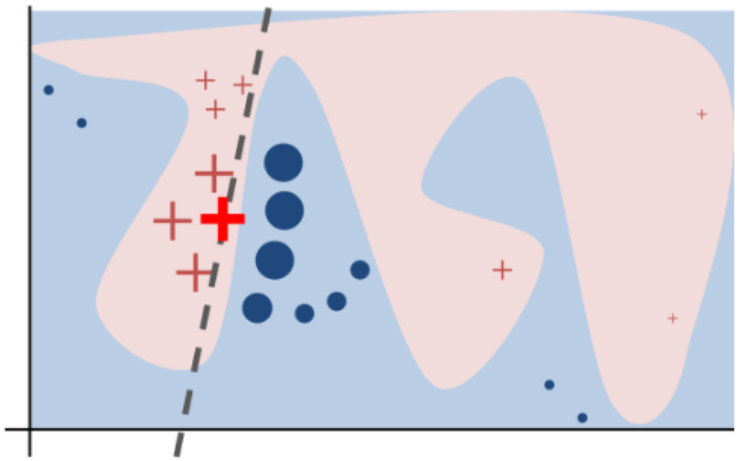
Adicionalmente, a proposta de CAM foi expandida para torná-la viável para diferentes arquiteturas e garantir um resultado mais preciso no mapeamento (SELVARAJU et al., 2017). A proposta foi nomeada grad-CAM e visa explorar os gradientes presentes nas conexões das camadas convolucionais ao invés da camada final da rede. Para tal, os autores generalizaram a regressão para ser aplicável à qualquer camada de uma CNN, consequentemente houve uma generalização do método CAM. Apesar da proposta ser aplicável em qualquer camada, os autores notaram que as camadas convolucionais mais profundas forneceram maior valor semântico em seus valores e conseguiram aprimorar os mapeamentos resultantes.

2.4.2 *Local Interpretable Model-Agnostic Explanations*

Outra estratégia que pode ser utilizada para explicar soluções obtidas via modelos CNNs é a proposta de Ribeiro, Singh e Guestrin (2016), conhecida como LIME. Esta técnica aplica uma busca local em relação a um dado resultado para tentar encontrar uma simplificação do que pode ter contribuído para o valor obtido. Os proponentes indicaram que esta técnica é independente de qualquer modelo.

Um exemplo da obtenção de uma explicação LIME está ilustrado na Figura 2.12, em que se busca uma separação linear que explique a classificação por um modelo. Neste exemplo, as cores de fundo indicam a percepção do modelo treinado em relação ao espaço e a cruz vermelha em destaque é a classificação para qual é buscada uma explicação. Para realização da busca local, são criados outros n exemplos considerando perturbações nos valores de entrada, ilustrados pelas demais cruzes e círculos. Com isto, pode ser realizado um levantamento de quais são os resultados locais mais relevantes para cada uma das classes, conforme a busca realizada. Na Figura 2.12, os pontos mais relevantes são aqueles mais próximos da classificação de interesse, com uma representação equivalente ao tamanho do sinal. Finalmente, a partir da observação da relevância de cada uma das perturbações, é possível encontrar uma explicação para os resultados. No caso da ilustração, a explicação é uma separação linear do espaço que justifica a distinção entre cruz e círculo. Este exemplo ilustra bem o aspecto local desta técnica, de forma que a divisão linear encontrada para a explicação é precisa para a região do ponto de interesse, mas globalmente se distancia do espaço treinado para o modelo original.

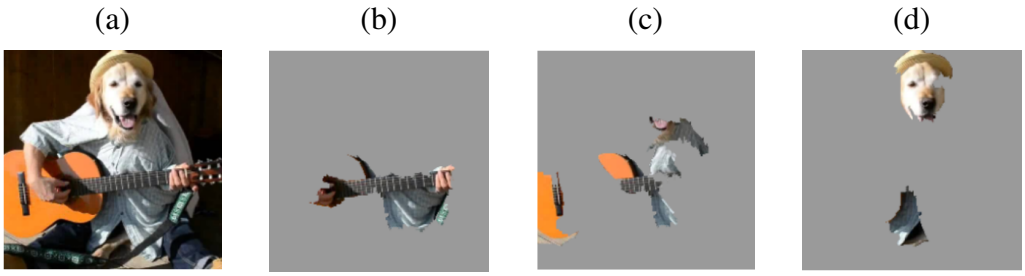
Figura 2.12: Ilustração da aplicação do LIME para explicar um modelo treinado na classificação de amostras um espaço bi-dimensional.



Fonte: (RIBEIRO; SINGH; GUESTRIN, 2016)

No contexto de imagens, a abordagem LIME tenta encontrar os super-pixels que contribuíram para o resultado. Os super-pixels são regiões da imagem resultantes da aplicação de uma técnica de segmentação. Desta forma, para ser possível investigar a contribuição destes segmentos, os exemplos resultantes das perturbações da imagem de entrada são obtidos a partir da remoção aleatória de alguns destes super-pixels. Assim, estes exemplos alimentam um modelo que será utilizado para ranquear quais regiões foram mais importantes e, então, definir os k super-pixels que explicam o resultado obtido. Na Figura 2.13 esta uma ilustração da aplicação desta técnica em uma imagem arbitrária e os resultados obtidos para três classificações prováveis em razão da CNN utilizada.

Figura 2.13: Explicações LIME para a classificação da CNN Inception-V3 de uma imagem arbitrária (a). Foram selecionados os $k = 10$ super-pixels que explicam três classes prováveis segundo o modelo: guitarra Elétrica (b); violão acústico (c); e Labrador (d).



Fonte: (RIBEIRO; SINGH; GUESTRIN, 2016)

2.5 *Ensemble* de Descritores

Existem diversas estratégias para quantificar padrões em imagens. Consequentemente, o uso de *ensemble* de descritores busca, de alguma forma, combinar múltiplos grupos de técnicas para compor um único vetor de características e, assim, aumentar a capacidade de discriminação dos conjuntos sob investigação. As abordagens *ensemble* podem ser divididas em técnicas de agrupamento e técnicas de fusão (YANG et al., 2003).

Os métodos de agrupamento são os mais comuns e consistem em concatenar os descritores em um único vetor, permitindo agregar informações sobre os diferentes aspectos dos dados para melhorar os resultados desejados. No entanto, os vetores resultantes desta estratégia podem apresentar alta dimensionalidade e redundâncias de descritores, implicando no uso de métodos para contornar estes problemas (DASIGI; MANN; PROTOPOPESCU, 2001). Por outro lado, os métodos de fusão aplicam operações entre os conjuntos de descritores e obtém novos valores. Os valores resultantes são utilizados para compor um novo vetor de características. As operações aplicadas podem ser executadas para que as informações obtidas sejam mais relevantes, o que permite eliminar o processo de seleção de descritores. Todavia, as operações interferem na capacidade semântica da informação utilizada para descrever a entrada. Como consequência, um *ensemble* fundamentado em fusão é altamente sensível às operações escolhidas e ao contexto de aplicação (MUHAMMAD; WANG; HADID, 2018).

2.6 Seleção de descritores

O processo de classificação requer um conjunto de descritores que são analisados e, consequentemente, utilizados nas tomadas de decisões. Porém, um problema frequente é a existência da maldição da dimensionalidade (WATANABE, 1985), quando o número de descritores é maior que o número de amostras, fato que pode tornar os grupos semelhantes para os algoritmos. Para evitar que este tipo de problema ocorra, a composição deve ser pelo menos 10 vezes maior que o número de descritores (JAIN; DUIN; MAO, 2000). Por outro lado, Hua et al. (2004) descrevem que, para n elementos, é desejável manter no máximo $n-1$ descritores não correlacionadas. Para descritores correlacionados, o limite indicado pelos autores é próximo a $\sqrt{n}$.

Aumentar o número de amostras pode ser difícil e/ou inviável em determinados contextos. Uma alternativa para solucionar este problema é implementar métodos capazes de identificar os descritores mais relevantes para o processo de classificação e,

consequentemente, reduzir a dimensão do vetor de características. Entre os métodos disponíveis na literatura, existem as categorias *wrapper*, embutido e filtro. Na categoria *wrapper*, os métodos realizam a seleção de descritores a partir de um algoritmo de treinamento para verificar a melhor combinação entre todos os descritores disponíveis, propriedade que outros métodos não conseguem fornecer. O principal problema deste tipo de abordagem é o custo computacional para analisar todas as combinações possíveis. Na categoria de métodos embutidos, as técnicas incorporam a seleção de descritores ao processo de classificação para realizar o aprendizado dos melhores descritores a cada execução. A principal desvantagem também está no custo computacional com o processo de identificação dos descritores. Por fim, as técnicas da categoria filtro são aplicadas independentemente da classificação. Assim, os descritores são selecionadas a partir de análises estatísticas e de correlação, capazes de fornecer resultados expressivos para definir a relevância da informação contida no vetor de características. Nesta categoria, um exemplo de técnica amplamente utilizada para analisar e identificar os melhores descritores é o *Relief* (LE et al., 2018).

2.6.1 *Relief*

O método *Relief* é derivado da proposta descrita por Kira e Rendell (1992), visando apresentar abordagens de seleção de descritores eficientes e que conseguem evitar o emprego de buscas com heurísticas. Especificamente, o algoritmo *Relief* apresenta complexidade linear e utiliza métodos estatísticos para determinar os melhores descritores. A ideia é que cada uma dos i descritores recebe um peso W, inicialmente zero, que tem como significado a sua relevância no conjunto. Este peso então é atualizado por meio de múltiplas iterações para comparar três amostras a cada etapa. Para isto, são selecionadas aleatoriamente três vetores como amostras. Estas amostras constituem um vetor de características qualquer denotado por x, um vetor que descreve um dado da classe A e outro da classe B, ambos com os valores mais próximos de x. Em

sequência, cada iteração atualiza os pesos a partir das distâncias dos valores de x em relação aos outros dois vetores, conforme descrito na Equação 2.2, sendo que *nearHit* representa o vetor selecionado de mesma classe que x e *nearMiss* indica o vetor selecionado de classe diferente.

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2. \tag{2.2}$$

Após o cálculo dos pesos, os valores são normalizados dividindo-os pelo número total de amostras. O resultado é um vetor que representa a relevância de cada um dos descritores. Por fim, os descritores com relevância maior podem ser selecionadas por meio de um limiar τ.

Uma das variações notáveis da técnica *Relief* foi proposta por (KONONENKO; ŠIMEC; ROBNIK-ŠIKONJA, 1997), denominada *ReliefF*, que visa solucionar alguns problemas encontrados no modelo original. Assim, os vetores *nearHit* e *nearMiss* são solucionados considerando uma média de k vetores vizinhos mais próximos da mesma classe ou de classes diferentes. Isso permite que o conjunto de treinamento seja reduzido e os descritores com ruído ou redundantes sejam filtrados. Adicionalmente, o método foi expandido para avaliações de conjuntos multi-classes e dados incompletos.

2.7 Métodos de Classificação

A classificação de padrões tem a finalidade de relacionar as informações extraídas de amostras sob avaliação com um conjunto de rótulos, sendo que vetores de características que apresentam descritores semelhantes devem ser vinculados ao mesmo rótulo (PEDRINI; SCHWARTZ, 2008). Assim, as estratégias utilizadas podem explorar diferentes heurísticas, tais como abordagens baseadas em árvores de decisão, funções, *Lazy Learning* e *Teorema de Bayes*, por exemplo.

O classificador *Decision-Tree* (DT) utiliza o modelo de árvore de decisão como

referência, em que cada ramificação representa uma verificação dos valores que descrevem a entrada e cada folha uma classificação. Sua construção se dá por meio de verificações recursivas em uma estrutura da árvore, de forma que cada rota determine um limiar otimizado para a classificação dos dados de entrada (QUINLAN, 1987).

O classificador *Random Forest* (RF) se baseia em uma combinação de diversas *Decision Trees* (BREIMAN, 2001), visando aproveitar o conhecimento divergente obtido por múltiplos classificadores diferentes. Este algoritmo usa da técnica de *bagging* de composição de *ensemble* de classificadores, em que múltiplas DT são construídas e treinadas a partir de diferentes amostras do conjunto de treinamento. Assim, os classificadores internos conseguem distinguir classes via diferentes características do conjunto de entrada. Isso permite que os dados sejam avaliados por meio de análises distintas. A classificação final é decidida a partir de votação simples, em que a classe obtida com mais frequência pelas classificações internas indica o resultado.

O classificador K-*Nearest-Neighbors* (KNN) é um método fundamentado em instância. Seu funcionamento considera cada vetor de características como um ponto em um espaço n-dimensional e os k vizinhos mais próximos determinam, por votação, a classe para o mapeamento. A heurística utilizada por este método pode garantir bons resultados, mas o processo de classificação é altamente sensível a ruído, situação que pode ocorrer em vetores de características com alta dimensionalidade (CUNNINGHAM; DELANY, 2020).

O classificador *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995) utiliza a abordagem de classificação estatística, por meio da aplicação de teorias de regressão. O treinamento deste algoritmo visa aplicar este tipo de operação para encontrar uma margem que separe adequadamente os elementos entre seus respectivos rótulos/grupos. A ideia é considerar os vetores de características como pontos em um hiperespaço multidimensional. Assim, o algoritmo visa definir um hiperplano para distinguir os pontos (amostras), explorando operações para otimizar o polinômio que define o hi-

perplano. Este processo ocorre até encontrar uma situação que define margens seguras de classificação.

O classificador *Naive* Bayes (NB) faz parte do conjunto de classificadores Bayesianos, pautados em cálculos estatísticos elaborados a partir do teorema de Bayes (PEDRINI; SCHWARTZ, 2008). Este teorema recorre à probabilidade condicional. Desta forma, dado dois indivíduos X e Y, é calculada a probabilidade de X fazer parte de um grupo A, dado que Y é parte de um grupo B, ou, em termos matemáticos $P(X \epsilon A | Y \epsilon B)$. Então, para a classificação dos métodos dessa família, é calculada a probabilidade de um vetor de características X pertencer a cada uma das classes, dado que outros vetores conhecidos fazem parte de suas respectivas classes. Por fim, a probabilidade mais alta determina o resultado. O *Naive* Bayes ou Bayesiano Ingênuo recebe este nome porque, para calcular estas probabilidades, considera os descritores como não correlacionados.

2.8 *Ensemble* de Classificadores

Um *ensemble* de classificadores é uma técnica de classificação que combina os resultados de diferentes classificadores a fim de obter uma classificação final mais confiável quando comparada com a fornecida por um único algoritmo de classificação (LIMA, 2005). A teoria que pode explicar este comportamento é a chamada sabedoria de multidão, em que, para uma pergunta complexa, a combinação de respostas de muitos indivíduos tende de ser mais precisa do que a de um único, mesmo que este seja um especialista. Um exemplo desta estratégia é o método de classificação *Random Forest*, que agrega a classificação de múltiplas *Decision-Trees* para sua decisão final (GERON, 2019). Logo, de um *ensemble* de classificadores é explorar os pontos fortes de cada algoritmo para que os possíveis pontos fracos individuais sejam reduzidos (KUNCHEVA, 2014). Existem diferentes formas para compor um *ensemble*, mas que podem ser observadas a partir de duas categorias principais, os *ensembles* homogêneos

e os heterogêneos.

2.8.1 *Ensembles* Homogêneos

Os *ensembles* homogêneos são aqueles compostos por múltiplas instâncias de classificadores com o mesmo método de classificação, como ocorre na abordagem *Random Forest*. Nestes casos, os classificadores devem diferir entre si a partir de variações nos métodos de treinamento. Com isso, os classificadores podem fornecer diferentes padrões com uma mesma entrada. Dentre os métodos mais populares de composição de um *ensemble* desta categoria, *Bootstrapping AGGregatING* (*Bagging*) e *Boosting* merecem destaque (KUNCHEVA, 2014).

Método *Bagging*

O *Bagging* foi proposto por (BREIMAN, 1996), inspirado pelo método de amostragem *Bootstrapping* para treinamento de classificadores. Neste método, o *ensemble* é construído a partir de classificadores treinados via diferentes conjuntos de treinamentos. Isto pode ser realizado com a aplicação do treinamento a partir de amostras selecionadas aleatoriamente do conjunto de dados de entrada, com reposição. A reposição indica que o mesmo dado pode aparecer mais de uma vez no conjunto selecionado. Esta forma de constituir o conjunto de treinamento é o que garante a diversidade necessária para o *ensemble*, visto que a reposição de indivíduos pode resultar em reforço na detecção de algum padrão específico. Outro benefício é que alguns indivíduos podem não fazer parte de qualquer conjunto de treinamento, o que pode contribuir na prevenção de *overfitting* na classificação (KUNCHEVA, 2014).

Usualmente, a classificação do *ensemble* que utilizam *Bagging* é obtida por meio de uma votação entre os classificadores. Desta forma, a média ou maioria simples de classificações para uma mesma classe determina o resultado. Contudo, embora seja geralmente empregada a maioria simples, é possível encontrar modelos que aplicam

algumas modificações na votação.

Método *Boosting*

Descrito por Schapire (1990) e aperfeiçoado por Freund, Schapire et al. (1996), os métodos de *Boosting* buscam combinar múltiplos classificadores "fracos" em um único mais preciso. Esta combinação ocorre ao treiná-los sequencialmente, de forma que cada um supere o seu antecessor (GERON, 2019). Assim, somente o primeiro classificador é treinado por um conjunto de amostras selecionadas via uma distribuição probabilística uniforme. Os demais são treinados por meio de conjunto de treinamento com amostras que podem maximar o erro. Com isso, os dados que foram classificados incorretamente pelos classificadores têm uma probabilidade maior de serem selecionados (LIMA, 2005).

No método *Boosting*, a classificação final é obtida por meio do cálculo da média ponderada, ou votação com pesos, considerando o resultado de cada um dos classificadores. Os pesos atribuídos aos votos são diretamente relacionados à precisão do classificador. Esta prática garante que classificadores mais precisos tenham maior impacto na decisão final.

2.8.2 *Ensembles* Heterogêneos

A categoria de *ensembles* heterogêneos considera um conjunto de classificadores com heurísticas distintas de classificação. Este tipo de *ensemble* evita os cuidados na fase de treinamento, como ocorre na estratégia de *ensemble* homogêneo, visto exploração a combinação de classificações individuais, tais como *voting* e *stacking*.

Voting

As abordagens de *voting* são aquelas que o *ensemble* obtém a sua classificação final por meio do sistema de votos, em que cada algoritmo indica uma decisão de rótulo

para uma amostra. No fim, a classe com mais votos (maioria simples) é definida como a classificação final. Em complemento, a classificação final também pode ser definida a partir de um processo de votação ponderada, votação com predição de probabilidade, entre outras (KUNCHEVA, 2014). Mesmo assim, independente dos sistemas de votação, os *ensembles* que utilizam este sistema podem não conseguir aproveitar tão bem as vantagens individuais de cada classificador. Além disso, a seleção de cada classificador deve ser criteriosa, uma vez que a presença de métodos com desempenhos globais ruins impactam negativamente a classificação final (LIMA, 2005).

Stacking

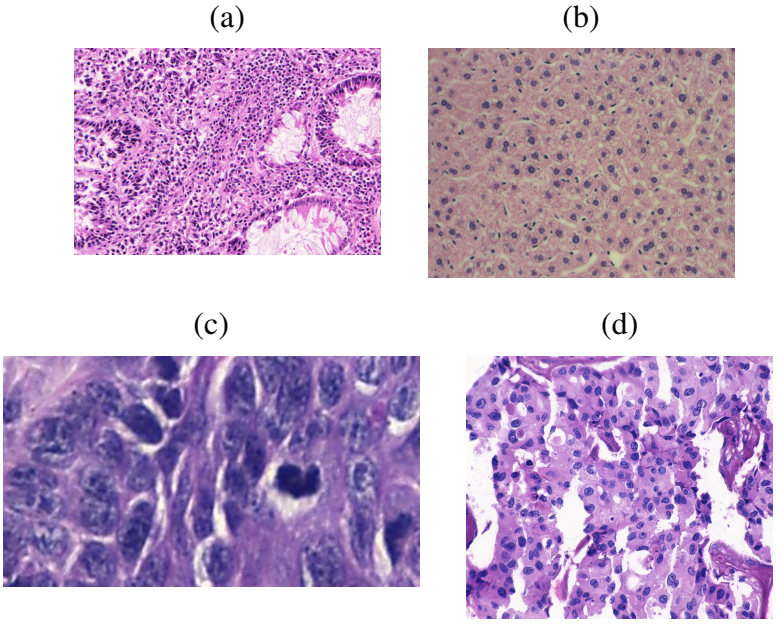
Para resolver os problemas gerados pela simplicidade do sistema de votos, a estratégia *stacking* define uma classificação final considerando que alguns classificadores podem ser mais confiáveis (KUNCHEVA, 2014). Esta abordagem, proposta por Wolpert (1992), consiste em aplicar um segundo classificador, ou meta classificador, para agregar as classificações do conjunto de base. Dessa forma, este tipo de *ensemble* pode ser dividido em classificadores de níveis 0 e 1, sendo eles os classificadores que compõem a base e o classificador responsável por combinar as classificações, respectivamente. Esta construção permite bons resultados, pois no nível 1 são aplicadas técnicas de aprendizado de máquina para tentar maximizar o aproveitamento da contribuição que cada classificador no nível 0 pode fornecer.

2.9 Imagens histológicas e Coloração H&E

Histologia é a área da biologia que permite estudar a estrutura microscópica de tecidos biológicos e, no que lhe concerne, a histopatologia é o estudo de amostras de tecidos biológicos doentes (STEDMAN, 1920), a fim de fornecer, por exemplo, diagnósticos mais precisos (ROSAI, 2007). Para tanto, as amostras são preparadas e tingidas com corantes, tais como hematoxilina e eosina (H&E), para evidenciar os basófilos

e eosinófilos dos tecidos via microscopia ótica (CARLETON; DRURY; WALLINGTON, 1980; TITFORD, 2005). O primeiro corante garante que o núcleo das células tome uma cor azulada, enquanto o segundo torna o citoplasma avermelhado conforme o ilustrado pelas Imagens 2.14. Além disso, as outras estruturas da célula acabam obtendo uma coloração derivada da mistura das duas substâncias (CHAN, 2014). A combinação destas cores é interessante, pois destaca as estruturas celulares de maior interesse, além de permitir a visualização da disposição geral das células no tecido. Estas condições contribuem para análises de regiões de interesse, seja por especialistas ou por sistemas automatizados direcionados para a classificação e o reconhecimento de padrões.

Figura 2.14: Imagens de lâminas histológicas coloridas com H&E de diferentes contextos, sendo eles respectivamente extraídos do (a) tecido colorretal(SIRINUKUNWATTANA et al., 2017), (b) figado (AGEMAP, 2020), (c) boca (SILVA et al., 2022) e (b) mamas (GELASCA et al., 2008).



Capítulo 3

Trabalhos Relacionados

Neste capítulo são apresentados os trabalhos relacionados com as abordagens que fazem o uso de métodos de *ensemble* de descritores e de classificadores, envolvendo diferentes contextos de imagens. Portanto, nas seções 3.1 e 3.2 estão descritos os principais trabalhos que fundamentam nossa proposta.

