



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

Cléber Ivo de Oliveira

Modelos híbridos para classificar
imagens histológicas: uma associação de
deep features por transferência de
aprendizado com comitê de
classificadores

São José do Rio Preto
2021

Cléber Ivo de Oliveira

Modelos híbridos para classificar
imagens histológicas: uma associação de
deep features por transferência de
aprendizado com comitê de
classificadores

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto 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.

Financiadora: CNPq - Número: 132940/2019-1
CAPES

Orientador:

Prof. Dr. Leandro Alves Neves

São José do Rio Preto
2021

O48m

Oliveira, Cléber Ivo de

Modelos híbridos para classificar imagens histológicas: uma associação de deep features por transferência de aprendizado com comitê de classificadores / Cléber Ivo de Oliveira. -- São José do Rio Preto, 2021

80 f.

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

Orientador: Leandro Alves Neves

1. Inteligência artificial Aplicações médicas. 2. Aprendizado do computador. 3. Diagnóstico por imagem. 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.

Cléber Ivo de Oliveira

Modelos híbridos para classificar
imagens histológicas: uma associação de
deep features por transferência de
aprendizado com comitê de
classificadores

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto 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.

Financiadora: CNPq - Número: 132940/2019-1
CAPES

Comissão Examinadora:

Prof. Dr. Leandro Alves Neves
Unesp - Câmpus São José do Rio Preto
Orientador

Prof. Dr. Adriano Mauro Cansian
Unesp - Câmpus São José do Rio Preto

Prof. Dr. Bruno Augusto Nassif Travençolo
UFU - Universidade Federal de Uberlândia

São José do Rio Preto
05 de agosto de 2021

Ao progresso da ciência e da divulgação científica.

Agradecimentos

Agradeço à minha família e a meus amigos pelos incentivos e pelo apoio incondicional nos momentos difíceis. À eles minha profunda gratidão por compreenderem minha ausência enquanto eu me dedicava à realização deste trabalho.

Agradeço aos meus colegas de turma e de laboratório pelo companheirismo e pela troca de experiências. O ambiente amistoso no qual convivemos e solidificamos os nossos conhecimentos foi uma experiência que não me esquecerei.

Ao Prof. Dr. Leandro Alves Neves meu agradecimento pelo voto de confiança ao aceitar ser meu orientador e guiar o desenvolvimento desse trabalho. Seus ensinamentos e correções foram fundamentais em minha formação acadêmica.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de estudos, cota GM, número 132940/2019-1, que financiou parte do desenvolvimento deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço.

"Lights will guide you home

And ignite your bones

And I will try to fix you. - Cris Martin"

Coldplay (2005)

Resumo

O uso de modelos CNN explorando transferência de aprendizado profundo é uma estratégia que pode ser aplicada para definir atributos de alto nível e, consequentemente, permitir investigações de padrões em conjuntos reduzidos de imagens, situação comumente evidenciada no contexto de imagens médicas. A principal vantagem dessa estratégia está em projetar modelos que minimizam a ocorrência de *overfitting*, tornando-os mais úteis para a prática clínica. Portanto, neste trabalho, descrevemos uma proposta capaz de definir modelos híbridos para classificar imagens histológicas de tecido mamário, colorretal e hepático, por meio da associação de *deep features* via transferência de aprendizado, seleção por ranqueamento e classificação via comitê. Os atributos foram definidos a partir de camadas das arquiteturas AlexNet e ResNet-50. Os atributos foram organizadas em subconjuntos de características mais relevantes e avaliados por meio de validação cruzada *k-fold*. Os principais modelos híbridos foram definidos com *deep features* fornecidas pela rede ResNet-50, utilizando as camadas *activation_48_relu* e *avg_pool*. Os melhores resultados foram valores de acurácia de 98,00% e 99,32%, ao utilizar no máximo 35 *deep features*. O modelo foi capaz de reduzir o total de *deep features* em até 99,86% para obter os melhores valores de acurácia. Nossa proposta com informações detalhadas sobre os métodos, características e melhores associações são contribuições relevantes para a comunidade interessada no estudo de técnicas de *machine learning* para o reconhecimento de padrões.

Palavras-chave: *Deep features*. Transferência de aprendizado. Modelos híbridos. Imagens histológicas. Reconhecimento de padrões.

Abstract

The use of CNN models exploring deep learning transfer is a strategy that can be applied to define high-level attributes and make them feasible to carry out pattern investigations in reduced image sets, a situation commonly evidenced in the context of medical images. The obtained models can minimize the occurrence of overfitting and make them more useful for clinical practice. In this work, we