3.1 *Ensemble* de Descritores

Nanni, Lumini e Ghidoni (2018) aplicaram técnicas de *ensemble* de descritores, *ensemble* de classificadores e transferência de conhecimento no contexto de imagens histológicas de doença de pele. O objetivo foi determinar o melhor conjunto de descritores para as classes representadas pelas imagens sob investigação. O estudo definiu combinações de descritores obtidos tanto a partir das últimas camadas de seis modelos diferentes de CNN, quanto com as obtidas por meio de análises de profissionais da área. Os autores consideraram arquiteturas CNN com duas estratégias: aplicar o treinamento sobre um conjunto contendo somente as imagens de doença de pele; e, expandir o conjunto de treinamento com outros tipos de imagens histológicas. Os autores observaram que o *ensemble* dos descritores resultou em classificações com acurácias mais altas do que as utilizando conjuntos individuais. Além disso, a combinação de

diversas redes convolucionais obteve um desempenho melhor que a aplicação de apenas uma única rede. Os autores concluíram que os métodos aplicados podem produzir bons resultados, uma vez que foi obtida uma acurácia de aproximadamente 85% nas classificações.

Ainda no contexto de doenças de pele, Hagerty et al. (2019) realizaram experimentos que combinaram descritores *handcrafted*, extraídos de imagens de melanoma, e descritores *deep learned* da penúltima camada de uma ResNet-50. Para a classificação, foi empregado um cálculo de regressão logística nos vetores de características, resultantes do *ensemble*. Os autores identificaram que, enquanto combinados, os resultados atingiam uma precisão de 90%. Este desempenho foi melhor que o fornecido por vetores de características antes do *ensemble*. Por fim, os autores concluíram que, apesar da existência de diferentes estratégias para obtenção de descritores, o uso de *ensemble* destes contribuiu para o incremento da precisão da classificação das imagens sob investigação.

Wang et al. (2019b) compuseram vetores de características por meio de diferentes técnicas *deep learned* e *handcrafted* (morfologia e densidade das imagens). Os autores consideraram classificações realizadas por meio dos algoritmos *SVM* e *ELM*, utilizando os descritores individualmente e com todas as combinações possíveis de *ensemble*. No contexto de aplicação, os *ensembles* forneceram resultados melhores que os demais testes. Além disso, os autores concluíram que a seleção de descritores foi fundamental para garantir resultados relevantes.

Ainda, Almaraz-Damian et al. (2020) investigaram o uso de *ensemble* de descritores *handcrafted* e *deep learned* para classificar doenças de pele. Aliado ao *ensemble*, os autores aplicaram diferentes métodos de seleção de descritores a fim de reduzir a dimensionalidade de cada conjunto de entrada. Os resultados obtidos apresentaram acurácia de 92% nos melhores casos. Os autores observaram que este valor é compatível com outras propostas disponíveis na literatura.

O *ensemble* de descritores *deep learned* também foi aplicado na classificação de doenças de ouvido por Zafer (2020). O método proposto pelos autores fez a extração de descritores a partir de 7 arquiteturas diferentes de CNN, sendo elas AlexNet, VGG-16, VGG-19, GoogleNet, ResNet-18, ResNet-50 e ResNet-101, todas com refinamento do modelo a partir da transferência de conhecimento. Para a classificação dos descritores obtidos, foram empregados quatro classificadores diferentes: rede neural, KNN, Decision Tree e SVM. Os autores conduziram os experimentos para comparar os resultados obtidos com a classificação direta pelas CNN, dos vetores extraídos individualmente e o *ensemble* dos vetores, sendo que o *ensemble* obteve uma acurácia superior aos outros. Além disto, o método proposto permitiu obter acurácia de 99.47% ao classificar o *ensemble* com o SVM, superando outros métodos da literatura aplicados ao mesmo contexto.

No contexto de imagens de ressonância magnética, Kang, Ullah e Gwak (2021) propuseram um método de *ensemble* de descritores para identificação de tumores cerebrais. O experimento realizado contou com diferentes combinações *ensemble* com três descritores mais relevantes dentre cada uma de 13 arquiteturas de CNN com transferência de aprendizado, ResNet-50, ResNet-101, DenseNet-121, DenseNet-169, VGG-16, VGG-19, AlexNet, Inception-V3, ResNeXt-50, ResNeXt-101, ShuffleNet V2, MobileNet V2 e MnasNet. Os resultados obtidos mostraram que algumas das arquiteturas favoreciam determinadas bases de imagens e que as classificações via SVM resultaram nos melhores desempenhos quando comparados aos conquistados com outros classificadores.

Também em busca de explorar um *ensemble* de *deep features* com descritores *handcrafted*, Roberto et al. (2021) exploraram esta combinação para fundamentar uma arquitetura de CNN direcionada para a classificação de imagens médicas. Os autores produziram imagens sintéticas derivadas descritores como percolação, dimensão fractal e lacunaridade. O método recorreu a dois modelos CNN para extrair *deep features*

das últimas camadas. Um modelo CNN foi aplicado sobre as imagens originais e outro sobre as imagens sintéticas. A classificação foi obtida via regra da soma, considerando os resultados de ambos os modelos de CNN. Os autores relataram que os resultados foram acurácias superiores a 89% para imagens histológicas colorretais e linfomas não-Hodgkin. Para imagens histológicas de tecidos hepáticos e representativas de câncer de mama, os resultados foram superiores a 99%.

Conforme descrição apresentada por Turkoglu (2021), foi realizado um *ensemble* de *deep features* a partir de valores extraídos da saída de cada da camada da rede AlexNet para investigar radiografias pulmonares. As classes estudadas foram definidas como saudável, pneumonia e COVID. Para os experimentos, a técnica de transferência de aprendizado foi aplicada para a obtenção das *deep features*, explorando a arquitetura AlexNet pré-treinada com a base ImageNet. O *ensemble* foi definido com 10568 descritores, obtidos via algoritmo de ranqueamento Relief. O processo de classificação foi realizado com o algoritmo SVM. A partir dos resultados, os autores identificaram que o *ensemble* superou os resultados dos descritores de cada camada individualmente. Além disto, a acurácia obtida foi de 99.18%, superando os demais trabalhos no contexto aqui explorado.

Ainda com *ensemble* de *deep features*, Koc et al. (2022) exploraram as saídas das três camadas *fully-connected* das redes VGG-16 e VGG-19, como forma de investigar imagens de ressonância magnética representativas de câncer prostático. A proposta também usou transferência de aprendizado para possibilitar o emprego de CNN, ambas as VGG-16 e VGG-19 foram pré-treinadas na base ImageNet. A extração dos valores *deep-learned* forneceram 18384 descritores, dos quais foram selecionados apenas os 500 mais relevantes para compor o vetor de características das imagens. Este ranqueamento foi obtido com a aplicação do algoritmo *Neighbourhood Componentes Analysis* (NCA) e, para a classificação dos vetores, foi aplicado o algoritmo K-*Nearest Neighbours* (KNN), sendo que estes algoritmos exploram o espaço na vizinhança dos dados.

Os resultados indicaram que o *ensemble* conseguiu superar os resultados obtidos por cada conjunto de descritores. Além disto, o modelo proposto forneceu acurácia de 98.01%, tornando viável sua aplicação para o diagnóstico automático de câncer prostático. Os autores concluíram ser possível expandir a técnica com a exploração de outras arquiteturas CNN.

3.2 *Ensemble* de Classificadores

Com a finalidade de verificar a possibilidade de melhora no desempenho da classificação de câncer de mama, Gupta e Gupta (2018) conduziram um experimento que comparou as taxas de classificações obtidas com e sem *ensemble* de classificadores. Os autores utilizaram quatro classificadores diferentes, KNN, SVM, *Decision Tree* e Regressão Logística. O vetor de características foi definido por meio de 52 descritores obtidos de imagens histológicas. O *ensemble* de classificadores utilizou votação ponderada. Na votação, os pesos de cada classificador foram calculados por meio do algoritmo de otimização de *Sequential Least Squares Programming*. A partir dos resultados obtidos, os autores identificaram que houve uma melhora de desempenho na classificação fornecida pelo *ensemble* em relação aos classificadores independentes. Além disso, também foram comparados os desempenhos obtidos a partir da regra de votação simples e ponderada. A votação ponderada indicou os melhores resultados. No entanto, os autores destacaram que votação simples superou a precisão dos classificadores independentes.

Nanni, Lumini e Zaffonato (2018) realizaram um estudo com *ensemble* de classificadores para investigar imagens de ressonância magnética representativas da doença de Alzheimer em estado inicial. O estudo foi realizado analisando combinações distintas de *ensembles* heterogêneos e seletores de descritores. Para tanto, foram utilizados quatro classificadores para a composição dos *ensembles*: um método baseado em função; um Bayesiano; e, dois baseados em *ensembles* homogêneos com *boosting*. Os

ensembles foram formados a partir de uma variação do método *Static Classifier Selection*. Neste método, cada um dos componentes do *ensemble* utiliza somente uma parte do vetor de características para obter sua classificação. Os resultados obtidos indicaram que a aplicação de *ensembles* na classificação forneceu desempenho superior ao dos classificadores individuais. Contudo, a acurácia obtida no melhor caso, aproximadamente 55%, é equivalente ao valor fornecido por outros métodos propostos para a mesma finalidade. Os autores destacaram que os resultados foram relevantes por considerar uma redução da dimensão do vetor de características em mais de 90% e obter desempenhos compatíveis com a literatura.

Na área de imagens hiperspectrais, Chen et al. (2019) propuseram um modelo de *ensemble* de classificadores a partir de diferentes modelos de CNN, por exemplo, da família ResNet. Para fins de comparação, os autores incluíram uma estratégia de classificação independente com variações aplicadas sobre o classificador SVM. A classificação por meio de *ensemble* foi obtida aplicando a regra de votação simples. Os experimentos mostraram que os *ensembles* de modelos de CNN forneceram desempenhos superiores aos conquistados via classificadores individuais e *ensembles* de SVM. Além disso, os autores observaram que os desempenhos não foram afetados com o uso de transferência de conhecimento.

Com um *ensemble* composto por cinco classificadores diferentes, Das e Biswas (2019) propuseram um método para classificação de imagens de câncer de mama, divididas em grupos maligno e benigno. O *ensemble* proposto foi heterogêneo, definidos pelos classificadores *Naive Bayes*, KNN, *Decision Tree*, *Random Forest* e SVM. A classificação foi realizada a partir votação simples. Os resultados obtidos mostraram que a avaliação do *ensemble* conseguiu indicar precisão e acurácia superiores aos resultados conquistados por meio de estratégias individuais.

Brunese et al. (2020) exploraram um método para classificar imagens de ressonância magnética representativas de cinco classes de tumores cerebrais. O método utilizou

ensemble de classificadores para analisar vetores de características definidos por meio de seção randômica de atributos. O *ensemble* proposto foi constituído por 10 diferentes classificadores, a saber: KNN; SVM; SVM com função de base radial; *GPC*, *C 4.5*, *RaF*, Rede Neural, *Quadratic Discriminant Analysis*, Logístico e *Naive* Bayes. O *ensemble* proposto foi definido com uma votação ponderada, considerando a capacidade de distinção de cada um dos classificadores. O peso foi calculado com base na certeza das predições individuais. Assim, os classificadores que apresentaram maior precisão receberam um peso maior em seu voto. Segundo os autores, os resultados fornecidos com esta estratégia superaram os resultados disponíveis na literatura.

A fim de classificar imagens histopatológicas cervicais, Xue et al. (2020) propuseram um *ensemble* constituído por modelos distintos de CNN. Os autores utilizaram as arquiteturas Inception-V3, Xception, VGG-16 e ResNet-50 para composição do *ensemble*. A classificação foi definida por meio da votação com pesos, determinados por um algoritmo de otimização. Os autores exploraram imagens representativas de diferentes grupos e com múltiplos métodos de coloração, além de explorar a estratégia de transferência de conhecimento. O uso da transferência de conhecimento foi definido para identificar sua influência no desempenho do modelo. Os resultados obtidos indicaram que o *ensemble* proposto superou a precisão de todos os outros métodos. Os autores concluíram também que a transferência de conhecimento conseguiu indicar bons resultados, além de reduzir o tempo de execução do modelo.

Ainda com um *ensemble* de classificadores constituídos apenas por CNN, Sułot et al. (2021) propuseram um *ensemble* homogêneo com uma arquitetura própria para classificar imagens de escaneamento oftalmológico a *laser*, buscando identificar imagens representativas de glaucoma. A arquitetura proposta considerou três blocos com duas ou três camadas convolucionais, seguidos de duas camadas *fully-connected*. O modo de treinamento realizado foi baseado na validação cruzada, na qual o conjunto de dados foi divido em k blocos. Desta forma, cada modelo foi treinado utilizando

$k-1$ destes blocos, sendo que cada modelo excluiu um bloco diferente. Foram verificados múltiplos métodos para a obtenção da classificação final. A votação simples superou as demais propostas em todos os casos. O *ensemble* conseguiu obter uma acurácia de 96,20%, superando outros métodos de aprendizado de máquina. Este desempenho também superou o resultado obtido pelo mesmo *ensemble* via arquitetura Inception-V3. Apesar dos bons resultados, os autores identificaram limitações referentes ao número de imagens disponíveis. Como proposta futura, os autores sugerem contornar esta limitação com o uso de *data-augmentation*.

No contexto de radiografias pulmonares, Islam e Nahiduzzaman (2022) propuseram um método para classificar imagens pulmonares saudáveis e representativas de casos de COVID. Para tal tarefa, foi proposta uma CNN que permitiu extrair 100 descritores das imagens. Então, para a classificação, foi construído um *ensemble* heterogêneo com cinco técnicas (SVM, *Naive* Bayes, *Random Forest*, *Decision Tree* e Regressão Linear). O modelo elaborado resultou em uma acurácia de 99,73%, valor superior aos indicados em outras técnicas propostas na literatura. Além disto, os autores notaram que a proposta evitou casos de falsos positivos. Apesar destes destaques, os autores apontam a etapa de treinamento da CNN como principal limitação da proposta.

Finalmente, Arjmand et al. (2022) propuseram um *ensemble* explorando CNNs para classificar imagens de representativas de quatro estágios de Doença Gordurosa Não Alcoólica do Pâncreas (NAFPD). O *ensemble* foi definido com quatro arquiteturas CNN comuns, sendo AlexNet, ResNet-50, VGG-16 e VGG-19. O método de votação simples foi o escolhido. A aplicação destas arquiteturas explorou a estratégia de transferência de aprendizado via ImageNet e refinamento a partir da base sob investigação. Além disto, os modelos CNN foram empregados para segmentar regiões de interesse com as técnicas de IA explicável, tais como LIME e CAM. Os resultados observados indicaram que o uso de *ensembles* permitiu acurácia de 98,25%, superando

os valores obtidos vi arquiteturas aplicadas individualmente. Entretanto, os autores notaram um aumento no custo computacional para a aplicação do modelo, tanto na etapa de treinamento, quanto para a classificação.

3.3 Considerações sobre os Trabalhos Relacionados

Os trabalhos citados neste capítulo permitem uma visão geral sobre o uso de técnicas *ensemble*, tanto para a composição de vetores de descritores (NANNI; LUMINI; ZAFFONATO, 2018; WANG et al., 2019b; ZAFER, 2020; ROBERTO et al., 2021), quanto para a classificação de imagens (CHEN et al., 2019; XUE et al., 2020). Por exemplo, na elaboração de *ensemble* de classificadores, nota-se uma tendência de aplicação de algoritmos menos custosos, tais como (GUPTA; GUPTA, 2018; NANNI; LUMINI; ZAFFONATO, 2018; DAS; BISWAS, 2019; ISLAM; NAHIDUZZAMAN, 2022). Esses trabalhos efetuaram o uso de estratégias heterogêneas com o emprego de votação e regra da soma. No entanto, foram propostos *ensembles* construídos apenas com CNNs (SUŁOT et al., 2021; ISLAM; NAHIDUZZAMAN, 2022). Nota-se que o uso desta estratégia permitiu obter os melhores resultados. Contudo, cabe ressaltar que arquiteturas de CNNs, especialmente as muito profundas, têm alto custo computacional para seu treinamento. Simultaneamente, a aplicação de algoritmos como KNN, *Random Forest*, SVM e *Naive* Bayes forneceram resultados relevantes sem a necessidade de um alto custo de processamento. Estas observações são úteis para embasar a metodologia proposta.

Em relação ao *ensemble* de descritores, nota-se uma tendência no uso de *deep features* extraídas das camadas finais de uma CNN (NANNI; LUMINI; ZAFFONATO, 2018; HAGERTY et al., 2019; KOC et al., 2022). Outro destaque é o uso da base ImageNet para aplicação da estratégia de transferência de aprendizado (TURKOGLU, 2021; KOC et al., 2022). Adicionalmente, nota-se que a aplicação de técnicas *handcrafted* com *deep learned* não são descartadas e também garantem uma melhora signi-

ficativa ao desempenho da classificação (WANG et al., 2019a; ALMARAZ-DAMIAN et al., 2020).

Contudo, até o momento, não foram exploradas as combinações envolvendo os descritores *deep learned* de distintas arquiteturas com as técnicas fractais, como as indicadas neste trabalho. Mais ainda, o uso de representações geradas a partir de inteligência artificial explicável, apesar de utilizadas com sucesso para compreensão dos modelos de classificação, ainda não foram investigadas como parte do processo de composição de vetores de características. Desta forma, torna-se interessante explorar os atributos *handcrafted* baseados em técnicas fractais para a classificação destas imagens, uma vez que as regiões de interesse foram visualmente destacadas.

Também, apesar das diferentes aplicações, a classificação e o reconhecimento de padrões em amostras H&E representativas do câncer colorretal, câncer de mama, tecido hepático e displasia oral não foram totalmente exploradas, com questões ainda pertinentes, tais como: se é possível determinar padrões de combinações de técnicas (descritores e ensemble) para diferentes tipos de imagens H&E; se descritores fractais multidimensionais permitem aprimorar os resultados conquistados via *deep learned features* com *transfer learning* no contexto aqui explorado; se os descritores *handcrafted* fractais obtidos das imagens resultantes de técnicas de IAx permitem um melhor desempenho para a classificação; e se a associação de ensembles (descritores e classificadores) pode fornecer resultados competitivos em relação aos disponíveis na literatura.

Por fim, algumas arquiteturas como DenseNet-121 (HUANG et al., 2017), Inception-V3 (SZEGEDY et al., 2016), ResNet-50 (HE et al., 2016), e a VGG-19 (SIMONYAN; ZISSERMAN, 2014) ainda podem ser exploradas para fornecer *deep learned features*, mesmo com o sucesso fornecido na classificação direta de alguns tipos de imagens histológicas, tais como os observados em (REDDY; JULIET, 2019; GANGULY; DAS; SETUA, 2020; AL-HAIJA; ADEBANJO, 2020; HAGERTY et al., 2019). Modelos

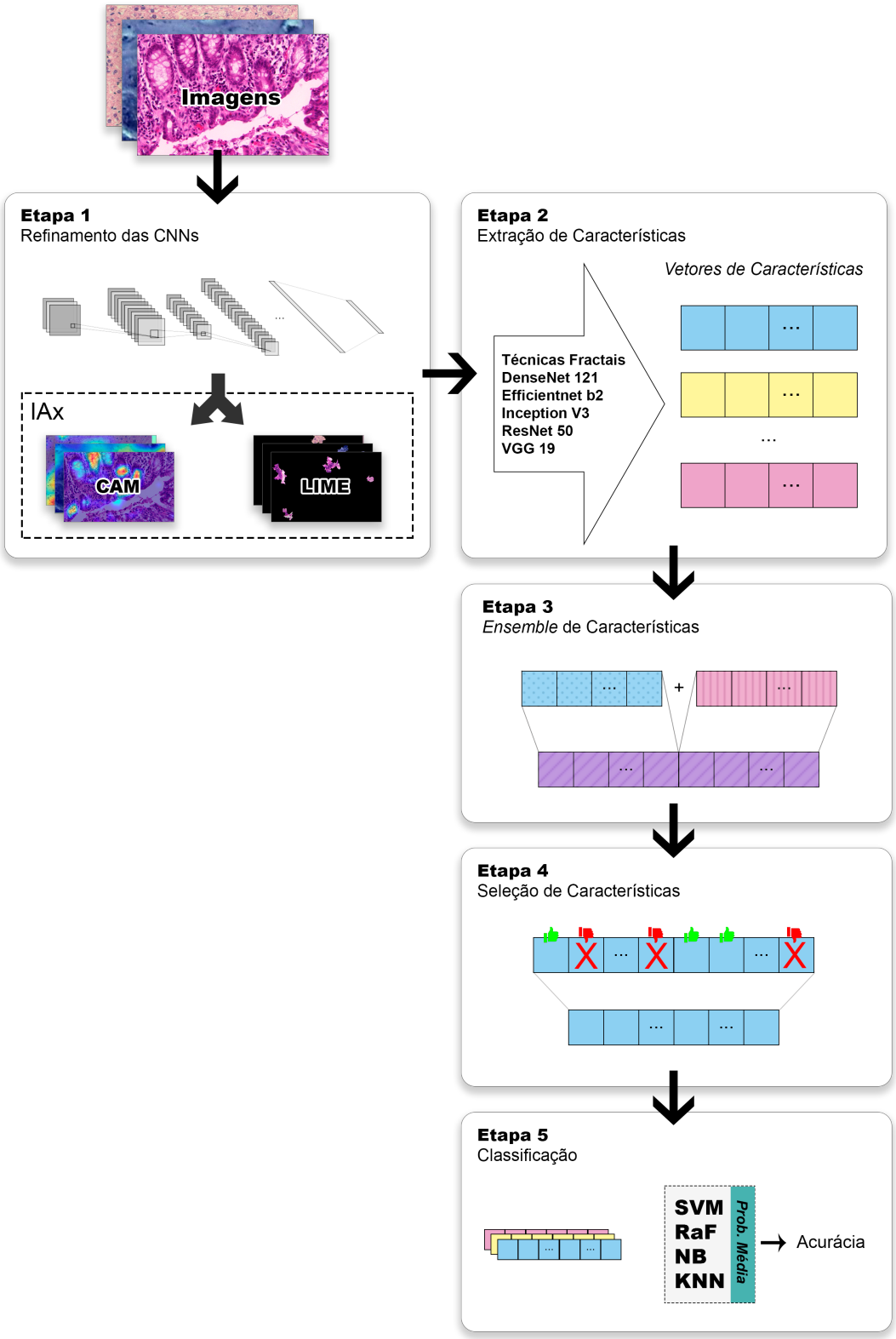
mais recentes, como a EfficientNet (TAN; LE, 2019), ainda não foram totalmente explorados no contexto de *deep features* como os aqui expostos.

Capítulo 4

Metodologia

Neste capítulo estão descritas as etapas propostas para definir o modelo fundamentado no uso combinado de *deep features* via transfer Learning com descritores fractais obtidos a partir de imagens H&E e as representações CAM e LIME correspondentes. Na primeira etapa, foram definidas as arquiteturas de CNN e realizados os refinamentos nos parâmetros da última camada de cada modelo. Também, as representações via LIME e CAM foram extraídas com a análise das interpretações fornecidas pelas redes refinadas. Em seguida, a segunda etapa foi definida para obter os descritores fractais multiescalas e multidimensionais das imagens H&E e de suas representações IAx, bem como as *deep features* de múltiplas arquiteturas de CNN. Os resultados são vetores de características representativos de cada origem. Na etapa 3, os valores extraídos são organizados em *ensembles* de descritores, via agrupamento, para verificar as relevâncias das diferentes combinações. Na etapa 4, os vetores e *ensembles* de descritores são avaliados aplicando uma técnica para selecionar os descritores mais relevantes. Na etapa 5, os descritores selecionados são utilizados para distinguir as imagens H&E por meio de um *ensemble* de classificadores. Nas próximas seções são apresentados os detalhes de cada uma das etapas, bem como informações sobre os conjuntos de imagens (seção 4.1) e dos pacotes de softwares para viabilizar esta proposta (seção 4.7). Um resumo esquemático do modelo está ilustrado na Figura 4.1.

Figura 4.1: Diagrama ilustrativo das etapas que estruturam o método proposto.



Fonte: Elaborado pelo Autor

4.1 Contextos de aplicação

A investigação proposta neste trabalho foi realizada em quatro bases de imagens histológicas H&E, a saber: tumores de mama (UCSB) (GELASCA et al., 2008); tumores colorretais (CR) (SIRINUKUNWATTANA et al., 2017), tecido hepático (LG) (AGEMAP, 2020) ; e displasia epitelial oral (OED) (SILVA et al., 2022). Um resumo da composição de cada base está disponível na Tabela 4.1 e, na Figura 4.2, estão exemplos de amostras que constituem cada conjunto.

Tabela 4.1: Informações sobre as bases de imagens histológicas H&E.

Base	Tipo de imagem	Nº de Classes	Classes	Amostras	Resolução
UCSB	Tumores de mama	2	Maligno e Benigno	58 (32/26)	896×768
CR	Tumores colorretais	2	Maligno e Benigno	165 (74/91)	de 567×430 a 775×552
LG	Tecido hepático	2	Macho e Fêmea	265 (150/115)	417×312
OED	Displasia Epitelial Oral	2	Saudável e Severo	148 (74/74)	450×250

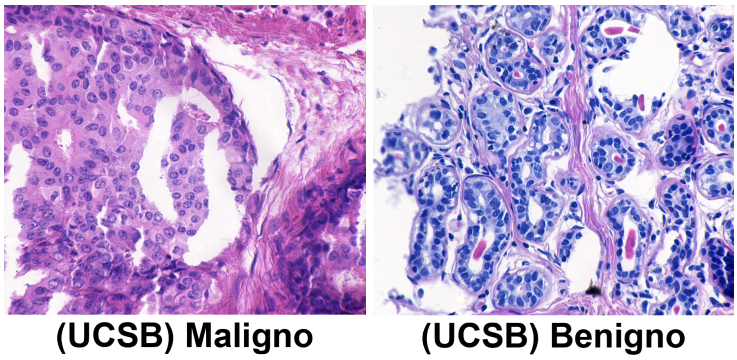
4.2 Etapa 1 - Refinamento das CNNs e Obtenção das Representações IAx

Foram selecionadas cinco arquiteturas de CNN neste estudo: DenseNet-121 (HUANG et al., 2017); EfficientNet-b2 (TAN; LE, 2019); Inception-V3 (SZEGEDY et al., 2016); ResNet-50 (HE et al., 2016); VGG-19 (SIMONYAN; ZISSERMAN, 2014). Estas arquiteturas foram selecionadas pelo uso de estratégias distintas no processamento e classificação das imagens de entrada, fato que pode garantir uma investigação mais ampla de técnicas com seus descritores *deep learned*. As redes neurais utilizadas foram recuperadas da biblioteca PyTorch e estão sumarizados na Tabela 4.2, incluindo os valores de acurácias obtidos para cada uma das arquiteturas quando aplicadas no contexto em que foram treinadas, a base de imagens ImageNet (DENG et al., 2009).

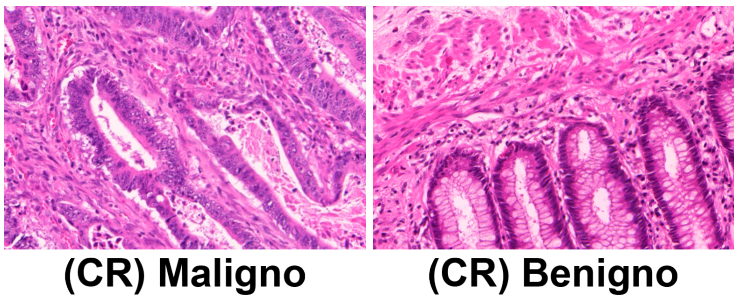
O refinamento no aprendizado de cada CNN foi realizado para alterar o resultado de cada camada de saída que, em razão do uso de transferência de aprendizado via

Figura 4.2: Exemplos de imagens H&E das bases USBC (a), CR (b), LG (c) e OED (d), disponibilizadas por (GELASCA et al., 2008; SIRINUKUNWATTANA et al., 2017; AGEMAP, 2020; SILVA et al., 2022), respectivamente.

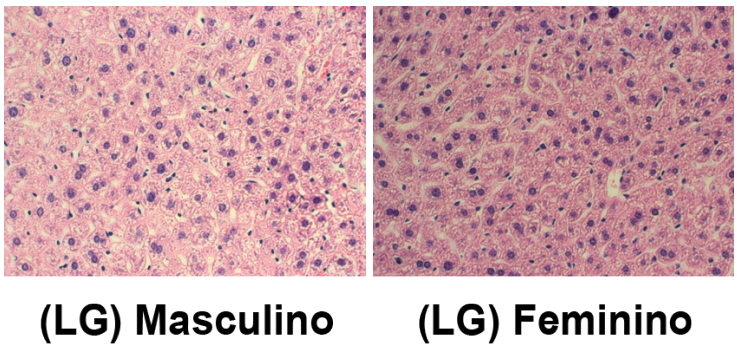
(a)



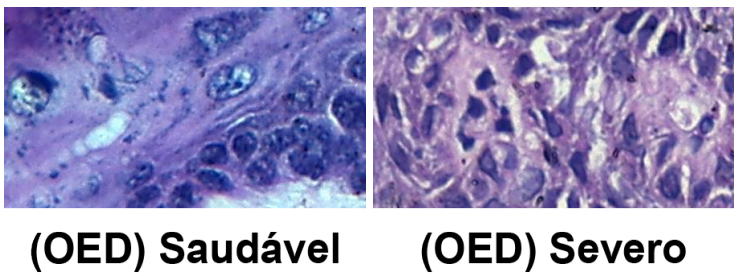
(b)



(c)



(d)



Fonte: Elaborado pelo Autor

Tabela 4.2: Informações das arquiteturas pré-treinadas utilizadas no processo de definição das *deep learned*.

Arquitetura	**Parâmetros**	**Camadas**	**Acurácia (%) (ImageNet)**
DenseNet-121	8×10^6	121	91,97%
EfficientNet-b2	$9{,}1 \times 10^6$	324	95,31%
Inception-V3	$2{,}7 \times 10^7$	48	93,45%
ResNet-50	$2{,}6 \times 10^7$	50	92,86%
VGG-19	$1{,}4 \times 10^8$	19	90,87%

ImageNet, com 1000 valores mapeados às classes reconhecíveis originalmente, foi ajustado para identificar somente duas saídas: número de classes existentes nas bases de imagens H&E sob investigação, conforme Tabela 4.1. Essa estratégia evitou a etapa de treinamento total da rede e viabilizou a investigação de conjuntos com um número reduzidos de imagens. Assim, essa modificação alterou as conexões finais e pesos correspondentes ao total de grupos de cada base H&E. É importante ressaltar que foi preservado o conhecimento contido em cada camada que antecede a camada de saída, mantendo os pesos obtidos via ImageNet.

O processo de treinamento (refinamento) foi aplicado com o uso da técnica *cross-validation* k-*folds*. Com esta técnica, as amostras foram divididas em $k = 10$ *folds*. Em seguida, foram realizadas k iterações do processo, sendo que, em cada um deles, um dos *folds* foi utilizado como conjunto de teste, enquanto os demais foram considerados para treinamento. Esta técnica foi aplicada para reduzir a possibilidade de ocorrência de sobreajuste.

Os treinamentos foram realizados considerando 10 iterações, ou épocas, com a aplicação da técnica de descida de gradiente estocástico (SGDM); taxa de aprendizado inicial $lr = 0.01$ com redução a cada 2 épocas em um fator de 0.75; e a função de entropia cruzada para a obtenção do erro/ajuste em cada época. Este processo foi realizado para cada combinação de arquitetura e base de imagens. É importante destacar que foi aplicada uma normalização nos valores nos canais de cores de cada imagem, conforme os valores de média e desvio padrão obtidos da base de treinamento, para

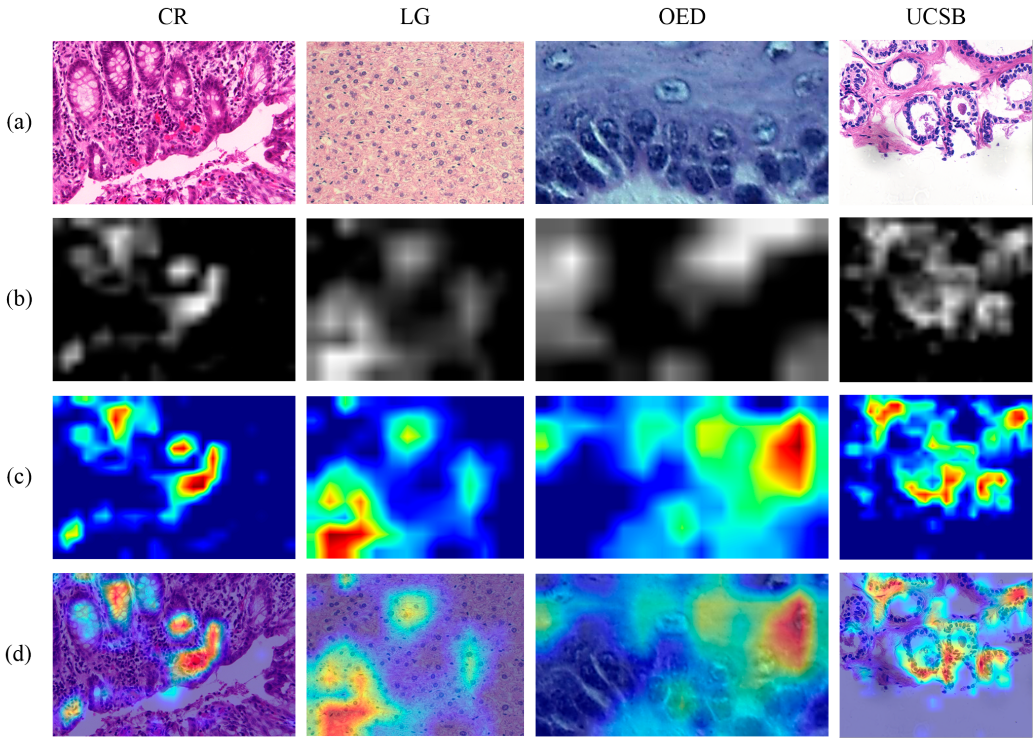
tornar as entradas compatíveis com as utilizadas em seu treinamento original (TORCHVISION..., 2021). Finalmente, a rede obtida como resultado do refinamento foi definida durante o processo, selecionando a que apresentou maior valor de acurácia ao classificar o conjunto de teste, independente da época.

4.2.1 Geração das representações IAx: LIME e CAM

A geração de cada representação CAM e LIME foi realizada via cada CNN refinada para todas as imagens H&E correspondentes ao conjunto que realizou o processo de refinamento. As representações CAM foram obtidas da abordagem Grad-CAM (SELVARAJU et al., 2017). Apesar desta abordagem ser aplicável à qualquer camada convolucional de uma CNN, neste trabalho, a última camada convolucional foi definida para fornecer as representações. Esta escolha foi fundamentada na ideia de que as camadas mais profundas tendem a conter informações mais relacionadas com os padrões globais presentes nos conjuntos sob investigação (SANTOS; PONTI, 2019). O retorno deste processo foi um mapeamento com a medida de contribuição de cada pixel, explorando uma escala de 0 a 1 para a classificação da rede. Este mapeamento foi então transformado em um mapa de calor e sobreposto à imagem original com opacidade de 50%. O resultado representou a imagem definida como explicação do processo de classificação de cada CNN em questão. Na Figura 4.3 são ilustradas representações Grad-CAM de algumas amostras de imagens H&E.

Em relação ao método LIME, foram utilizadas 1000 perturbações para cada imagem com a aplicação do método de segmentação *quickshift*. Este é o método de segmentação padrão aplicado pela biblioteca lime v0.2.0.1 (RIBEIRO; CONTRIBUTORS, 2016), implementado pelo pacote scikit-image (WALT et al., 2014) que faz parte de suas dependências. Como resultado esperado, o método forneceu os cinco super-pixels (regiões de interesse) mais relevantes para o processo de classificação. Assim, o resultado é uma explicação com as principais regiões de interesse que fo-

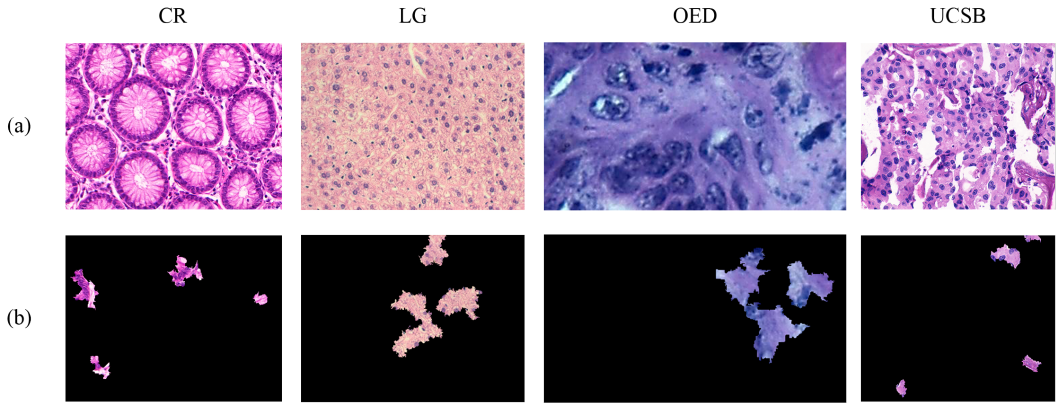
Figura 4.3: Exemplo da obtenção da explicação com a técnica Grad-CAM, sendo (a) a imagem original, (b) o peso da contribuição de cada pixel, (c) a imagem em (b) transformada em mapa de calor e (d) a sobreposição do mapa de calor na imagem original.



Fonte: Elaborado pelo Autor

ram exploradas a partir da imagem original. Na Figura 4.4 estão ilustrados alguns exemplos de resultados obtidos a partir das imagens H&E.

Figura 4.4: Exemplo de explicação obtida via técnica LIME, sendo (a) a imagem original e (b) o recorte dos 5 super-pixels selecionados como explicação.



Fonte: Elaborado pelo Autor

4.3 Etapa 2 - Extração de descritores

Na segunda etapa, os descritores foram extraídos a partir de cada uma das abordagens para compor os *ensembles* de descritores. Neste caso, os descritores foram definidas a partir de três processos de quantificações: descritores fractais das imagens H&E; valores extraídos da camada final de diferentes CNNs (*deep learned*); descritores fractais das representações LIME e Grad-CAM. Esses processos foram identificados como quantificações *handcrafted*, *deep learned* e IAx, respectivamente.

4.3.1 Descritores *Handcrafted*: Técnicas fractais multiescalas e multidimensionais

As técnicas fractais de dimensão fractal e lacunaridade foram calculadas a partir de matrizes de probabilidades, responsáveis por armazenar as probabilidades da imagem conter uma região quadrada preenchida com cada uma das formas que a constitui. A

verificação desse preenchimento ocorreu via uma relação de distância entre valores de pixels sob análise e tamanho da área em questão (CANTRELL, 2000). Assim, a matriz de probabilidades foi obtida por meio do método *Gliding-Box*, que consistiu em deslizar uma caixa, ou janela, quadrada de lado r por toda a imagem e verificar se os pixeis correspondentes estão dentro ou fora da caixa em questão (ALLAIN; CLOITRE, 1991). Assim, dada uma caixa de lado r com um pixel central p_c, um pixel p foi considerado na caixa quando sua distância d em relação a p_c foi menor ou igual a r. Esse processo resulta em uma matriz de distribuição de frequência $N(m,r)$, com m representando o número de pixels dentro de uma caixa de lado r. A matriz de probabilidades $P(m,r)$ foi obtida por meio da normalização de $N(m,r)$, Equação 4.1, dividindo o valor de cada contagem por n_r, que representa o total de caixas de lado r contidos na imagem com a aplicação do *gliding-box*, Equação 4.2.

$$P(m,r) = \frac{N(m,r)}{n_r}. \tag{4.1}$$

$$n_r = (largura - r + 1) * (altura - r + 1). \tag{4.2}$$

A estratégia descrita previamente caracteriza uma quantificação multiescala em razão da variação de r. De maneira complementar, assim como descrito por (IVANOVICI; RICHARD, 2010), neste trabalho foi aplicada a distância de Chebyshev ou *Chessboard* para calcular d. Por fim, como o contexto de aplicação usa imagens coloridas no padrões de cores RGB, a estratégia multidimensional foi aplicada considerando um pixel representado por um espaço 5-dimensional, tal como (x,y,r,g,b) (IVANOVICI; RICHARD, 2010).

A partir destes procedimentos, os descritores de dimensão fractal e lacunaridade, ambos multiescalas e multidimensionais, foram calculados por meio da estratégia de matriz de probabilidades (IVANOVICI; RICHARD, 2010). A dimensão fractal D foi calculada considerando a estimativa de caixas de lado r necessárias para cobrir uma

imagem, Equação 4.3. A lacunaridade L foi calculada a partir do primeiro e segundo momentos da matriz de probabilidades, obtidos conforme as Equações 4.4 e 4.5, respectivamente. Esses momentos foram combinados por meio da Equação 4.6 e resultaram na medida da lacunaridade em uma escala específica de observação.

$$D(r) = \sum_{m=1}^{r^2} \frac{P(m,r)}{m}. \tag{4.3}$$

$$\mu(r) = \sum_{m=1}^{r^2} mP(m,r). \tag{4.4}$$

$$\mu^2(r) = \sum_{m=1}^{r^2} m^2P(m,r). \tag{4.5}$$

$$L(r) = \frac{\mu^2(r) - (\mu(r))^2}{(\mu(r))^2}. \tag{4.6}$$

As medidas de percolação também foram extraídas por meio do algoritmo de *gliding-box*, seguindo a estratégia descrita por (ROBERTO et al., 2017). Logo, dada uma caixa de lado r, os pixels contidos na caixa foram considerados poros e rotulados com a aplicação do algoritmo Hoshen-Kopelman (HOSHEN; KOPELMAN, 1976). Os poros com um mesmo rótulo foram entendidos como parte de um mesmo aglomerado. Este processo foi repetido para cada caixa de lado r. Com isso, foi possível computar as métricas: $C(r)$, que indica a média de aglomerados presentes em cada caixa; $Q(r)$, que define a razão de cobertura média do maior aglomerado; e $P(r)$, que fornece a razão de caixas percolantes.

A métrica $C(r)$ é obtida via uma média da contagem do número de aglomerados c, presente em cada caixa i, conforme a Equação 4.7.

$$C(r) = \frac{\sum_{i=1}^{n_r} c_i}{n_r} \tag{4.7}$$

A segunda, $Q(r)$, é obtida considerando uma média do tamanho do maior aglomerado $|c_{max}|$ em cada uma das i caixas, conforme a Equação 4.8.

$$Q(r) = \frac{\sum_{i=1}^{n_r} |c_{max_i}|}{n_r} \tag{4.8}$$

Por fim, a razão de caixas percolantes é obtida dividindo o número de caixas em que ocorreu a percolação pelo número de caixas computadas em uma imagem. A percolação em uma caixa p_i ocorreu quando a razão entre o número de poros Ω_i e o número total de pixels na caixa r^2 excedeu o limiar de percolação 0,59275, conforme a Equação 4.9. Assim, $P(r)$ foi definida conforme a Equação 4.10

$$p_i = \begin{cases} 1, & \frac{\Omega_i}{r^2} \geq 0{,}59275 \\ 0, & \frac{\Omega_i}{r^2} < 0{,}59275 \end{cases} \tag{4.9}$$

$$P(r) = \frac{\sum_{i=1}^{n_r} p_i}{n_r} \tag{4.10}$$

É importante destacar que as quantificações realizadas neste trabalho utilizaram matrizes construídas a partir de caixas r, respeitando o intervalo $3 \leq r \leq 43$, conforme descrições apresentadas por (IVANOVICI; RICHARD, 2010; ROBERTO et al., 2021). Além disso, o parâmetro r foi definido como um valor ímpar para garantir a existência de pixel central em cada caixa. Logo, o incremento no valor de r foi de duas unidades em cada iteração. A construção da matriz de probabilidades a partir destes parâmetros garantiu quantificações em 20 escalas diferentes.

Adicionalmente, cada uma das funções de lacunaridade e medidas de percolação também foi representada por escalares a fim de obter descritores representativos de possíveis padrões existentes em cada curva (IVANOVICI; RICHARD, 2010; CĂLIMAN; IVANOVICI, 2012; ROBERTO et al., 2017; CĂLIMAN; IVANOVICI, 2012; Segato dos Santos et al., 2018; TOSTA et al., 2019; RIBEIRO et al., 2019; ROBERTO et al., 2021). Os descritores escalares foram o valor máximo, assimetria, área sob a

curva e média entre as metades da área sob a curva, de forma que:

- **Área sob a curva (A):** possibilita quantificar a complexidade da textura, visto que texturas semelhantes podem apresentar valores próximos. Para uma função discreta constituída por N pontos definidos em $x_1,...,x_n$, este descritor pode ser obtido via Equação 4.11, com a e b como os índices dos pontos que delimitam o intervalo de análise;

$$A(a,b) = \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(x_n) + f(x_{n+1})). \tag{4.11}$$

- **Assimetria (O):** indica que imagens de uma mesma classe têm assimetrias no mesmo sentido. Este descritor é determinado a partir da observação do terceiro momento da função, mas que para uma função discreta é possível obter vi Equação 4.12, em que N é o número de pontos na função, x_i é o i-ésimo ponto na função está definida, $\bar{x}$ é a média dos valores da função, e a e b os índices dos pontos que delimitam o intervalo;

$$O(a,b) = \frac{\frac{1}{N}\sum_{i=a}^{b}(x_i - \bar{x})^3}{\sqrt{[\frac{1}{N}\sum_{i=a}^{b}(x_i - \bar{x})^2]^3}}. \tag{4.12}$$

- **Razão entre áreas (R):** como consequência da assimetria, a razão entre as metades da área sob a curva também deve apresentar valores próximos para classes semelhantes. Portanto, este descritor é obtido com a aplicação da Equação 4.11, sendo a e b os índices dos pontos que delimitam o intervalo;

$$R(a,b) = \frac{A(\frac{b}{2}+1, b)}{A(a, \frac{b}{2})}. \tag{4.13}$$

- **Ponto de Máximo:** indica o valor na maior área heterogênea da curva. Assim, imagens de uma mesma classe podem apresentar valores próximos, tanto para

$f(x)$, quanto para x. Para classes distintas, espera-se que os valores também sejam totalmente distintos.

Finalmente, os descritores fractais citados foram organizados em um vetor de características com 116 descritores: 20 dimensões fractais; 20 lacunaridades; 20 médias de número aglomerados presentes em cada caixa; 20 razões de cobertura média do maior aglomerado; 20 razões de caixas percolantes; e 16 descritores de curva, sendo quatro de cada uma das funções $L(r)$, $C(r)$, $Q(r)$ e $P(r)$.

4.3.2 Descritores *Deep Learned*

Neste trabalho, os descritores *Deep Learned* foram extraídos de cinco arquiteturas CNNs: DenseNet-121 (HUANG et al., 2017); EfficientNet-b2 (TAN; LE, 2019); Inception-V3 (SZEGEDY et al., 2016); ResNet-50 (HE et al., 2016); VGG-19 (SIMONYAN; ZISSERMAN, 2014). Como resultado, foram obtidos cinco vetores de características *deep learned*, um para cada uma das redes. É importante destacar que, de uma maneira geral, os valores nas camadas iniciais de uma CNN definem quantificações de padrões locais, tais como forma, borda e cor. As camadas mais profundas são úteis para identificar padrões globais, tais como textura e semântica (SANTOS; PONTI, 2019). Assim, para explorar os padrões globais, da arquitetura DenseNet-121, a camada de normalização após o último bloco denso foi a escolhida neste trabalho. Essa camada forneceu 1024 valores. Das arquiteturas EfficientNet-b2, Inception-V3 e ResNet-50, cada camada final *average pooling* contribuiu com 1408, 2048 e 2048, respectivamente. Da VGG-19, a extração ocorreu na última camada *fully-connected*, fornecendo um vetor com 4096.

4.4 Etapa 3 - *Ensemble* de Descritores

Esta etapa foi definida para organizar os vetores de características e construir os *ensembles*. Para isso, foram consideradas combinações entre cada um dos conjuntos de descritores *handcrafted* e *deep learned*, via agrupamento dos valores, a fim de analisar suas capacidades discriminativas em cada uma das bases de imagens H&E (DASIGI; MANN; PROTOPOPESCU, 2001). É importante reforçar que cada representação obtida por meio das técnicas de IAx foi quantificada com as abordagens fractais (subseção 4.3.1), completando o conjunto *handcrafted*. Desta forma, para cada uma das cinco arquiteturas CNNs, foram definidos dois vetores de características de representações IAx via técnicas fractais, sendo um derivado das explicações com CAM e outro com LIME. Cada vetor foi constituído por 116 descritores: 20 valores de dimensões fractais; 20 valores de lacunaridades; 20 valores médios de números aglomerados presentes em cada caixa; 20 valores de razões de cobertura média do maior aglomerado; 20 valores de razões de caixas percolantes; e 16 descritores de curva, sendo quatro para cada uma das métricas $L(r)$, $C(r)$, $Q(r)$ e $P(r)$.

Os descritores explorados neste trabalho foram organizados em 55 composições distintas. Destas, 16 composições são vetores individuais de acordo com suas origens e número de descritores disponíveis. Dentre os vetores, temos três grupos de origens distintas, sendo *handcrafted*, *deep learned* e IAx. Os *handcrafted* são vetores obtidos com a aplicação das técnicas *handcrafted* fractais, faz parte deste grupo o vetor nomeado Fractais (F), com 116 descritores. Os *deep learned* são vetores obtidos da extração dos valores da camada final de uma CNN, são eles: DenseNet-121 (D) com 1024 descritores; EfficientNet-b2 (E) com 1408 descritores; Inception-V3 (I) com 2048 descritores; ResNet-50 (R) com 2048 descritores; VGG-19 (V) com 4096 descritores. Por fim, os IAx são vetores compostos de descritores *handcrafted* obtidos via aplicação das técnicas fractais para quantificar as representações de IAx. Neste caso, os IAx são constituídos por 116 descritores e podem ser divididos ainda em dois

grupos: resultantes das explicações CAM, com CAM DenseNet-121 (D_{CAM}), CAM EfficientNet-b2 (E_{CAM}), CAM Inception-V3 (I_{CAM}), CAM ResNet-50 (R_{CAM}) e CAM VGG-19 (V_{CAM}); e os resultantes das explicações LIME, com LIME DenseNet-121 (D_{LIME}), LIME EfficientNet-b2 (E_{LIME}), LIME Inception-V3 (I_{LIME}), LIME ResNet-50 (R_{LIME}) e LIME VGG-19 (V_{LIME}). As 16 composições de vetores com suas origens e números de descritores estão indicados na Tabela 4.3.

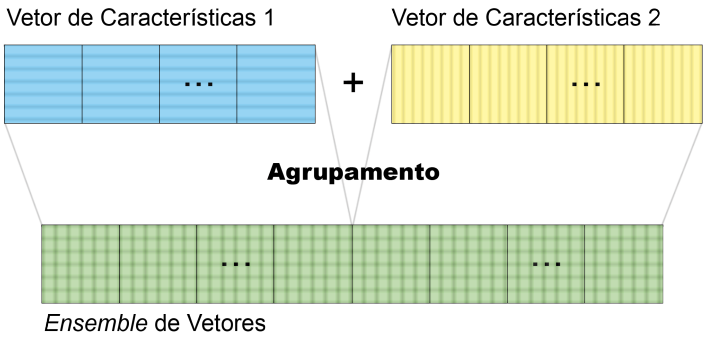
Tabela 4.3: Vetores de características obtidos a partir de diferentes técnicas.

Origem	**Composição**	**Totais de Descritores**
Handcrafted	Fractais (F)	116
Deep Learned	DenseNet-121 (D)	1024
	EfficientNet-b2 (E)	1408
	Inception-V3 (I)	2048
	ResNet-50 (R)	2048
	VGG-19 (V)	4096
IAx	CAM DenseNet-121 (D_{CAM})	116
	CAM EfficientNet-b2 (E_{CAM})	116
	CAM Inception-V3 (I_{CAM})	116
	CAM ResNet-50 (R_{CAM})	116
	CAM VGG-19 (V_{CAM})	116
	LIME DenseNet-121 (D_{LIME})	116
	LIME EfficientNet-b2 (E_{LIME})	116
	LIME Inception-V3 (I_{LIME})	116
	LIME ResNet-50 (R_{LIME})	116
	LIME VGG-19 (V_{LIME})	116

Os 39 vetores remanescentes podem ser divididos em três grupos: *ensemble* de *handcrafted* e *deep features*; ensemble de *deep feaures*; e *ensemble* de representações IAx com técnicas fractais. Uma ilustração esquemática do processo de estruturação destes vetores de características está na Figura 4.5. O primeiro grupo é composto por seis vetores, sendo cinco combinações dos fractais com os *deep learned* de uma das arquiteturas e um deles combinando os fractais com todos os outros *deep learned*, conforme resumido na Tabela 4.4. O grupo de *ensemble* de *deep features* é composto por 11 vetores, sendo 10 permutações 2 a 2 de descritores *deep learned* com origem

em cada uma das arquiteturas e um com todas as *deep features* disponíveis, Tabela 4.5. O último grupo contém 22 vetores com *ensembles* de descritores das representações IAx, organizados segundo a metodologia aplicada aos *ensembles deep learned*, Tabela 4.6.

Figura 4.5: Ilustração de combinações por agrupamento para definir os *ensembles* de descritores.



Fonte: Elaborado pelo Autor

Tabela 4.4: *Ensembles* realizadas com descritores *Handcrafted* e *Deep Learned*.

Composição	Nº de Descritores
F + D	1140
F + E	1524
F + I	2164
F + R	2164
F + V	4212
F + D + E + I + R + V	10740

4.5 Etapa 4 - Seleção de Descritores

Os resultados fornecidos pela Etapa 3 foram vetores com altas dimensionalidades, fato que pode acarretar classificações com overfitting (WATANABE, 1985). Para evitar este problema, a solução foi aplicar uma redução de dimensionalidade via técnica ReliefF (KONONENKO; ŠIMEC; ROBNIK-ŠIKONJA, 1997). Essa técnica foi considerada por permitir controlar o número de descritores sob análise, bem como pelo

Tabela 4.5: *Ensembles* envolvendo descritores *Deep Learned*.

Composição	**Nº de Descritores**
D + E	2432
D + I	3072
D + R	3072
D + V	5120
E + I	3456
E + R	3456
E + V	5504
I + R	4096
I + V	6144
R + V	6144
D + E + I + R + V	10624

Tabela 4.6: *Ensembles* considerando descritores das representações *IAx*.

Composição	**Nº de Descritores**
$D_{CAM} + E_{CAM}$	232
$D_{CAM} + I_{CAM}$	232
$D_{CAM} + R_{CAM}$	232
$D_{CAM} + V_{CAM}$	232
$E_{CAM} + I_{CAM}$	232
$E_{CAM} + R_{CAM}$	232
$E_{CAM} + V_{CAM}$	232
$I_{CAM} + R_{CAM}$	232
$I_{CAM} + V_{CAM}$	232
$R_{CAM} + V_{CAM}$	232
$D_{CAM} + E_{CAM} + I_{CAM} + R_{CAM} + V_{CAM}$	580
$D_{LIME} + E_{LIME}$	232
$D_{LIME} + I_{LIME}$	232
$D_{LIME} + R_{LIME}$	232
$D_{LIME} + V_{LIME}$	232
$E_{LIME} + I_{LIME}$	232
$E_{LIME} + R_{LIME}$	232
$E_{LIME} + V_{LIME}$	232
$I_{LIME} + R_{LIME}$	232
$I_{LIME} + V_{LIME}$	232
$R_{LIME} + V_{LIME}$	232
$D_{LIME} + E_{LIME} + I_{LIME} + R_{LIME} + V_{LIME}$	580

sucesso constatado em outros estudos (URBANOWICZ et al., 2018; GHOSH et al., 2021; DEMIR et al., 2022). É importante destacar que os vetores analisados têm di-

mensões entre 116 e 10740, sendo que existem diferentes recomendações para limitar o tamanho do espaço de descritores. Uma delas é que, para um conjunto com n amostras, o vetor deve ser composto por no máximo $\frac{n}{10}$ descritores (JAIN; DUIN; MAO, 2000). Outra recomendação indica que, para um conjunto de descritores não correlatos, o vetor deve ser composto de no máximo $n-1$ descritores (HUA et al., 2004). Neste contexto, considerando que os totais de amostras disponíveis em nossos experimentos variam de 58 a 265 imagens, os testes foram realizados com totais entre 5 e 25 descritores, inciando o processo de análise com o valor máximo de descritores e explorando decrementos de cinco descritores, até atingir a menor dimensão (RIBEIRO et al., 2019).

4.6 Etapa 5 - *Ensemble* de classificadores e Métrica de Verificação

A última etapa consistiu em realizar as classificações a partir dos vetores com um *ensemble* de classificadores heterogêneos, representativos de diferentes categorias: SVM (CORTES; VAPNIK, 1995), baseado em função; *Naive Bayes* (PEDRINI; SCHWARTZ, 2008), baseado em probabilidade; *Random Forest* (BREIMAN, 2001), baseado em árvore de decisão; e K-*Nearest Neighbors* (CUNNINGHAM; DELANY, 2020), baseado em instância. As classificações foram combinadas a partir da média de probabilidades, nessa estratégia é calculada a média entre todas as probabilidades resultantes de cada um dos classificadores, assim a classe com maior média é a selecionada como resposta.

Os resultados foram analisados considerando a métrica acurácia, capaz de indicar o desempenho global do modelo (dentre todas as classificações, quantas o modelo classificou corretamente) (MARTINEZ; LOUZADA; PEREIRA, 2003). A acurácia é dada pela Equação 4.14, em que: verdadeiros positivos (VP), valores positivos que

o modelo classificou corretamente como positivos; verdadeiros negativos (VN), valores negativos que o modelo classificou corretamente como negativos; falsos positivos (FP), valores negativos que o modelo classificou incorretamente como positivos; e falsos negativos (FN), valores positivos que o modelo classificou incorretamente como negativos.

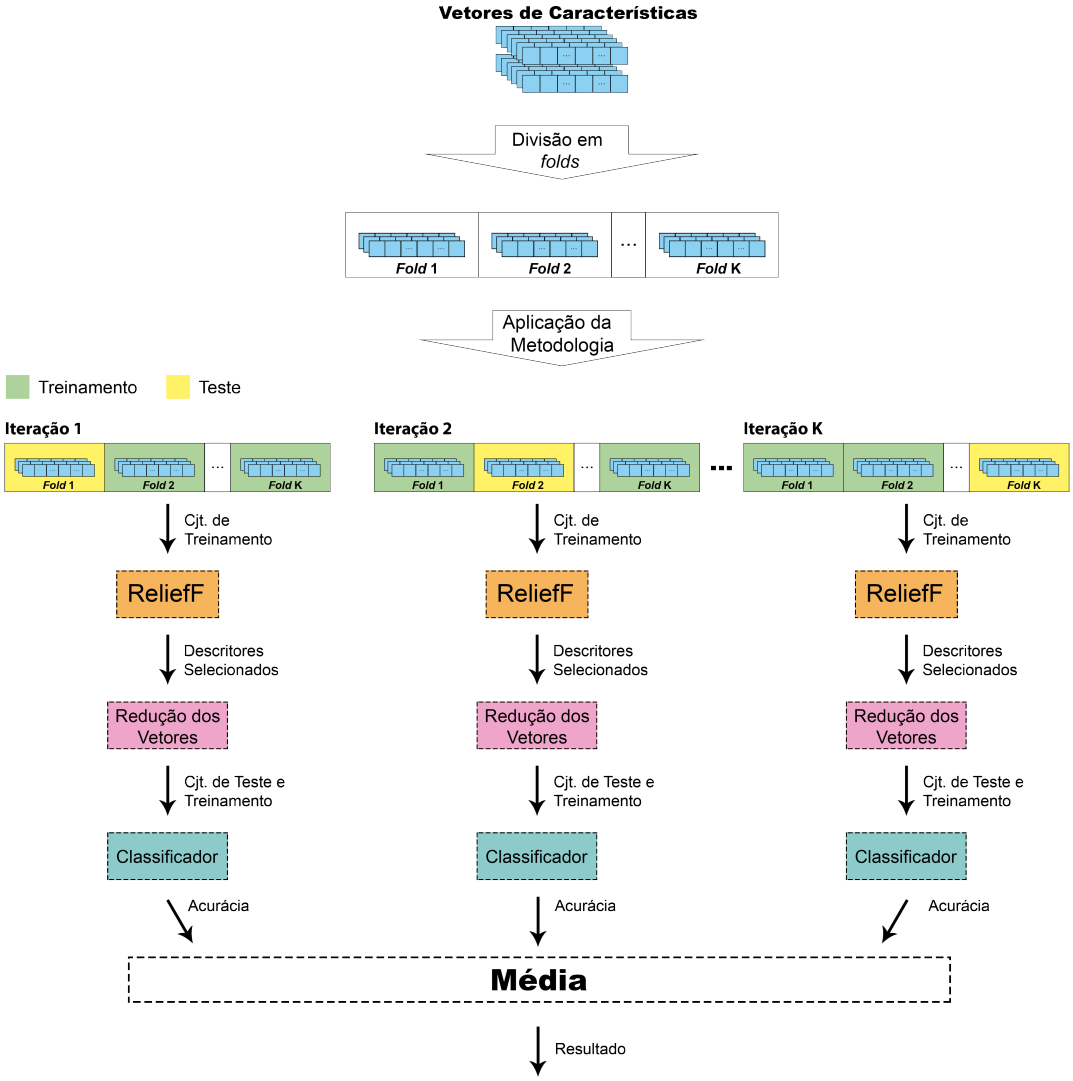
$$Acuracia = \frac{VP + VN}{VP + FP + VN + FN}. \tag{4.14}$$

Além disso, o processo de classificação também considerou a abordagem *cross-validation* k-*folds*, com $k = 10$. É importante destacar que o processo de seleção de descritores, conforme descrição apresentada na Etapa 4 desta metodologia, foi aplicado em cada conjunto de treinamento de cada *fold*. Os descritores selecionados foram utilizados para realizar a classificação no conjunto de teste correspondente. Esta técnica foi aplicada para reduzir a possibilidade de ocorrência de *overfitting*, uma vez que o resultado foi obtido por meio da média de classificações (KOHAVI et al., 1995). Na Figura 4.6 está ilustrado este processo de *cross-validation* com a seleção de descritores.

4.7 Pacotes de *software* e ambiente de execução para os experimentos

Neste trabalho foi utilizada a biblioteca de aprendizado de máquina Pytorch 1.9.0 (PASZKE et al., 2019) para definir os modelos de redes convolucionais. As explicações CAM foram obtidas via biblioteca pytorch-grad-cam 1.3.1 (GILDENBLAT; CONTRIBUTORS, 2021). As explicações LIME foram obtidas por meio da biblioteca lime 0.2.0.1 (RIBEIRO; CONTRIBUTORS, 2016), implementada em python pelos autores do artigo que propõe a técnica (RIBEIRO; SINGH; GUESTRIN, 2016). A aplicação de cada modelo foi realizada em um ambiente remoto no Google Co-

Figura 4.6: Ilustração do processo aplicado com o uso da técnica de *cross-validation* k-*folds*.



Fonte: Elaborado pelo autor

lab, que forneceu gratuitamente um console para execução de códigos em python3 e uma GPU para processamento com 12GB de memória disponível. Os algoritmos para extração dos descritores fractais foram implementados e executados em partes no *software* Matlab R2019b (MATLAB, 2019), Go e python3. A seleção de características e classificações foram realizadas com o uso da plataforma Weka v3.8.5 (FRANK et al., 2005). Nesses casos, as execuções ocorreram em um ambiente com processador Intel(R) Core(TM) i5-8265U 1.60GHz e 8GB de memória RAM.

Capítulo 5

Resultados e Discussão

Neste capítulo são apresentados os resultados conquistados com a aplicação do modelo proposto, explorando as diferentes composições de vetores de características (seção 5.1). A proposta foi testada nos conjuntos de imagens histológicas, conforme descrição apresentada na seção 4.1 da metodologia. As comparações realizadas foram: benigno e maligno para as bases UCSB e CR; saudável e severo para OED; macho e fêmea para a base LG. Por fim, uma visão geral dos desempenhos proporcionados pelas principais associações, frente aos resultados das redes aplicadas diretamente (seção 5.2) e dos estudos disponíveis na Literatura (seção 5.3), também é descrita.

5.1 Desempenhos das Combinações de Descritores

As combinações de descritores foram testadas conforme destalhes apresentados na seção 4.4. Assim, as médias das 10 melhores acurácias estão na Tabela 5.1, considerando como referência a maior acurácia via menor número de descritores para cada categoria de descritor. As maiores médias estão em negrito. Os detalhes dos experimentos que fundamentaram os 10 melhores resultados estão no apêndice A, com indicações das taxas de acurácias alcançadas a partir de cada combinação e em função de variações nos totais de descritores.

Tabela 5.1: Médias das acurácias (%) computadas a partir dos 10 melhores resultados em cada combinação e para cada tipo de imagem H&E.

	CR	**LG**	**OED**	**UCSB**
Handcrafted	84,48% ± 2,05	90,42% ± 5,23	87,03% ± 1,78	72,41% ± 1,89
Deep learned	99,27% ± 0,76	98,49% ± 0,65	96,28% ± 0,55	91,38% ± 1,54
IAx	83,82% ± 0,90	86,98% ± 2,54	80,88% ± 0,43	78,28% ± 1,14
Ensemble* de *handcrafted* e *Deep learned	99,76% ± 0,30	99,02% ± 0,51	96,49% ± 0,66	90,52% ± 0,86
Ensemble* de *deep learned	**100%**	**99,66% ± 0,11**	**97,23% ± 0,47**	**92,93% ± 0,93**
***Ensemble* de IAx**	86,18% ± 1,45	89,62% ± 0,42	78,85% ± 0,61	78,45% ± 1,16

A partir das médias, nota-se que *ensemble* de *deep learned* forneceu os maiores valores em todos os conjuntos de dados, com destaque para a base de imagens CR, com uma acurácia de 100%. Também, *ensemble* de *handcrafted* com *deep learned* foi a segunda associação com os melhores resultados, indicando que esta composição é outra importante estratégia para alcançar os melhores desempenhos. Mais ainda, os resultados via *ensemble* de *deep learned* confirmam a viabilidade do uso da estratégia *transfer learning* para a obtenção de descritores e classificação de imagens H&E. Por outro lado, vetores de características construídos com representações IAx apresentaram resultados menos expressivos, com acurácias variando de 78% a 89% (aproximadamente), mas indicando o potencial de informação existente neste tipo composição e que pode ser utilizada em outras frentes de investigação. Esses apontamentos são contribuições relevantes para fundamentar abordagens metodológicas nos contextos aqui explorados.

Como parte das análises dos resultados, o teste de Friedman foi aplicado para verificar se há diferença estatisticamente relevante entre as combinações. Neste teste, o resultado obtido, chamado *p-value*, foi comparado a uma variável $\alpha = 0{,}05$. Logo, se *p-value* $< \alpha$, a diferença foi considerada significativa. Como fator de análise, o teste foi efetuado observando as combinações de vetores que proporcionaram as médias das 10 melhores acurácias em cada uma das bases H&E, conforme Tabela 5.1. Assim, considerando uma comparação entre os seis tipos de vetores de características, o *p-value*

geral foi de 0,0033, permitindo inferir que existem diferenças estatisticamente importantes, especialmente do *ensemble* de *deep learned* contra as associações *handcrafted*, IAx e *ensemble* de IAx. Cada *p-value* par-a-par está na Tabela 5.2. Por outro lado, as diferenças entre as taxas de acurácias atingidas por meio das combinações *ensemble* de *deep learned* e *ensemble* de *handcrafted* com *deep learned* não são estatisticamente relevantes. Apesar disso, com a aplicação do teste de Friedman, foi possível obter um ranqueamento médio entre os desempenhos das associações aqui testadas. O resultado desta comparação apontou que a melhor associação foi definida por meio de *ensemble* de *deep learned*, sendo ranqueada em primeiro lugar dentre todas as combinações aqui investigadas. Neste tipo de combinação, as acurácias médias foram de 92,93% (UCSB) a 100% (CR), conforme já destacado na Tabela 5.1.

Tabela 5.2: Ranqueamento das melhores associações, segundo as acurácias médias e teste de Friedman, com os *p-values* correspondentes obtidos via comparações par-a-par entre as associações.

p-value	***Ensemble* de *deep learned***	**Ensemble de *handcrafted* e *Deep learned***	***Deep learned***	***Handcrafted***	***Ensemble* de IAx**	**IAx**	**Ranqueamento Médio**
Ensemble* de *deep learned	-	0,3955	0,2396	**0,0191**	**0,0191**	**0,0066**	1
Ensemble* de *handcrafted* e *Deep learned	0,3955	-	0,7313	0,1006	0,1006	**0,0380**	2,25
Deep learned	0,2396	0,7313	-	0,1819	0,1819	0,0735	2,75
Handcrafted	**0,0191**	0,1006	0,1819	**-**	1,0000	0,6073	4,75
***Ensemble* de IAx**	**0,0191**	0,1006	0,1819	1,0000	**-**	0,6073	4,75
IAx	**0,0066**	**0,0380**	0,0735	0,6073	0,6073	-	5.5

5.1.1 Detalhamento das *top* 10 soluções que fundamentaram os melhores resultados

Considerando que a melhor combinação foi definida pela maior acurácia com o menor número de descritores, nas Tabelas 5.3 a 5.6 estão listados, em ordem decrescente, os 10 melhores desempenhos que fundamentaram cada *ensemble* de *deep learned*. É importante observar que o ranqueamento médio das soluções ratifica essa situação.

Em relação aos dados coletados a partir da base CR (Tabela 5.3), é possível ob-

Tabela 5.3: Indicação dos 10 melhores resultados para classificar a base de imagens CR, via vetores constituídos do *ensemble* de *deep learned.*

Vetor	Tamanho	Acurácia (%)
D + E	10	100
E + V	10	100
D + E + I + R + V	10	100
D + E	15	100
D + I	15	100
E + V	15	100
D + E + I + R + V	15	100
D + E	20	100
D + I	20	100
D + V	20	100

Tabela 5.4: Apresentação dos 10 melhores resultados obtidos a partir da classificação da base de imagens LG, explorando vetores constituídos do *ensemble* de *deep learned.*

Vetor	Tamanho	Acurácia (%)
D + R	25	100
D + E	10	99,62
D + E	15	99,62
D + I	15	99,62
D + R	15	99,62
D + E	20	99,62
D + I	20	99,62
D + R	20	99,62
D + E	25	99,62
D + I	25	99,62

servar que os 10 melhores resultados indicaram uma acurácia de 100%. Nota-se que os descritores das redes DenseNet-121 (D) e EfficientNet-b2 (E) foram os que mais contribuíram para estes resultados. Nestes casos, os vetores foram constituídos por no máximo 20 descritores, minimizando a possibilidade de *overfitting*.

Quando o conjunto LG é considerado (Tabela 5.4), nota-se que somente uma combinação indicou o valor de acurácia de 100%: *ensemble* de *deep learned* das arquiteturas DenseNet-121 (D) e ResNet-50 (R), explorando 25 descritores. Para esta base de imagens, é visível a contribuição dos descritores da CNN DenseNet-121, uma vez que

Tabela 5.5: Indicação dos 10 melhores resultados na base de imagens OED, utilizando vetores constituídos do *ensemble* de *deep learned.*

Vetor	**Tamanho**	**Acurácia (%)**
I + V	20	97,97
E + I	25	97,97
D + E	20	97,30
D + I	20	97,30
I + R	20	97,30
D + R	25	97,30
I + V	25	97,30
D + I	10	96,62
D + E	15	96,62
I + V	15	96,62

Tabela 5.6: Apontamento dos 10 melhores resultados para classifica imagens da base UCSB, considerando vetores do *ensemble* de *deep learned.*

Vetor	**Tamanho**	**Acurácia (%)**
D + E	25	94,83
D + E	15	93,10
E + R	15	93,10
I + R	15	93,10
D + E	20	93,10
D + I	25	93,10
E + I	25	93,10
I + R	25	93,10
E + R	10	91,38
D + R	15	91,38

estes compõem todos os vetores que constam entre os melhores resultados, utilizando até 25 descritores.

Para as imagens OED (Tabela 5.5), a maior acurácia foi de 97,97% com o *ensemble* de descritores das redes Inception-V3 (I) e VGG-19 (V), utilizando 20 atributos *deep learned*. Outra combinação que atingiu o mesmo valor foi por meio dos descritores das redes EfficientNet-b2 (E) e Inception-V3 (I). Entretanto, neste último caso, foram necessários 25 atributos. Ao contrário do comportamento observado nas bases CR e LG, na base OED, a Inception-V3 (I) foi a rede que mais contribuiu para os melho-

res resultados, seguida pela DenseNet-121 (D). Os principais vetores também foram definidos com até 25 descritores.

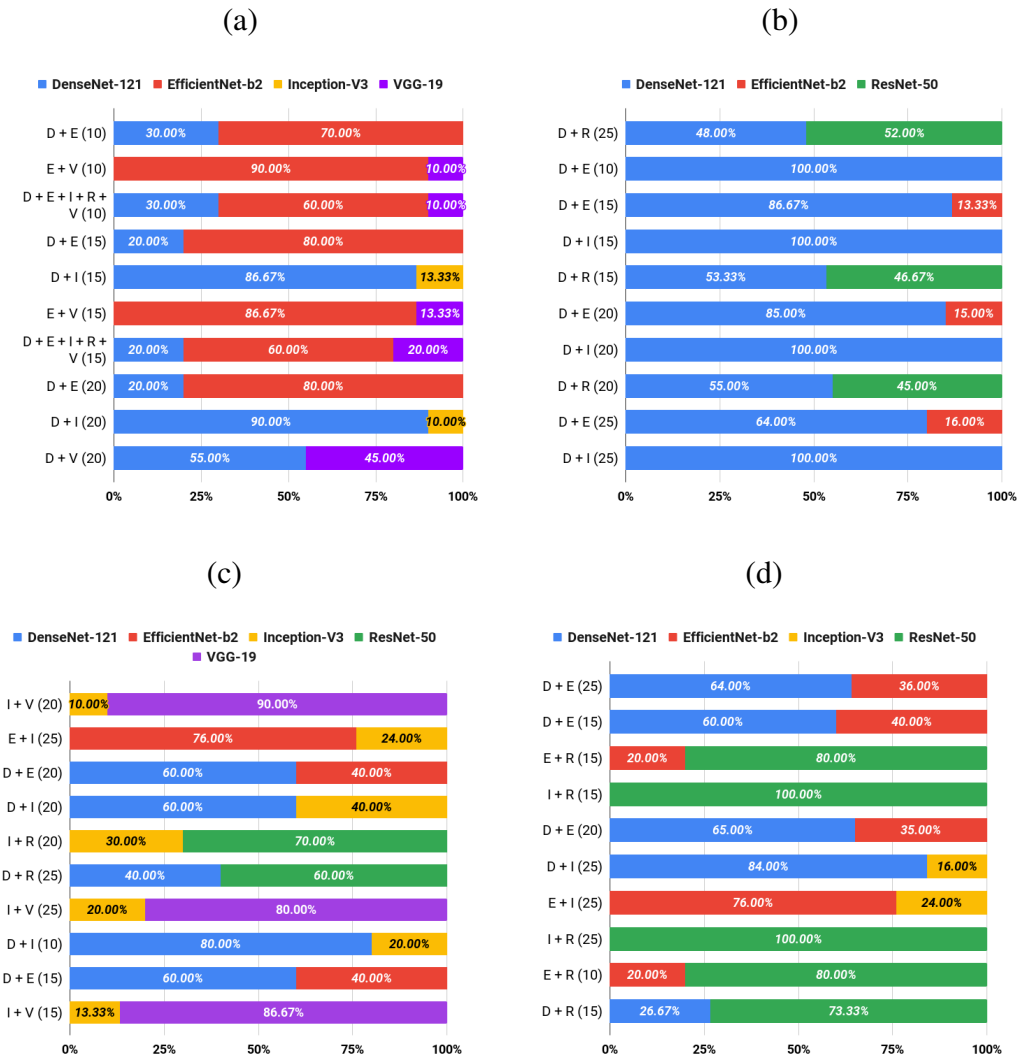
Por fim, considerando os valores na base UCSB, Tabela 5.6, nota-se que a maior taxa de acurácia foi de 94,83%. Este desempenho foi definido na associação de *deep learned* da DenseNet-121 (D) com EfficientNet-b2 (E), explorando um total de 25 descritores. Por outro lado, das 10 principais soluções, nota-se que nove composições foram definidas a partir de descritores das arquiteturas DenseNet-121 ou ResNet-50. O total de descritores presentes nos vetores segue o padrão encontrado previamente, soluções obtidas com no máximo 25 atributos.

5.1.2 Composições dos vetores que definem os melhores resultados

Outro ponto de vista interessante é observar a composição de cada *ensemble* de *deep learned* presente em cada lista de 10 *top* soluções. As composições dos vetores para cada base de imagens H&E estão na Figura 5.1.

Para a base de imagens CR (Figura 5.1a), observa-se que a maioria dos vetores tem a arquitetura EfficientNet-b2 (E) como origem, 7 das 10 combinações que contribuíram para os melhores resultados, com composições variando de 60% a 90% nos descritores. Para LG (Figura 5.1b), os vetores obtidos da arquitetura DenseNet-121 (D) predominam (9 em um total 10 vetores). Nestes casos, as composições nos vetores foram de 48% a 100% dos descritores. Já para OED (Figura 5.1c), ao contrário das anteriores, é possível notar uma frequência maior de descritores da Inception-V3 (I), presentes em 7 dos 10 vetores. Entretanto, os percentuais de descritores nessas composições foram menores dos destacados previamente, variando de 10% a 40%. Por fim, para UCSB (Figura 5.1d), apesar de uma distribuição mais homogênea entre DenseNet-121 (D), EfficientNet-b2 (E) e ResNet-50 (R), as composições de vetores mais significativas foram a partir de descritores da ResNet-50 (R), de 73% a 100% dos atributos.

Figura 5.1: Proporção do número de descritores selecionados em cada composição, observando os 10 melhores resultados que fundamentaram as principais soluções para as bases CR (a), LG (b), OED (c) e UCSB (d).



Fonte: Elaborado pelo Autor

Com base nas observações realizadas, foi possível indicar alguns padrões e/ou comportamentos obtidos com a aplicação da técnica de *ensemble* de descritores no contexto deste trabalho. Nota-se que os *ensembles* contribuíram efetivamente para distinções dos grupos de imagens H&E, superando os resultados fornecidos via descritores de uma única origem. O *ensemble* de *deep learned* foi a principal associação, com destaques, em termos de ocorrência, para os descritores *deep learned* extraídos das redes DenseNet-121 (em 28 de 40 vetores) e EfficientNet-b2 (em 19 de 40 veto-

res). Por fim, considerando as dimensões dos vetores de características, composições envolvendo de 15 a 25 descritores *deep learned* foram suficientes para determinar os melhores resultados.

5.2 Comparação com os desempenhos via aplicação direta das CNNs

Para verificar possíveis ganhos da nossa proposta em relação aos modelos de redes aplicados diretamente sobre cada conjunto de imagens H&E, os desempenhos das arquiteturas (DenseNet-121 (HUANG et al., 2017), EfficientNet-b2 (TAN; LE, 2019), Inception-V3 (SZEGEDY et al., 2016), ResNet-50 (HE et al., 2016) e VGG-19 (SIMONYAN; ZISSERMAN, 2014)) foram coletados após o refinamento da camada final. Os detalhes metodológicos foram apresentados na seção 4.2 e a ilustração de cada experimento está no apêndice B, sumarizados nas Figuras B.1 a B.5. Para fins comparativos, nas Tabelas 5.7 a 5.10 estão as melhores taxas de acurácias fornecidas pelas redes e os melhores resultados da nossa abordagem. É importante ressaltar que outros testes poderiam ser explorados, envolvendo diferentes condições e limites, mas acreditamos que os expostos aqui fornecem informações suficientes para fundamentar as contribuições das nossas investigações em diferentes bases de imagens H&E.

Tabela 5.7: Acurácias (%) ranqueadas dos resultados de classificação da base de imagens CR por diferentes métodos

	Acurácia (%)
***Ensemble* de *deep learned* (D + E)**	100%
ResNet-50	91,26%
VGG-19	86,52%
EfficientNet-b2	80,35%
Inception-V3	77,74%
DenseNet-121	67,81%

Quando o processo direto de classificação é considerado em cada modelo de CNN,

Tabela 5.8: Acurácias (%) ranqueadas dos resultados de classificação da base de imagens LG por diferentes métodos

	Acurácia (%)
***Ensemble* de *deep learned* (D + R)**	100%
VGG-19	88,29%
ResNet-50	71,94%
DenseNet-121	68,39%
Inception-V3	61,79%
EfficientNet-b2	54,35%

Tabela 5.9: Acurácias (%) ranqueadas dos resultados de classificação da base de imagens OED por diferentes métodos

	Acurácia (%)
***Ensemble* de *deep learned* (I + V)**	97,97%
VGG-19	91,59%
ResNet-50	79,89%
Inception-V3	75,19%
DenseNet-121	72,91%
EfficientNet-b2	66,35%

Tabela 5.10: Acurácias (%) ranqueadas dos resultados de classificação da base de imagens UCSB por diferentes métodos

	Acurácia (%)
***Ensemble* de *deep learned* (D + E)**	94,83%
ResNet-50	88,52%
DenseNet-121	79,26%
VGG-19	77,04%
EfficientNet-b2	77,04%
Inception-V3	60,74%

nota-se que as arquiteturas DenseNet-121 e EfficientNet-b2 indicaram os menores valores de acurácias, variando de 67,81% a 79,26% e de 54,35% a 80,35%, respectivamente. Neste caso, as redes VGG-19, para as bases de imagem LG (88,29%) e OED (97,59%), e ResNet-50, para as bases CR (91,26%) e UCSB (88,52%), forneceram os maiores valores de acurácias. Contudo, estas taxas são menores que as conquistadas via abordagens *ensemble* de *deep learned*. Também, por meio das in-

formações aqui detalhadas, observou-se que o desempenho de classificação com um modelo CNN não está diretamente relacionado com a garantia de bom desempenho via análises independentes dos descritores *deep learned* correspondentes, uma vez que a maioria dos melhores descritores foram de combinações com as redes DenseNet-121 e EfficientNet-b2. Estas arquiteturas forneceram as menores taxas de desempenhos nos contextos de imagens aqui explorados. Estes resultados ratificam as contribuições obtidas neste estudo.

5.3 Visão ilustrativa de desempenhos

Diferentes técnicas estão disponíveis para investigar padrões em imagens histológicas, tais como para as bases CR, LG, OED e UCSB (KAUSAR et al., 2019; NANNI et al., 2019; TAVOLARA et al., 2019; YU et al., 2019; ROBERTO et al., 2021; SILVA et al., 2022). Assim, uma visão geral ilustrativa é importante para mostrar a qualidade da nossa proposta, indicada nas Tabelas 5.11 a 5.14 para cada uma das bases.

Tabela 5.11: Acurácias (%) em conjuntos de imagens colorretais a partir de diferentes abordagens na literatura.

Método	**Abordagem**	**Acc (%)**
Proposto	DenseNet-121 e EfficientNet-b2 (*Ensemble de deep learned*)	100%
ROBERTO et al. (2021)	ResNet-50, Dimensão Fractal, Lacunaridade e Percolação (*Ensemble de deep learned + Handcrafted*)	99,39%
DABASS; VIG; VASHISTH (2019)	CNN com 31 camadas (*Deep Learning*)	96,97%
TAVOLARA et al. (2019)	GAN e U-Net (*Deep Learning*)	94,02%
SENA et al. (2019)	CNN com 12 camadas (*Deep Learning*)	93,28%
SANTOS (2018)	Entropia amostral e lógica fuzzy (*Handcrafted*)	91,39%
ROBERTO et al. (2019)	Percolação (*Handcrafted*)	90,90%
BENTAIEB; HAMARNEH (2018)	U-Net e AlexNet (*Deep Learning*)	87,50%
AWAN et al. (2020)	Normalização de cores, U-Net e GoogLeNet (*Deep Learning*)	85,00%

A partir dessa visão ilustrativa, é possível verificar que a proposta conseguiu fornecer soluções com resultados dentre os relatados na literatura para classificar padrões em imagens histológicas H&E. Nota-se ainda que o *ensemble* de *deep learned* forneceram resultados que superaram um único tipo de descritor ou outros tipos de associações (ANDREARCZYK; WHELAN, 2017; PAPASTERGIOU et al., 2018; KAUSAR et

Tabela 5.12: Acurácias (%) em imagens histológicas de tecido hepático considerando distintas abordagens disponíveis na literatura.

Método	**Abordagem**	**Acc (%)**
Proposto	DenseNet-121 e ResNet-50 (*Ensemble de deep learned*)	100%
RUBERTO et al. (2016)	Análise Estatísticas e Descritores de Textura (*Handcrafted*)	100%
NANNI et al. (2019)	6 modelos de CNN e descritores *handcrafted* (*Ensemble de deep learned* + *Handcrafted*)	100%
ROBERTO et al. (2021)	ResNet-50, Dimensão Fractal, Lacunaridade e Percolação (*Ensemble de deep learned* + *Handcrafted*)	99,62%
ANDREARCZYK; WHELAN (2017)	*Texture* CNN (*Deep Learning*)	99,10%
WATANABE; KOBAYASHI; WADA (2016)	Descritores GIST, PCA e LDA (*Handcrafted*)	93,70%

Tabela 5.13: Acurácias (%) em conjuntos de imagens de displasia oral a partir de diferentes abordagens na literatura.

Método	**Abordagem**	**Acc (%)**
Proposto	Inception-V3 e VGG-19 (*Ensemble de deep learned*)	97,97%
ADEL et al. (2018)	SIFT, SURF, ORB (*Handcrafted*)	92,80%
SILVA et al. (2022)	Descritores morfológicos e não-morfológicos (*Handcrafted*)	92,40%
KRISHNAN et al. (2011)	Dimensão Fractal, *wavelet*, Movimento Browniano e Filtros de Gabor (*Handcrafted*)	88,38%

Tabela 5.14: Acurácias (%) em imagens histológicas de mama considerando diferentes métodos disponíveis na literatura.

Método	**Abordagem**	**Acc (%)**
LI et al. (2019b)	RefineNet e Atrous DenseNet (Deep Learning)	97,63%
YU et al. (2019)	CNN, LBP, SURF, GLCM e outros descritores handcrafted (Ensemble de deep learned + Handcrafted)	96,67%
Proposto	DenseNet-121 e EfficientNet-b2 (*Ensemble de deep learned*)	94,83%
FENG; ZHANG; YI (2018)	Stacked denoising autoencoder (Deep Learning)	94,41%
KAUSAR et al. (2019)	Normalização de cor, decomposição de wavelet de Haar e CNN com 16 camadas (Deep Learning)	91,00%
ROBERTO et al. (2021)	ResNet-50, Dimensão Fractal, Lacunaridade e Percolação (Ensemble de deep learned + Handcrafted)	89,66%
ROBERTO et al. (2019)	Percolação (Handcrafted)	86,20%
PAPASTERGIOU et al. (2018)	Decomposição espacial e tensors (Deep Learning)	84,67%
ARAÚJO et al. (2017)	Nomalização de cor, CNN com 13 camadas e SVM (Deep Learning)	83,30%

al., 2019; TAVOLARA et al., 2019; SILVA et al., 2022). Mais ainda, o uso combinado de *deep learned* por transferência de aprendizado e um *ensemble* de classificadores permitiu identificar soluções relevantes para cada conjunto de imagens, complementando os trabalhos aqui explorados nessa visão ilustrativa.

Capítulo 6

Conclusão

Neste trabalho, foram exploradas múltiplas técnicas de *ensemble* de descritores e classificadores para verificar os desempenhos de diferentes descritores fractais, descritores *deep learned* com origens distintas, e quantificações a partir de representações IAx. No contexto explorado, os autores (YU et al., 2019; NANNI et al., 2019; ROBERTO et al., 2021; DAS; BISWAS, 2019; BRUNESE et al., 2020; KUNCHEVA, 2014) também investigaram *deep features* de múltiplos modelos de CNNs, inclusive desenvolvendo suas próprias arquiteturas, com *handcrafted* de diferentes origens, tais como descritores morfológicos e medidas fundamentadas em percolação. Apesar de avanços importantes proporcionados por esses estudos, consideramos que a investigação aqui desenvolvida apontou outros destaques, como o uso de uma etapa de ranqueamento para selecionar os atributos mais relevantes para o processo de classificação, garantindo resultados, no mínimo, compatíveis com a Literatura especializada.

Diferentemente de outras abordagens que utilizaram apenas técnicas de *deep learning* (ANDREARCZYK; WHELAN, 2017; BENTAIEB; HAMARNEH, 2018; PAPASTERGIOU et al., 2018; SENA et al., 2019) ou apenas técnicas *handcrafted* (WATANABE; KOBAYASHI; WADA, 2016; SANTOS; PONTI, 2019; ROBERTO et al., 2019; SILVA et al., 2022), mostramos que o uso combinado de técnicas de *ensemble* de descritores e classificadores forneceram distinções relevantes em contextos de

imagens histológicas H&E, complementando, inclusive, importantes abordagens disponíveis na Literatura. Também, foi possível constatar a contribuição do processo de transferência de aprendizado, uma vez que os descritores *deep learned* definiram os melhores resultados. Notou-se, ainda, distinções expressivas via outros *ensembles*, incluindo o composto de descritores *handcrafted* e *deep learned*. Mesmo que essa combinação não tenha superado os valores conquistados com *ensembles* compostos apenas por descritores *deep learned*, a associação de atributos *handcrafted* com *deep learned* indicou desempenhos melhores do que os conquistados em suas análises individualizadas.

Em relação aos experimentos explorando representações IAx, é importante evidenciar resultados menos expressivos, com médias de acurácias de 78% a 89%, aproximadamente. Por outro lado, foi possível constatar que este tipo de composição forneceu resultados similares aos das redes aplicadas diretamente, responsáveis por fornecer as representações IAx. Diante deste fato, acreditamos que ainda há várias frentes de estudos para conhecer a capacidade plena de informação presente neste tipo de representação, buscando aprimorar projetos de sistemas de apoio ao diagnóstico para imagens H&E.

Por fim, nosso trabalho possibilitou analisar algumas situações ainda pertinentes: indicou possíveis padrões de combinações de técnicas para diferentes tipos de imagens H&E; mostrou que descritores fractais multiescalas e multidimensionais permitem aumentar os desempenhos dos resultados conquistados via *deep features* com *transfer learning* em imagens H&E; e apontou que a associação de *ensembles* (descritores e classificadores) forneceu resultados competitivos em relação aos trabalhos correlatos, com a vantagem de usar um número reduzido de descritores.

6.1 Trabalhos Futuros

É possível obter avanços em relação aos modelos aqui investigados, visando complementar ou aprimorar as principais combinações. Por exemplo, é interessante ampliar o escopo de técnicas *handcrafted*, especialmente para conhecer possíveis limites envolvendo as representações IAx. Abordagens *multi-view learning* é outra frente que pode ser utilizada para explorar as múltiplas representações, inclusive com investigações sobre possíveis ganhos a partir do enriquecimento de *deep features* em uma CNN. Novas arquiteturas de CNN podem também contribuir neste ponto, bem como ajustar os parâmetros empregados na fase de refinamento.

6.2 Contribuições

O desenvolvimento deste trabalho permitiu publicações relacionadas direta e indiretamente com o tema, a saber:

- Artigo *Colour feature extraction and polynomial algorithm for classification of lymphoma images*. *Iberoamerican Congress on Pattern Recognition* (pp. 262-271). Springer, Cham, 2019 (MARTINS et al., 2019);

- Artigo *Analysis of cancer in histological images: employing an approach based on genetic algorithm*. *Pattern Analysis and Applications*, 24(2), 483-496, 2021 (TAINO et al., 2021);

- Artigo *A Hermite polynomial algorithm for detection of lesions in lymphoma images*. *Pattern Analysis and Applications*, 24(2), 523-535, 2021 (MARTINS et al., 2021);

- Artigo *Ensembles of fractal descriptors with multiple deep learned features for classification of histological images*. Em *29th International Conference on Sys-*

tems, Signals and Image Processing (IWSSIP) (pp. 1-4). IEEE. 2022 (LONGO et al., 2022);

- Artigo *Classification of H&E images exploring ensemble learning with two-stage feature selection*. Em *29th International Conference on Systems, Signals and Image Processing* (IWSSIP) (pp. 1-4). IEEE. 2022 (TENGUAM et al., 2022);

- Artigo *Multidimensional shannon entropy (HM) as an approach to classify H&E colorectal images*. Em *29th International Conference on Systems, Signals and Image Processing* (IWSSIP) (pp. 1-4). IEEE. 2022 (SANTOS et al., 2022);

- Artigo *Percolation Features: An approach for evaluating fractal properties in colour images*. *Software Impacts*, 14, 100387, 2022 (ROBERTO et al., 2022).

Referências

ADEL, D.; MOUNIR, J.; EL-SHAFEY, 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.

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.

AL-HAIJA, Q. A.; ADEBANJO, A. Breast cancer diagnosis in histopathological images using resnet-50 convolutional neural network. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). [S.l.: s.n.], 2020. p. 1–7.

ALICIOGLU, G.; SUN, B. A survey of visual analytics for explainable artificial intelligence methods. Computers & Graphics, Elsevier, v. 102, p. 502–520, 2022.

ALINSAIF, S.; LANG, J. Histological image classification using deep features and transfer learning. In: 2020 17th Conference on Computer and Robot Vision (CRV). [S.l.: s.n.], 2020. p. 101–108.

ALLAIN, C.; CLOITRE, M. Characterizing the lacunarity of random and deterministic fractal sets. Physical review A, APS, v. 44, n. 6, p. 3552, 1991.

ALMARAZ-DAMIAN, J.-A.; PONOMARYOV, V.; SADOVNYCHIY, S.; CASTILLEJOS-FERNANDEZ, H. Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. Entropy, Multidisciplinary Digital Publishing Institute, v. 22, n. 4, p. 484, 2020.

ANDREARCZYK, V.; WHELAN, P. F. Deep learning for biomedical texture image analysis. In: IRISH PATTERN RECOGNITION & CLASSIFICATION SOCIETY (IPRCS). Proceedings of the Irish Machine Vision & Image Processing Conference. [S.l.], 2017.

ARALICA, G.; IVELJ, M. Š.; PAČIĆ, A.; BAKOVIĆ, J.; PERIŠA, M. M.; KRIŠTIĆ, A.; KONJEVODA, P. Prognostic significance of lacunarity in preoperative biopsy of colorectal cancer. Pathology & Oncology Research, Springer, v. 26, n. 4, p. 2567–2576, 2020.

ARAÚJO, T.; ARESTA, G.; CASTRO, E.; ROUCO, J.; AGUIAR, P.; ELOY, C.; POLÓNIA, A.; CAMPILHO, A. Classification of breast cancer histology images using convolutional neural networks. PloS one, Public Library of Science San Francisco, CA USA, v. 12, n. 6, p. e0177544, 2017.

ARJMAND, A.; TSAKAI, O.; CHRISTOU, V.; TZALLAS, A. T.; TSIPOURAS, M. G.; FORLANO, R.; MANOUSOU, P.; GOLDIN, R. D.; GOGOS, C.; GLAVAS, E. et al. Ensemble convolutional neural network classification for pancreatic steatosis assessment in biopsy images. Information, MDPI, v. 13, n. 4, p. 160, 2022.

AWAN, R.; AL-MAADEED, S.; AL-SAADY, R.; BOURIDANE, A. Glandular structure-guided classification of microscopic colorectal images using deep learning. Computers & Electrical Engineering, Elsevier, v. 85, p. 106450, 2020. ISSN 0045-7906. Disponível em: https://www.sciencedirect.com/science/article/pii/S0045790618332087.

BENTAIEB, A.; HAMARNEH, G. Adversarial stain transfer for histopathology image analysis. IEEE Transactions on Medical Imaging, v. 37, n. 3, p. 792–802, March 2018. ISSN 1558-254X.

BISHOP, C. M. Pattern recognition and machine learning. [S.l.]: springer, 2006.

BREIMAN, L. Bagging predictors. Machine learning, Springer, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L. Random forests. Machine learning, Springer, v. 45, n. 1, p. 5–32, 2001.

BROADBENT, S. R.; HAMMERSLEY, J. M. Percolation processes: I. crystals and mazes. In: CAMBRIDGE UNIVERSITY PRESS. Mathematical proceedings of the Cambridge philosophical society. [S.l.], 1957. v. 53, n. 3, p. 629–641.

BRUNESE, L.; MERCALDO, F.; REGINELLI, A.; SANTONE, A. An ensemble learning approach for brain cancer detection exploiting radiomic features. Computer Methods and Programs in Biomedicine, v. 185, p. 105134, 2020. ISSN 0169-2607.

CĂLIMAN, A.; IVANOVICI, M. Psoriasis image analysis using color lacunarity. In: IEEE. Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on. [S.l.], 2012. p. 1401–1406.

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.

CANTRELL, C. D. Modern mathematical methods for physicists and engineers. [S.l.]: Cambridge University Press, 2000.

CARLETON, H. M.; DRURY, R. A. B.; WALLINGTON, E. A. Carleton's histological technique. [S.l.]: Oxford University Press, USA, 1980.

CHAN, J. K. C. The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology. International Journal of Surgical Pathology, v. 22, n. 1, p. 12–32, 2014. PMID: 24406626. Disponível em: https://doi.org/10.1177/1066896913517939.

CHEN, Y.; WANG, Y.; GU, Y.; HE, X.; GHAMISI, P.; JIA, X. Deep learning ensemble for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE, v. 12, n. 6, p. 1882–1897, 2019.

CORTES, C.; VAPNIK, V. Support-vector networks. Machine learning, Springer, v. 20, n. 3, p. 273–297, 1995.

CUNNINGHAM, P.; DELANY, S. J. k-nearest neighbour classifiers–. arXiv preprint arXiv:2004.04523, 2020.

DABASS, M.; VIG, R.; VASHISTH, S. Five-grade cancer classification of colon histology images via deep learning. In: Communication and Computing Systems. [S.l.]: CRC Press, 2019. p. 18–24.

DAS, S.; BISWAS, D. Prediction of breast cancer using ensemble learning. In: 2019 5th International Conference on Advances in Electrical Engineering (ICAEE). [S.l.: s.n.], 2019. p. 804–808.

DASIGI, V.; MANN, R. C.; PROTOPOPESCU, V. A. Information fusion for text classification — an experimental comparison. Pattern Recognition, v. 34, n. 12, p. 2413 – 2425, 2001. ISSN 0031-3203. Disponível em: http://www.sciencedirect.com/science/article/pii/S0031320300001710.

DEMIR, S.; KEY, S.; BAYGIN, M.; TUNCER, T.; DOGAN, S.; BELHAOUARI, S. B.; POYRAZ, A. K.; GURGER, M. Automated knee ligament injuries classification method based on exemplar pyramid local binary pattern feature extraction and hybrid iterative feature selection. Biomedical Signal Processing and Control, Elsevier, v. 71, p. 103191, 2022.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2009. p. 248–255.

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. International workshop on multiple classifier systems. [S.l.], 2000. p. 1–15.

FENG, Y.; ZHANG, L.; YI, Z. Breast cancer cell nuclei classification in histopathology images using deep neural networks. International journal of computer assisted radiology and surgery, Springer, v. 13, n. 2, p. 179–191, 2018.

FONSECA-SILVA, T.; DINIZ, M. G.; SOUSA, S. F. de; GOMEZ, R. S.; GOMES, C. C. Association between histopathological features of dysplasia in oral leukoplakia and loss of heterozygosity. Histopathology, Wiley Online Library, v. 68, n. 3, p. 456–460, 2016.

FORTIN, C. S.; KUMARESAN, R.; OHLEY, W. J.; HOEFER, S. Fractal dimension in the analysis of medical images. IEEE Engineering in Medicine and Biology Magazine, v. 11, n. 2, p. 65–71, 1992.

FRANK, E.; HALL, M. A.; HOLMES, G.; KIRKBY, R.; PFAHRINGER, B.; WITTEN, I. H. Weka: A machine learning workbench for data mining. In: ____. Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers. Berlin: Springer, 2005. p. 1305–1314. Disponível em: http://researchcommons.waikato.ac.nz/handle/10289/1497.

FREUND, Y.; SCHAPIRE, R. E. et al. Experiments with a new boosting algorithm. In: CITESEER. icml. [S.l.], 1996. v. 96, p. 148–156.

GANGULY, A.; DAS, R.; SETUA, S. K. Histopathological image and lymphoma image classification using customized deep learning models and different optimization algorithms. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). [S.l.: s.n.], 2020. p. 1–7.

GELASCA, E. D.; BYUN, J.; OBARA, B.; MANJUNATH, B. Evaluation and benchmark for biological image segmentation. In: IEEE International Conference on Image Processing. [S.l.: s.n.], 2008.

GERON, A. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn TensorFlow. [S.l.]: Rio de Janeiro: Alta Books, 2019.

GHOSH, P.; AZAM, S.; JONKMAN, M.; KARIM, A.; SHAMRAT, F. J. M.; IGNATIOUS, E.; SHULTANA, S.; BEERAVOLU, A. R.; BOER, F. D. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. IEEE Access, IEEE, v. 9, p. 19304–19326, 2021.

GILDENBLAT, J.; CONTRIBUTORS. PyTorch library for CAM methods. [S.l.]: GitHub, 2021. https://github.com/jacobgil/pytorch-grad-cam.

GONZALEZ, R. C.; WOODS, R. C. Processamento digital de imagens . [S.l.]: Pearson Educación, 2009.

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

GUNNING, D.; AHA, D. Darpa's explainable artificial intelligence (xai) program. AI magazine, v. 40, n. 2, p. 44–58, 2019.

GUPTA, M.; GUPTA, B. An ensemble model for breast cancer prediction using sequential least squares programming method (slsqp). In: 2018 Eleventh International Conference on Contemporary Computing (IC3). [S.l.: s.n.], 2018. p. 1–3.

HAGERTY, J. R.; STANLEY, R. J.; ALMUBARAK, H. A.; LAMA, N.; KASMI, R.; GUO, P.; DRUGGE, R. J.; RABINOVITZ, H. S.; OLIVIERO, M.; STOECKER, W. V. Deep learning and handcrafted method fusion: higher diagnostic accuracy for

melanoma dermoscopy images. IEEE journal of biomedical and health informatics, IEEE, v. 23, n. 4, p. 1385–1391, 2019.

HASAN, A. M.; JALAB, H. A.; MEZIANE, F.; KAHTAN, H.; AL-AHMAD, A. S. Combining deep and handcrafted image features for mri brain scan classification. IEEE Access, v. 7, p. 79959–79967, 2019.

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.

HOLZINGER, A.; SARANTI, A.; MOLNAR, C.; BIECEK, P.; SAMEK, W. Explainable ai methods-a brief overview. In: SPRINGER. International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers. [S.l.], 2022. p. 13–38.

HOSHEN, J.; KOPELMAN, R. Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm. Physical Review B, APS, v. 14, n. 8, p. 3438, 1976.

HUA, J.; XIONG, Z.; LOWEY, J.; SUH, E.; DOUGHERTY, E. R. Optimal number of features as a function of sample size for various classification rules. Bioinformatics, v. 21, n. 8, p. 1509–1515, 11 2004. ISSN 1367-4803.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2017. p. 4700–4708.

ISLAM, M. R.; NAHIDUZZAMAN, M. Complex features extraction with deep learning model for the detection of covid19 from ct scan images using ensemble based machine learning approach. Expert Systems with Applications, v. 195, p. 116554, 2022. ISSN 0957-4174. Disponível em: https://www.sciencedirect.com/science/article/pii/S0957417422000537.

IVANOVICI, M.; RICHARD, N. Fractal dimension of color fractal images. IEEE Transactions on Image Processing, IEEE, v. 20, n. 1, p. 227–235, 2010.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 22, n. 1, p. 4–37, 2000.

KANG, J.; ULLAH, Z.; GWAK, J. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors, Multidisciplinary Digital Publishing Institute, v. 21, n. 6, p. 2222, 2021.

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, v. 39, n. 4, p. 967–982, 2019. ISSN 0208-5216. Disponível em: https://www.sciencedirect.com/science/article/pii/S0208521619301457.

KHAN, A.; SOHAIL, A.; ZAHOORA, U.; QURESHI, A. S. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, Springer, p. 1–62, 2020.

KHEDKAR, S.; SUBRAMANIAN, V.; SHINDE, G.; GANDHI, P. Explainable ai in healthcare. In: Healthcare (April 8, 2019). 2nd International Conference on Advances in Science & Technology (ICAST). [S.l.: s.n.], 2019.

KILICARSLAN, S.; ADEM, K.; CELIK, M. Diagnosis and classification of cancer using hybrid model based on relieff and convolutional neural network. Medical Hypotheses, Elsevier, v. 137, p. 109577, 2020.

KIRA, K.; RENDELL, L. A. A practical approach to feature selection. In: Machine Learning Proceedings 1992. [S.l.]: Elsevier, 1992. p. 249–256.

KOC, M.; SUT, S. K.; SERHATLIOGLU, I.; BAYGIN, M.; TUNCER, T. Automatic prostate cancer detection model based on ensemble vggnet feature generation and nca feature selection using magnetic resonance images. Multimedia Tools and Applications, Springer, v. 81, n. 5, p. 7125–7144, 2022.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. Ijcai. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

KONONENKO, I.; ŠIMEC, E.; ROBNIK-ŠIKONJA, M. Overcoming the myopia of inductive learning algorithms with relieff. Applied Intelligence, Springer, v. 7, n. 1, p. 39–55, 1997.

KRISHNAN, M. M. R.; SHAH, P.; CHOUDHARY, A.; CHAKRABORTY, C.; PAUL, R. R.; RAY, A. K. Textural characterization of histopathological images for oral submucous fibrosis detection. Tissue and Cell, Elsevier, v. 43, n. 5, p. 318–330, 2011.

KUMAR, G.; BHATIA, P. K. A detailed review of feature extraction in image processing systems. In: 2014 Fourth International Conference on Advanced Computing Communication Technologies. [S.l.: s.n.], 2014. p. 5–12.

KUMAR, N.; SHARMA, M.; SINGH, V. P.; MADAN, C.; MEHANDIA, S. An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. Biomedical Signal Processing and Control, Elsevier, v. 75, p. 103596, 2022.

KUMAR, V.; ABBAS, A. K.; FAUSTO, N.; MITCHELL, R. N. Robbins patologia básica. [S.l.]: Elsevier Brasil, 2008.

KUMAR, V.; ABBAS, A. K.; FAUSTO, N. R. Cotran–patologia–bases patológicas das doenças. São Paulo: Elselvier, 2005.

KUNCHEVA, L. I. Combining pattern classifiers: methods and algorithms. [S.l.]: John Wiley & Sons, 2014.

LE, T. T.; URBANOWICZ, R. J.; MOORE, J. H.; MCKINNEY, B. A. STatistical Inference Relief (STIR) feature selection. Bioinformatics, v. 35, n. 8, p. 1358–1365, 09 2018. ISSN 1367-4803. Disponível em: https://doi.org/10.1093/bioinformatics/bty788.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. nature, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BENGIO, Y. et al. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, v. 3361, n. 10, p. 1995, 1995.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, Ieee, v. 86, n. 11, p. 2278–2324, 1998.

LEI, J.; SONG, X.; SUN, L.; SONG, M.; LI, N.; CHEN, C. Learning deep classifiers with deep features. In: 2016 IEEE International Conference on Multimedia and Expo (ICME). [S.l.: s.n.], 2016. p. 1–6.

LENAIL, A. Nn-svg: Publication-ready neural network architecture schematics. J. Open Source Softw., v. 4, n. 33, p. 747, 2019.

LI, S.; XU, P.; LI, B.; CHEN, L.; ZHOU, Z.; HAO, H.; DUAN, Y.; FOLKERT, M.; MA, J.; HUANG, S. et al. Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. Physics in Medicine & Biology, IOP Publishing, v. 64, n. 17, p. 175012, 2019.

LI, Y.; XIE, X.; SHEN, L.; LIU, S. Reverse active learning based atrous densenet for pathological image classification. BMC bioinformatics, Springer, v. 20, n. 1, p. 1–15, 2019.

LIMA, C. A. de M. Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte. Tese (Doutorado) — Universidade Estadual de Campinas, 2005.

LONGO, L. H. D. C.; MARTINS, A. S.; NASCIMENTO, M. Z. D.; SANTOS, L. F. S. D.; ROBERTO, G. F.; NEVES, L. A. Ensembles of fractal descriptors with multiple deep learned features for classification of histological images. In: 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP). [S.l.: s.n.], 2022. CFP2255E-ART, p. 1–4.

LUMINI, A.; NANNI, L. Convolutional neural networks for atc classification. Current pharmaceutical design, Bentham Science Publishers, v. 24, n. 34, p. 4007–4012, 2018.

MAHBOD, A.; SCHAEFER, G.; ELLINGER, I.; ECKER, R.; PITIOT, A.; WANG, C. Fusing fine-tuned deep features for skin lesion classification. Computerized Medical Imaging and Graphics, Elsevier, v. 71, p. 19–29, 2019.

MANDELBROT, B. B.; PIGNONI, R. The fractal geometry of nature. [S.l.]: WH freeman New York, 1983. v. 173.

MARTINEZ, E.; LOUZADA, F.; PEREIRA, B. A curva roc para testes diagnósticos. Cad Saúde Coletiva, v. 11, p. 7–31, 01 2003.

MARTINS, A. S.; NEVES, L. A.; FARIA, P. R.; TOSTA, T. A.; BRUNO, D. O.; LONGO, L. C.; NASCIMENTO, M. Z. do. Colour feature extraction and polynomial algorithm for classification of lymphoma images. In: SPRINGER. Iberoamerican Congress on Pattern Recognition. [S.l.], 2019. p. 262–271.

MARTINS, A. S.; NEVES, L. A.; FARIA, P. R. de; TOSTA, T. A.; LONGO, L. C.; SILVA, A. B.; ROBERTO, G. F.; NASCIMENTO, M. Z. do. A hermite polynomial algorithm for detection of lesions in lymphoma images. Pattern Analysis and Applications, Springer, v. 24, n. 2, p. 523–535, 2021.

MATLAB. version 9.7.0 (R2019b). Natick, Massachusetts: The MathWorks Inc., 2019.

MUHAMMAD, U.; WANG, W.; HADID, A. Feature fusion with deep supervision for remote-sensing image scene classification. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). [S.l.: s.n.], 2018. p. 249–253.

NANNI, L.; BRAHNAM, S.; GHIDONI, S.; MAGUOLO, G. General purpose (genp) bioimage ensemble of handcrafted and learned features with data augmentation. arXiv preprint arXiv:1904.08084, 2019.

NANNI, L.; GHIDONI, S.; BRAHNAM, S. Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognition, Elsevier, v. 71, p. 158–172, 2017.

NANNI, L.; GHIDONI, S.; BRAHNAM, S.; LIU, S.; ZHANG, L. Ensemble of handcrafted and deep learned features for cervical cell classification. In: NANNI, L.; BRAHNAM, S.; BRATTIN, R.; GHIDONI, S.; JAIN, L. (Ed.). Deep Learners and Deep Learner Descriptors for Medical Applications. Intelligent Systems Reference Library. [S.l.]: Springer, 2020. v. 186, p. 117–135.

NANNI, L.; LUMINI, A.; GHIDONI, S. Ensemble of deep learned features for melanoma classification. arXiv preprint arXiv:1807.08008, 2018.

NANNI, L.; LUMINI, A.; ZAFFONATO, N. Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment. Journal of neuroscience methods, Elsevier, v. 302, p. 42–46, 2018.

NIKOLAIDIS, N.; NIKOLAIDIS, I. N.; TSOUROS, C. C. A variation of the box-counting algorithm applied to colour images. arXiv preprint arXiv:1107.2336, 2011.

NIXON, M.; AGUADO, A. Feature extraction and image processing for computer vision. [S.l.]: Academic press, 2019.

OHATA, E. F.; CHAGAS, J. V. S. d.; BEZERRA, G. M.; HASSAN, M. M.; ALBUQUERQUE, V. H. C. de et al. A novel transfer learning approach for the classification of histological images of colorectal cancer. The Journal of Supercomputing, Springer, v. 77, n. 9, p. 9494–9519, 2021.

PAPASTERGIOU, T. et al. Tensor decomposition for multiple-instance classification of high-order medical data. Complexity, Hindawi, v. 2018, 2018.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KOPF, A.; YANG, E.; DEVITO, Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.; STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32. Curran Associates, Inc., 2019. p. 8024–8035. Disponível em: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

PEDRINI, H.; SCHWARTZ, W. R. Análise de imagens digitais: princípios, algoritmos e aplicações. [S.l.]: Thomson Learning, 2008.

PENTLAND, A. P. Fractal-based description of natural scenes. IEEE transactions on pattern analysis and machine intelligence, IEEE, n. 6, p. 661–674, 1984.

PRATT, L. Y. Discriminability-based transfer between neural networks. In: Advances in neural information processing systems. [S.l.: s.n.], 1993. p. 204–211.

QIN, J.; PUCKETT, L.; QIAN, X. Image based fractal analysis for detection of cancer cells. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). [S.l.: s.n.], 2020. p. 1482–1486.

QUINLAN, J. R. Simplifying decision trees. International journal of man-machine studies, Elsevier, v. 27, n. 3, p. 221–234, 1987.

REDDY, A. S. B.; JULIET, D. S. Transfer learning with resnet-50 for malaria cell-image classification. In: 2019 International Conference on Communication and Signal Processing (ICCSP). [S.l.: s.n.], 2019. p. 0945–0949.

RIBEIRO, M. G.; NEVES, L. A.; do Nascimento, M. Z.; ROBERTO, G. F.; MARTINS, A. S.; Azevedo Tosta, T. A. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. Expert Systems with Applications, v. 120, p. 262 – 278, 2019. ISSN 0957-4174. Disponível em: http://www.sciencedirect.com/science/article/pii/S0957417418307541.

RIBEIRO, M. T.; CONTRIBUTORS. lime. [S.l.]: GitHub, 2016. https://github.com/marcotcr/lime.

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. [S.l.: s.n.], 2016. p. 1135–1144.

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.

ROBERTO, G. F.; NASCIMENTO, M. Z.; MARTINS, A. S.; TOSTA, T. A.; FARIA, P. R.; NEVES, L. A. Classification of breast and colorectal tumors based on percolation of color normalized images. Computers Graphics, v. 84, p. 134–143, 2019. ISSN 0097-8493. Disponível em: https://www.sciencedirect.com/science/article/pii/S0097849319301360.

ROBERTO, G. F.; NEVES, L. A.; LONGO, L. H. da C.; ROZENDO, G. B.; TOSTA, T. A. A.; FARIA, P. R. de; MARTINS, A. S.; NASCIMENTO, M. Z. do. Percolation features: An approach for evaluating fractal properties in colour images. Software Impacts, Elsevier, v. 14, p. 100387, 2022.

ROBERTO, G. F.; NEVES, L. A.; NASCIMENTO, M. Z.; TOSTA, T. A.; LONGO, L. C.; MARTINS, A. S.; FARIA, P. R. Features based on the percolation theory for quantification of non-hodgkin lymphomas. Computers in Biology and Medicine, v. 91, p. 135 – 147, 2017. ISSN 0010-4825. Disponível em: http://www.sciencedirect.com/science/article/pii/S0010482517303384.

ROSAI, J. Why microscopy will remain a cornerstone of surgical pathology. Laboratory investigation, Nature Publishing Group, v. 87, n. 5, p. 403–408, 2007.

RUBERTO, C. D.; PUTZU, L.; ARABNIA, H.; QUOC-NAM, T. A feature learning framework for histology images classification. Emerging trends in applications and infrastructures for computational biology, bioinformatics, and systems biology: systems and applications, Elsevier Press, p. 37–48, 2016.

SAGI, O.; ROKACH, L. Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery, v. 8, n. 4, p. e1249, 2018. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1249.

SAMEK, W.; MÜLLER, K.-R. Towards explainable artificial intelligence. In: Explainable AI: interpreting, explaining and visualizing deep learning. [S.l.]: Springer, 2019. p. 5–22.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2018. p. 4510–4520.

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.

SANTOS, G. T. dos. Lima e quaresma: uma análise da alteridade em "triste fim de policarpo quaresma" à luz da cosmologia semiótica de charles sanders peirce. FronteiraZ. Revista do Programa de Estudos Pós-Graduados em Literatura e Crítica Literária, n. 21, p. 148–162, 2018.

SANTOS, L. F. S. D.; ROZENDO, G. B.; NASCIMENTO, M. Z. D.; TOSTA, T. A. A.; LONGO, L. H. D. C.; NEVES, L. A. Multidimensional shannon entropy (hm) as an approach to classify he colorectal images. In: 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP). [S.l.: s.n.], 2022. CFP2255E-ART, p. 1–4.

SCHAPIRE, R. E. The strength of weak learnability. Machine learning, Springer, v. 5, n. 2, p. 197–227, 1990.

Segato dos Santos, L. F.; NEVES, L. A.; ROZENDO, G. B.; RIBEIRO, M. G.; Zanchetta do Nascimento, M.; Azevedo Tosta, T. A. Multidimensional and fuzzy sample entropy (sampenmf) for quantifying he histological images of colorectal cancer. Computers in Biology and Medicine, v. 103, p. 148 – 160, 2018. ISSN 0010-4825. Disponível em: http://www.sciencedirect.com/science/article/pii/S0010482518303123.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. [S.l.: s.n.], 2017. p. 618–626.

SENA, P.; FIORESI, R.; FAGLIONI, F.; LOSI, L.; FAGLIONI, G.; RONCUCCI, L. Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images. Oncology letters, Spandidos Publications, v. 18, n. 6, p. 6101–6107, 2019.

SIEGEL, R. L.; MILLER, K. D.; FUCHS, H. E.; JEMAL, A. Cancer statistics, 2022. CA: A Cancer Journal for Clinicians, v. 72, n. 1, p. 7–33, 2022. Disponível em: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21708.

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.

SRIVASTAVA, R. K.; GREFF, K.; SCHMIDHUBER, J. Highway networks. arXiv preprint arXiv:1505.00387, 2015.

STEDMAN, T. Stedman's medical dictionary. [S.l.]: Dalcassian publishing company, 1920.

STRELNIKER, Y. M.; HAVLIN, S.; BUNDE, A. Fractals and percolation. Encyclopedia of Complexity and Systems Science, Springer New York, p. 3847–3858, 2009.

SUŁOT, D.; ALONSO-CANEIRO, D.; KSIENIEWICZ, P.; KRZYZANOWSKA-BERKOWSKA, P.; ISKANDER, D. R. Glaucoma classification based on scanning laser ophthalmoscopic images using a deep learning ensemble method. Plos one, Public Library of Science San Francisco, CA USA, v. 16, n. 6, p. e0252339, 2021.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.: s.n.], 2017. v. 31, n. 1.

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.

TAINO, D. F.; RIBEIRO, M. G.; ROBERTO, G. F.; ZAFALON, G. F.; NASCIMENTO, M. Z. do; TOSTA, T. A.; MARTINS, A. S.; NEVES, L. A. Analysis of cancer in histological images: employing an approach based on genetic algorithm. Pattern Analysis and Applications, Springer, v. 24, n. 2, p. 483–496, 2021.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. International conference on machine learning. [S.l.], 2019. p. 6105–6114.

TAVOLARA, T. E.; NIAZI, M. K. K.; AROLE, V.; CHEN, W.; FRANKEL, W.; GURCAN, M. N. A modular cgan classification framework: Application to colorectal tumor detection. Scientific reports, Nature Publishing Group, v. 9, n. 1, p. 1–8, 2019.

TENGUAM, J. J.; LONGO, L. H. D. C.; SILVA, A. B.; FARIA, P. R. D.; NASCIMENTO, M. Z. D.; NEVES, L. A. Classification of he images exploring ensemble learning with two-stage feature selection. In: 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP). [S.l.: s.n.], 2022. CFP2255E-ART, p. 1–4.

TITFORD, M. The long history of hematoxylin. Biotechnic & Histochemistry, Taylor Francis, v. 80, n. 2, p. 73–78, 2005. Disponível em: https://doi.org/10.1080/10520290500138372.

TOĞAÇAR, M.; CÖMERT, Z.; ERGEN, B. Enhancing of dataset using deep-dream, 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.

TORCHVISION.MODELS. Torch Contributors, 2021. Disponível em: https://pytorch.org/vision/stable/models.html.

TORREY, L.; SHAVLIK, J. Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. [S.l.]: IGI global, 2010. p. 242–264.

TOSTA, T. A.; BRUNO, D. O.; LONGO, L. C.; NASCIMENTO, M. Z. do. Colour feature extraction and polynomial algorithm for classification of lymphoma images. In: SPRINGER NATURE. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019, Havana, Cuba, October 28-31, 2019, Proceedings. [S.l.], 2019. v. 11896, p. 262.

TRIPATHI, S.; SINGH, S. K. Ensembling handcrafted features with deep features: an analytical study for classification of routine colon cancer histopathological nuclei images. MULTIMEDIA TOOLS AND APPLICATIONS, Springer, 2020.

TURKOGLU, M. Covidetectionet: Covid-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble. Applied Intelligence, Springer, v. 51, n. 3, p. 1213–1226, 2021.

URBANOWICZ, R. J.; MEEKER, M.; La Cava, W.; OLSON, R. S.; MOORE, J. H. Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics, v. 85, p. 189–203, 2018. ISSN 1532-0464. Disponível em: https://www.sciencedirect.com/science/article/pii/S1532046418301400.

WALT, S. Van der; SCHÖNBERGER, J. L.; NUNEZ-IGLESIAS, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T. scikit-image: image processing in python. PeerJ, PeerJ Inc., v. 2, p. e453, 2014.

WANG, W.; YANG, Y.; WANG, X.; WANG, W.; LI, J. Development of convolutional neural network and its application in image classification: a survey. Optical Engineering, SPIE, v. 58, n. 4, p. 1 – 19, 2019.

WANG, Z.; LI, M.; WANG, H.; JIANG, H.; YAO, Y.; ZHANG, H.; XIN, J. Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features. IEEE Access, v. 7, p. 105146–105158, 2019.

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.

WATANABE, S. Pattern Recognition: Human and Mechanical. USA: John Wiley & Sons, Inc., 1985. ISBN 0471808156.

WEI, L.; SU, R.; WANG, B.; LI, X.; ZOU, Q.; GAO, X. Integration of deep feature representations and handcrafted features to improve the prediction of n6-methyladenosine sites. Neurocomputing, v. 324, p. 3 – 9, 2019. ISSN 0925-2312. Deep Learning for Biological/Clinical Data. Disponível em: http://www.sciencedirect.com/science/article/pii/S0925231218306325.

WELLS, L.; BEDNARZ, T. Explainable ai and reinforcement learning—a systematic review of current approaches and trends. Frontiers in artificial intelligence, Frontiers Media SA, v. 4, p. 550030, 2021.

WOLPERT, D. H. Stacked generalization. Neural networks, Elsevier, v. 5, n. 2, p. 241–259, 1992.

XUE, D.; ZHOU, X.; LI, C.; YAO, Y.; RAHAMAN, M. M.; ZHANG, J.; CHEN, H.; ZHANG, J.; QI, S.; SUN, H. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. IEEE Access, v. 8, p. 104603–104618, 2020.

YAMAGUCHI, T.; HASHIMOTO, S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. Machine Vision and Applications, v. 21, p. 797–809, 2010.

YANG, J.; YANG, J.-y.; ZHANG, D.; LU, J.-f. Feature fusion: parallel strategy vs. serial strategy. Pattern recognition, Elsevier, v. 36, n. 6, p. 1369–1381, 2003.

YU, C.; CHEN, H.; LI, Y.; PENG, Y.; LI, J.; YANG, F. Breast cancer classification in pathological images based on hybrid features. Multimedia Tools and Applications, Springer, v. 78, n. 15, p. 21325–21345, 2019.

ZAFER, C. Fusing fine-tuned deep features for recognizing different tympanic membranes. Biocybernetics and Biomedical Engineering, Elsevier, v. 40, n. 1, p. 40–51, 2020.

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.

Apêndice A

Resultados Detalhados dos Experimentos

Neste apêndice estão listados todos os resultados obtidos com as classificações resultantes da aplicação do método proposto. Para cada base de imagens H&E, os resultados foram coletados para os 52 vetores de características, analisados com incrementos consecutivos de 5 atributos, totalizando 260 resultados. Estes vetores foram divididos em grupos representantes da origem dos descritores que compõem os vetores de características, conforme metodologia proposta: *handcrafted*, descritores obtidos com a aplicação de técnicas fractais; *deep learned*, descritores obtidos com a extração dos valores das camadas de uma CNN; IAx, descritores com a aplicação de técnicas fractais sobre uma imagem obtida para explicação do resultado de uma arquitetura; *ensemble handcrafted* e *deep learned*, uma combinação de descritores do grupo *handcrafted* e *deep learned*; *ensemble* de *deep learned*, uma combinação de descritores do grupo *deep learned* obtidos por arquiteturas diferentes; e *ensemble* de IAx, uma combinação de descritores do grupo IAx obtidos de diferentes explicações. Nas Tabelas A.1, A.2, A.3 e A.4 estão os resultados para as bases CR, LG, OED e UCSB, respectivamente. Os melhores resultados foram destacados em negrito para cada tipo de combinação.

Tabela A.1: Taxas de Acurácias (%) obtidas para a base CR a partir de diferentes combinações de descritores.

Origem	**Composição**	**Nº de Descritores**				
		25	**20**	**15**	**10**	**5**
Handcrafted	Fractais (F)	86,06	**86,06**	85,45	84,24	80,61
Deep Learned	DenseNet-121 (D)	99,39	**100**	99,39	98,18	96,36
	EfficientNet-b2 (E)	100	100	**100**	99,39	97,58
	Inception-V3 (I)	98,18	**98,18**	95,15	95,15	95,15
	ResNet-50 (R)	97,58	97,58	**97,58**	95,15	93,94
	VGG-19 (V)	96,97	98,18	**98,18**	96,97	94,55
IAx	CAM DenseNet-121 (D_{CAM})	**84,85**	83,03	82,42	80	83,03
	CAM EfficientNet-b2 (E_{CAM})	**85,45**	84,85	83,64	82,42	84,24
	CAM Inception-V3 (I_{CAM})	75,76	75,76	76,97	**78,79**	76,36
	CAM ResNet-50 (R_{CAM})	73,94	**74,55**	72,73	70,30	72,12
	CAM VGG-19 (V_{CAM})	77,58	78,79	**79,39**	75,15	74,55
	LIME DenseNet-121 (D_{LIME})	67,88	65,45	**69,09**	66,67	67,88
	LIME EfficientNet-b2 (E_{LIME})	74,55	75,15	73,94	75,15	**76,36**
	LIME Inception-V3 (I_{LIME})	**58,18**	55,76	55,76	53,33	52,73
	LIME ResNet-50 (R_{LIME})	83,03	82,42	83,03	82,42	**83,03**
	LIME VGG-19 (V_{LIME})	62,42	62,42	**64,85**	62,42	63,64
Ensemble Handcrafted e *Deep Learned*	F + D	**100**	99,39	99,39	97,58	95,76
	F + E	99,39	99,39	**100**	99,39	95,15
	F + I	**98,18**	97,58	95,15	92,73	95,15
	F + R	97,58	**97,58**	96,97	95,15	94,55
	F + V	96,97	97,58	**98,18**	96,36	95,15
	F + D + E + I + R + V	100	100	100	**100**	98,18
Ensemble de *Deep Learned*	D + E	100	100	100	**100**	98,79
	D + I	98,79	100	**100**	97,58	96,36
	D + R	99,39	99,39	**99,39**	97,58	96,97
	D + V	100	**100**	98,18	96,97	95,76
	E + I	100	**100**	99,39	98,79	95,15
	E + R	100	**100**	99,39	98,18	95,76
	E + V	100	100	100	**100**	96,97
	I + R	97,58	**97,58**	96,97	96,97	95,76
	I + V	**98,18**	96,97	95,76	96,97	93,33
	R + V	**98,79**	97,58	98,18	96,97	95,76
	D + E + I + R + V	100	100	100	**100**	98,18
Ensemble IAx	D_{CAM} + E_{CAM}	**85,45**	81,21	81,82	82,42	83,03
	D_{CAM} + I_{CAM}	**84,24**	81,21	81,21	80	81,82
	D_{CAM} + R_{CAM}	80	**81,21**	78,18	80	80,61
	D_{CAM} + V_{CAM}	78,79	78,18	78,79	78,79	**79,39**
	E_{CAM} + I_{CAM}	87,88	**88,48**	86,67	86,06	76,36
	E_{CAM} + R_{CAM}	**87,88**	86,06	84,85	84,24	74,55
	E_{CAM} + V_{CAM}	81,82	**84,24**	83,03	83,03	76,97
	I_{CAM} + R_{CAM}	75,15	75,15	74,55	**77,58**	76,97
	I_{CAM} + V_{CAM}	77,58	76,97	75,76	76,97	**77,58**
	R_{CAM} + V_{CAM}	76,97	75,76	75,15	**79,39**	73,33
	D_{CAM} + E_{CAM} + I_{CAM} + R_{CAM} + V_{CAM}	**84,24**	83,64	81,82	82,42	80,61
	D_{LIME} + E_{LIME}	74,55	75,15	75,76	75,15	**76,36**
	D_{LIME} + I_{LIME}	65,45	63,03	64,24	**67,88**	67,27
	D_{LIME} + R_{LIME}	82,42	81,82	**83,03**	82,42	81,82
	D_{LIME} + V_{LIME}	67,27	**69,09**	67,27	65,45	64,24
	E_{LIME} + I_{LIME}	74,55	74,55	75,15	75,76	**76,97**
	E_{LIME} + R_{LIME}	**82,42**	80,61	80,61	80	80,61
	E_{LIME} + V_{LIME}	75,15	74,55	74,55	74,55	**75,15**
	I_{LIME} + R_{LIME}	**83,64**	83,03	82,42	83,03	81,21
	I_{LIME} + V_{LIME}	63,03	61,21	**64,24**	59,39	59,39
	R_{LIME} + V_{LIME}	**83,64**	83,03	83,03	83,03	83,03
	D_{LIME} + E_{LIME} + I_{CAM} + R_{LIME} + V_{LIME}	81,21	82,42	**82,42**	81,82	76,97

Tabela A.2: Taxas de Acurácias (%) obtidas para a base LG a partir de diferentes combinações de descritores.

Origem	**Composição**	**Nº de Descritores**				
		25	**20**	**15**	**10**	**5**
Handcrafted	Fractais (F)	93,21	93,96	**94,34**	90,19	80,38
Deep Learned	DenseNet-121 (D)	99,62	**99,62**	98,87	98,49	97,74
	EfficientNet-b2 (E)	**98,11**	97,36	97,74	96,98	95,47
	Inception-V3 (I)	**95,09**	94,72	94,34	90,94	89,06
	ResNet-50 (R)	98,11	**98,49**	97,74	98,11	95,09
	VGG-19 (V)	93,21	93,96	91,70	92,08	**94,34**
IAx	CAM DenseNet-121 (D_{CAM})	**85,28**	84,91	81,13	80	78,11
	CAM EfficientNet-b2 (E_{CAM})	84,15	83,77	**85,28**	76,98	71,32
	CAM Inception-V3 (I_{CAM})	**82,26**	81,51	78,87	78,87	74,34
	CAM ResNet-50 (R_{CAM})	78,49	79,62	**81,13**	79,62	79,62
	CAM VGG-19 (V_{CAM})	89,43	90,19	**91,32**	87,55	87,92
	LIME DenseNet-121 (D_{LIME})	**66,42**	65,66	64,53	63,77	64,15
	LIME EfficientNet-b2 (E_{LIME})	**67,55**	65,28	56,98	54,72	54,34
	LIME Inception-V3 (I_{LIME})	63,77	63,40	62,64	**63,77**	57,74
	LIME ResNet-50 (R_{LIME})	**73,96**	72,83	71,70	72,83	72,08
	LIME VGG-19 (V_{LIME})	81,51	81,13	**81,51**	80,75	81,13
Ensemble Handcrafted e *Deep Learned*	F + D	99,62	99,62	**99,62**	99,25	97,74
	F + E	**97,74**	96,23	95,47	94,34	93,96
	F + I	**96,60**	95,85	94,72	88,30	88,68
	F + R	98,11	98,11	97,74	**98,49**	95,85
	F + V	**95,47**	94,34	94,72	94,34	92,83
	F + D + E + I + R + V	98,87	**99,25**	98,87	98,49	97,36
Ensemble de *Deep Learned*	D + E	99,62	99,62	99,62	**99,62**	97,74
	D + I	99,62	99,62	**99,62**	99,25	97,74
	D + R	**100**	99,62	99,62	98,49	97,74
	D + V	**99,62**	99,25	98,87	96,98	96,60
	E + I	98,49	**98,49**	97,36	96,98	95,47
	E + R	98,87	98,49	98,49	**98,87**	93,58
	E + V	**96,23**	95,85	94,72	95,09	93,21
	I + R	98,49	**98,87**	97,74	98,11	96,60
	I + V	93,96	**93,96**	92,83	93,21	93,21
	R + V	98,11	**98,11**	97,74	96,23	94,72
	D + E + I + R + V	**99,62**	98,87	99,25	98,49	97,36
Ensemble IAx	$D_{CAM} + E_{CAM}$	**84,53**	81,13	81,51	80,38	81,51
	$D_{CAM} + I_{CAM}$	**86,42**	83,40	82,64	82,26	81,13
	$D_{CAM} + R_{CAM}$	**83,02**	80	81,13	76,98	78,11
	$D_{CAM} + V_{CAM}$	88,30	88,30	88,30	**88,30**	87,17
	$E_{CAM} + I_{CAM}$	**84,53**	82,26	80	78,49	76,23
	$E_{CAM} + R_{CAM}$	**84,91**	83,40	79,25	73,58	72,83
	$E_{CAM} + V_{CAM}$	88,30	89,43	89,06	88,68	**90,57**
	$I_{CAM} + R_{CAM}$	**81,51**	80,75	78,49	76,23	78,87
	$I_{CAM} + V_{CAM}$	89,43	89,06	**89,81**	89,43	89,06
	$R_{CAM} + V_{CAM}$	89,43	89,43	**90,19**	87,92	86,79
	$D_{CAM} + E_{CAM} + I_{CAM} + R_{CAM} + V_{CAM}$	85,28	86,79	88,30	87,55	**89,43**
	$D_{LIME} + E_{LIME}$	**64,15**	63,02	55,85	52,08	50,94
	$D_{LIME} + I_{LIME}$	**69,43**	68,68	67,17	61,89	60
	$D_{LIME} + R_{LIME}$	70,94	72,08	**72,45**	71,32	69,06
	$D_{LIME} + V_{LIME}$	79,25	79,62	78,11	**79,62**	79,25
	$E_{LIME} + I_{LIME}$	**70,19**	61,89	57,36	57,74	58,87
	$E_{LIME} + R_{LIME}$	**70,94**	68,30	65,28	65,66	65,28
	$E_{LIME} + V_{LIME}$	**80,75**	80	79,62	80	80,38
	$I_{LIME} + R_{LIME}$	**66,04**	62,64	61,51	57,74	54,72
	$I_{LIME} + V_{LIME}$	80	79,62	79,62	79,25	**80,38**
	$R_{LIME} + V_{LIME}$	78,49	78,49	78,49	79,62	**79,62**
	$D_{LIME} + E_{LIME} + I_{CAM} + R_{LIME} + V_{LIME}$	75,68	77,70	76,35	73,65	**75**

Tabela A.3: Taxas de Acurácias (%) obtidas para a base OED a partir de diferentes combinações de descritores.

Origem	**Composição**	**Nº de Descritores**				
		25	**20**	**15**	**10**	**5**
Handcrafted	Fractais (F)	87,84	88,51	**88,51**	86,49	83,78
Deep Learned	DenseNet-121 (D)	95,95	96,62	**97,30**	94,59	91,22
	EfficientNet-b2 (E)	**95,27**	93,92	93,92	93,24	89,86
	Inception-V3 (I)	95,95	94,59	**96,62**	95,95	94,59
	ResNet-50 (R)	**95,95**	95,27	94,59	91,22	89,19
	VGG-19 (V)	94,59	94,59	96,62	**96,62**	92,57
IAx	CAM DenseNet-121 (D_{CAM})	70,95	68,92	69,59	68,92	**71,62**
	CAM EfficientNet-b2 (E_{CAM})	**69,59**	66,22	64,86	64,19	60,81
	CAM Inception-V3 (I_{CAM})	79,05	78,38	78,38	78,38	**79,05**
	CAM ResNet-50 (R_{CAM})	72,30	75	73,65	74,32	**76,35**
	CAM VGG-19 (V_{CAM})	77,70	77,70	77,70	**79,05**	78,38
	LIME DenseNet-121 (D_{LIME})	77,70	**78,38**	76,35	75,68	73,65
	LIME EfficientNet-b2 (E_{LIME})	71,62	**72,30**	70,27	69,59	65,54
	LIME Inception-V3 (I_{LIME})	71,62	72,30	72,97	71,62	**73,65**
	LIME ResNet-50 (R_{LIME})	75	75,68	75,68	75,68	**79,05**
	LIME VGG-19 (V_{LIME})	**80,41**	78,38	76,35	76,35	77,03
Ensemble Handcrafted e *Deep Learned*	F + D	97,30	**97,97**	96,62	95,95	93,24
	F + E	**95,95**	94,59	95,27	94,59	91,89
	F + I	**96,62**	95,95	95,95	93,92	91,22
	F + R	**95,95**	94,59	91,89	87,84	87,84
	F + V	95,27	95,95	95,95	**96,62**	91,89
	F + D + E + I + R + V	**95,27**	93,24	93,92	93,92	89,19
Ensemble de *Deep Learned*	D + E	96,62	**97,30**	96,62	94,59	90,54
	D + I	96,62	**97,30**	95,95	96,62	90,54
	D + R	**97,30**	95,95	94,59	91,89	89,19
	D + V	94,59	**95,95**	95,27	94,59	91,89
	E + I	**97,97**	95,27	94,59	93,24	91,22
	E + R	**96,62**	95,95	94,59	91,89	90,54
	E + V	95,27	94,59	**95,95**	93,92	93,92
	I + R	96,62	**97,30**	94,59	91,22	86,49
	I + V	97,30	**97,97**	96,62	95,95	92,57
	R + V	**94,59**	93,92	92,57	93,92	90,54
	D + E + I + R + V	**94,59**	93,24	93,24	93,24	89,86
Ensemble IAx	$D_{CAM} + E_{CAM}$	68,24	68,92	69,59	**70,95**	70,27
	$D_{CAM} + I_{CAM}$	75,68	76,35	77,03	**78,38**	77,70
	$D_{CAM} + R_{CAM}$	**72,30**	70,27	70,27	68,24	69,59
	$D_{CAM} + V_{CAM}$	77,03	77,70	77,03	**79,05**	78,38
	$E_{CAM} + I_{CAM}$	75,68	77,03	77,70	**79,05**	78,38
	$E_{CAM} + R_{CAM}$	**75**	74,32	74,32	72,30	70,95
	$E_{CAM} + V_{CAM}$	77,03	77,70	77,70	**79,73**	79,05
	$I_{CAM} + R_{CAM}$	76,35	77,03	77,70	78,38	**79,05**
	$I_{CAM} + V_{CAM}$	78,38	79,05	**80,41**	79,73	79,05
	$R_{CAM} + V_{CAM}$	76,35	77,03	77,03	77,70	**78,38**
	$D_{CAM} + E_{CAM} + I_{CAM} + R_{CAM} + V_{CAM}$	79,05	78,38	81,08	**81,08**	80,41
	$D_{LIME} + E_{LIME}$	72,97	**74,32**	73,65	69,59	70,27
	$D_{LIME} + I_{LIME}$	76,35	77,70	**79,05**	77,03	75
	$D_{LIME} + R_{LIME}$	**80,41**	76,35	78,38	77,03	77,03
	$D_{LIME} + V_{LIME}$	**81,76**	77,70	78,38	77,03	77,70
	$E_{LIME} + I_{LIME}$	74,32	74,32	75	73,65	**75**
	$E_{LIME} + R_{LIME}$	**80,41**	79,05	79,73	78,38	77,03
	$E_{LIME} + V_{LIME}$	**77,03**	75	75,68	76,35	72,97
	$I_{LIME} + R_{LIME}$	77,03	76,35	77,70	76,35	**77,70**
	$I_{LIME} + V_{LIME}$	80,41	**81,08**	78,38	77,03	77,70
	$R_{LIME} + V_{LIME}$	81,08	**81,08**	73,65	69,59	69,59
	$D_{LIME} + E_{LIME} + I_{CAM} + R_{LIME} + V_{LIME}$	75,68	**77,70**	76,35	73,65	75

Tabela A.4: Taxas de Acurácias (%) obtidas para a base UCSB a partir de diferentes combinações de descritores.

Origem	**Composição**	**Nº de Descritores**				
		25	**20**	**15**	**10**	**5**
Handcrafted	Fractais (F)	72,41	74,14	**74,14**	68,97	72,41
Deep Learned	DenseNet-121 (D)	93,10	93,10	**93,10**	87,93	86,21
	EfficientNet-b2 (E)	**93,10**	89,66	87,93	86,21	84,48
	Inception-V3 (I)	89,66	89,66	**89,66**	87,93	82,76
	ResNet-50 (R)	89,66	91,38	**91,38**	89,66	84,48
	VGG-19 (V)	**86,21**	84,48	84,48	79,31	75,86
IAx	CAM DenseNet-121 (D_{CAM})	75,86	75,86	77,59	**79,31**	74,14
	CAM EfficientNet-b2 (E_{CAM})	75,86	74,14	75,86	75,86	**79,31**
	CAM Inception-V3 (I_{CAM})	72,41	74,14	70,69	**79,31**	75,86
	CAM ResNet-50 (R_{CAM})	70,69	70,69	72,41	72,41	**79,31**
	CAM VGG-19 (V_{CAM})	**79,31**	75,86	77,59	77,59	72,41
	LIME DenseNet-121 (D_{LIME})	68,97	65,52	65,52	**70,69**	63,79
	LIME EfficientNet-b2 (E_{LIME})	53,45	**62,07**	60,34	53,45	51,72
	LIME Inception-V3 (I_{LIME})	**58,62**	56,90	56,90	56,90	50
	LIME ResNet-50 (R_{LIME})	**77,59**	75,86	75,86	74,14	72,41
	LIME VGG-19 (V_{LIME})	50	51,72	50	**55,17**	48,28
Ensemble Handcrafted e *Deep Learned*	F + D	91,38	**91,38**	87,93	86,21	86,21
	F + E	89,66	89,66	**89,66**	84,48	79,31
	F + I	87,93	87,93	**89,66**	87,93	81,03
	F + R	89,66	91,38	91,38	**91,38**	86,21
	F + V	**87,93**	86,21	86,21	74,14	74,14
	F + D + E + I + R + V	**89,66**	84,48	86,21	82,76	75,86
Ensemble de *Deep Learned*	D + E	**94,83**	93,10	93,10	87,93	84,48
	D + I	**93,10**	91,38	89,66	86,21	84,48
	D + R	89,66	91,38	**91,38**	89,66	86,21
	D + V	**89,66**	86,21	84,48	84,48	81,03
	E + I	**93,10**	89,66	87,93	87,93	81,035
	E + R	91,38	89,66	**93,10**	91,38	79,31
	E + V	**89,66**	86,21	81,03	79,31	74,14
	I + R	93,10	89,66	**93,10**	87,93	87,93
	I + V	86,21	86,21	**87,93**	86,21	81,03
	R + V	89,66	89,66	89,66	**89,66**	79,31
	D + E + I + R + V	**91,38**	87,93	86,21	82,76	74,14
Ensemble IAx	$D_{CAM} + E_{CAM}$	77,59	77,59	**77,59**	72,41	74,14
	$D_{CAM} + I_{CAM}$	74,14	74,14	75,86	**77,59**	74,14
	$D_{CAM} + R_{CAM}$	72,41	75,86	75,86	74,14	**81,03**
	$D_{CAM} + V_{CAM}$	75,86	**75,86**	74,14	72,41	70,69
	$E_{CAM} + I_{CAM}$	77,59	**77,59**	75,86	75,86	74,14
	$E_{CAM} + R_{CAM}$	72,41	74,14	75,86	74,14	**79,31**
	$E_{CAM} + V_{CAM}$	75,86	**77,59**	75,86	75,86	74,14
	$I_{CAM} + R_{CAM}$	72,41	74,14	74,14	75,86	**79,31**
	$I_{CAM} + V_{CAM}$	**77,59**	75,86	72,41	74,14	74,14
	$R_{CAM} + V_{CAM}$	75,86	77,59	77,59	77,59	**77,59**
	$D_{CAM} + E_{CAM} + I_{CAM} + R_{CAM} + V_{CAM}$	74,14	74,14	74,14	74,14	**74,14**
	$D_{LIME} + E_{LIME}$	63,79	63,79	58,62	**65,52**	56,90
	$D_{LIME} + I_{LIME}$	62,07	**63,79**	56,90	58,62	56,90
	$D_{LIME} + R_{LIME}$	**74,14**	72,41	72,41	70,69	68,97
	$D_{LIME} + V_{LIME}$	65,52	**68,97**	65,52	67,24	63,79
	$E_{LIME} + I_{LIME}$	50	50	**56,90**	55,17	50
	$E_{LIME} + R_{LIME}$	70,69	**79,31**	75,86	75,86	72,41
	$E_{LIME} + V_{LIME}$	37,93	36,21	39,66	43,10	**44,83**
	$I_{LIME} + R_{LIME}$	75,86	77,59	**77,59**	75,86	70,69
	$I_{LIME} + V_{LIME}$	56,90	56,90	**60,34**	55,17	56,90
	$R_{LIME} + V_{LIME}$	74,14	75,86	**75,86**	74,14	72,41
	$D_{LIME} + E_{LIME} + I_{CAM} + R_{LIME} + V_{LIME}$	74,14	72,41	**74,14**	68,97	65,52

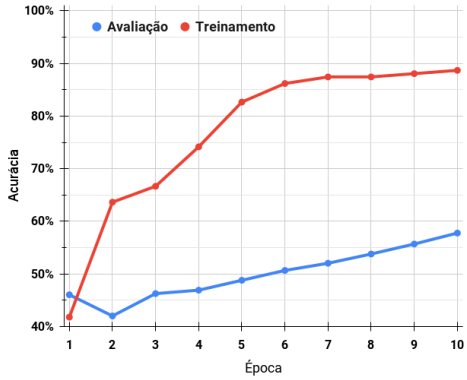
Apêndice B

Treinamento das CNNs

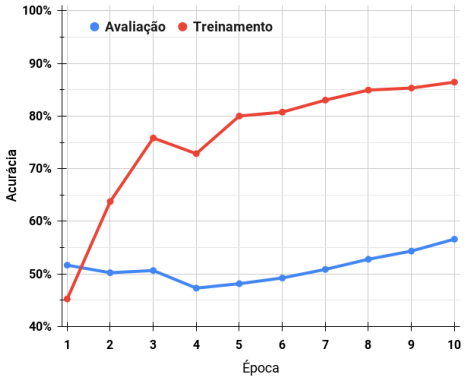
Neste apêndice estão os gráficos indicativos dos refinamentos aplicados nas CNNs, conforme a descrição na Seção 4.2. Cada figura contém os gráficos de treinamento para cada uma das quatro bases de imagens H&E. As curvas para a rede DenseNet-121, EfficientNet-b2, Inception-V3, ResNet-50 e VGG-19 estão ilustradas nas Figuras B.1, B.2, B.3, B.4 e B.5.

Figura B.1: Acurácia média obtida durante o treinamento da rede DenseNet-121 com as bases (a) CR, (b) LG, (c) OED e (e) UCSB

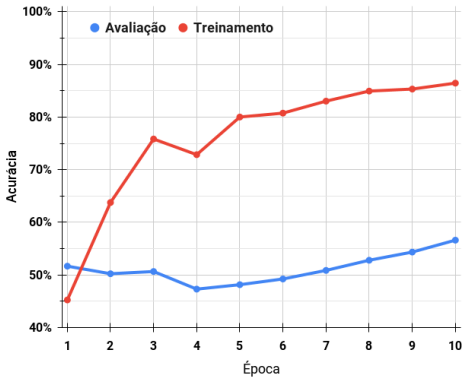
(a)



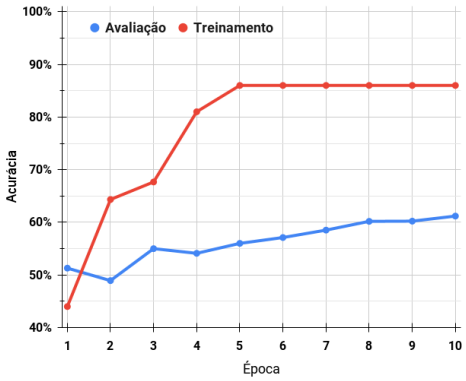
(b)



(c)



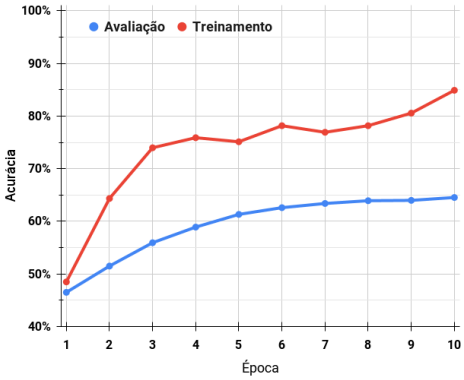
(d)



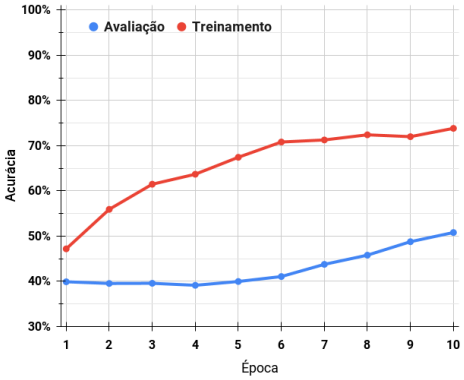
Fonte: Elaborado pelo Autor

Figura B.2: Acurácia média obtida durante o treinamento da rede EfficientNet-b2 com as bases (a) CR, (b) LG, (c) OED e (e) UCSB

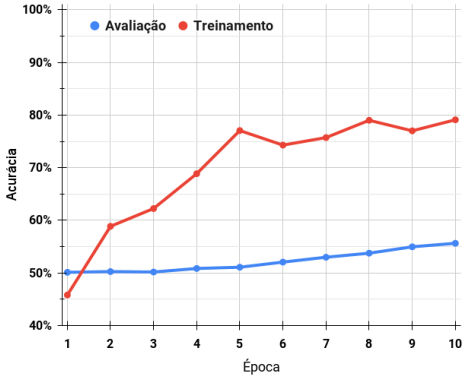
(a)



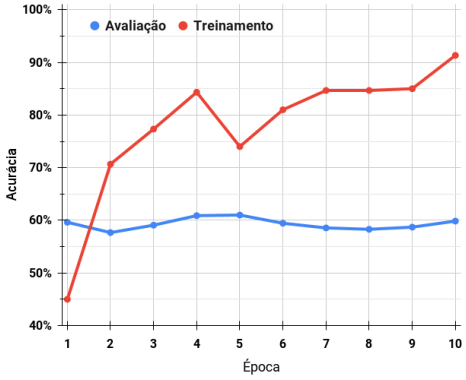
(b)



(c)



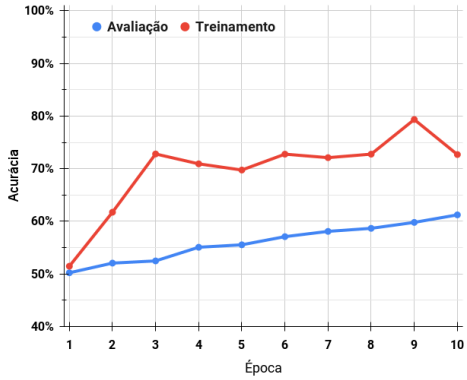
(d)



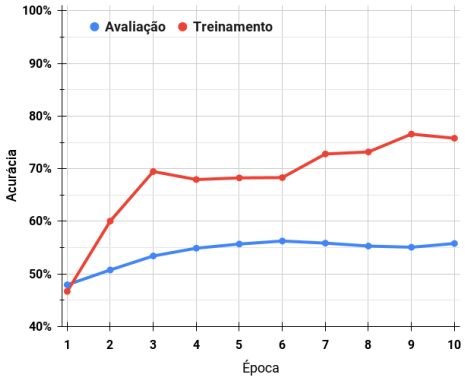
Fonte: Elaborado pelo Autor

Figura B.3: Acurácia média obtida durante o treinamento da rede Inception-V3 com as bases (a) CR, (b) LG, (c) OED e (e) UCSB

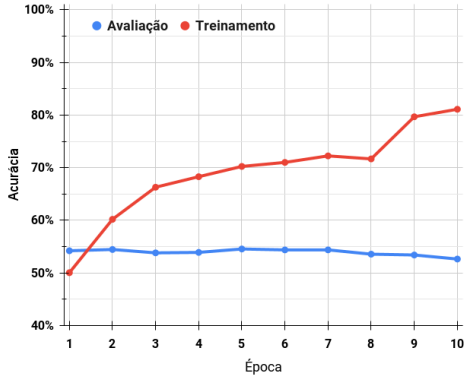
(a)



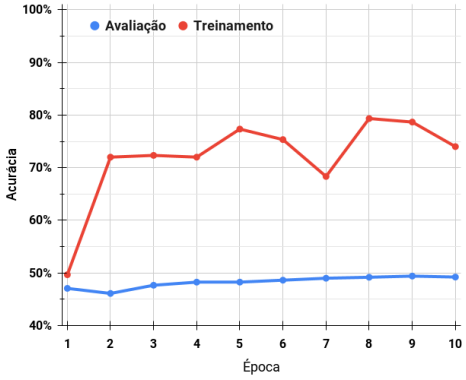
(b)



(c)



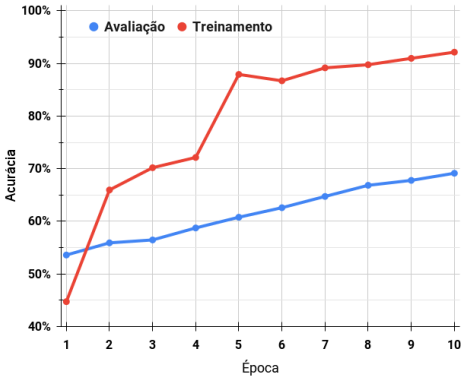
(d)



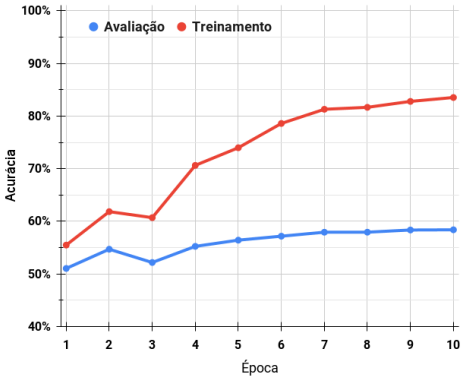
Fonte: Elaborado pelo Autor

Figura B.4: Acurácia média obtida durante o treinamento da rede ResNet-50 com as bases (a) CR, (b) LG, (c) OED e (e) UCSB

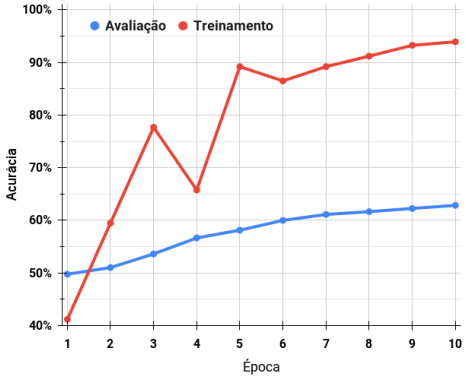
(a)



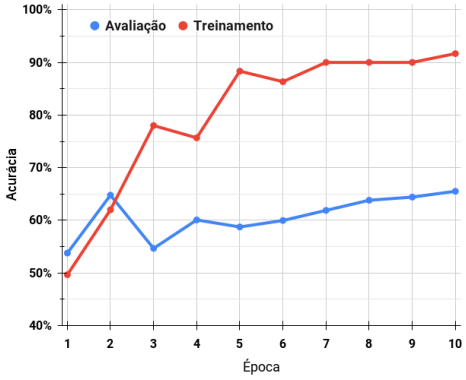
(b)



(c)

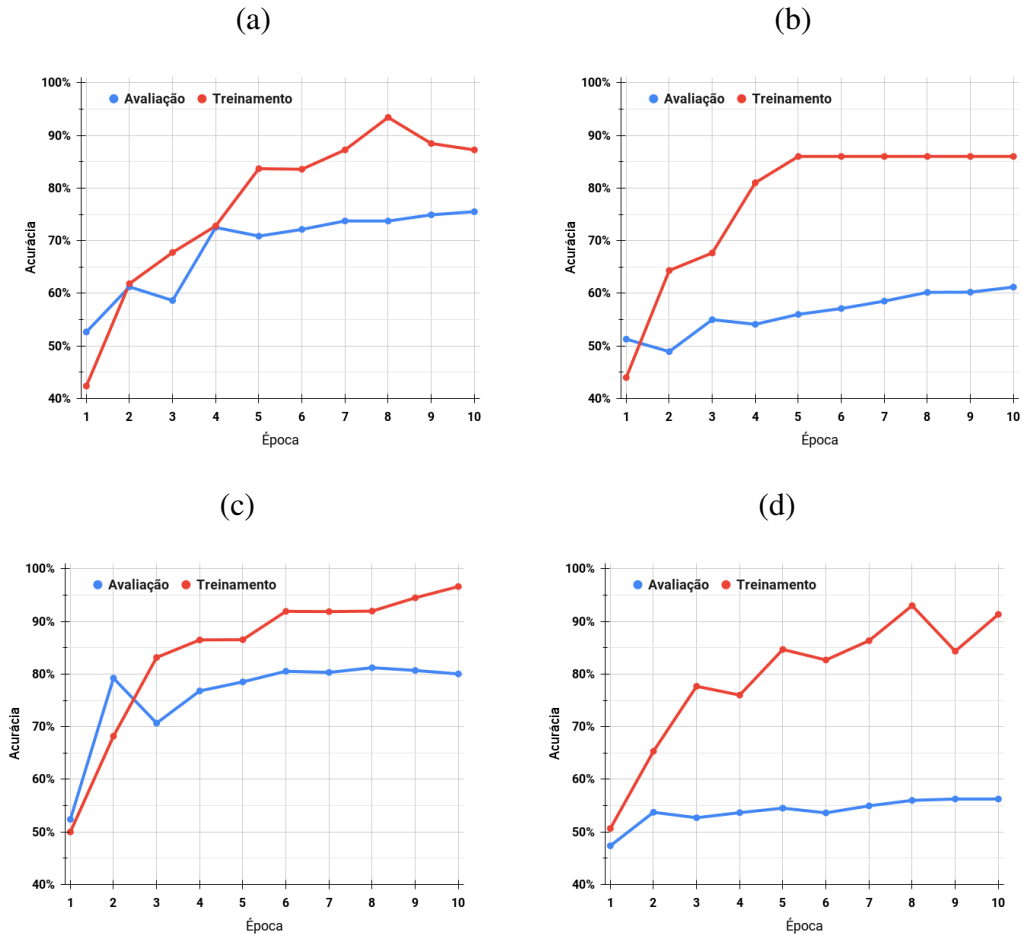


(d)



Fonte: Elaborado pelo Autor

Figura B.5: Acurácia média obtida durante o treinamento da rede VGG-19 com as bases (a) CR, (b) LG, (c) OED e (e) UCSB



Fonte: Elaborado pelo Autor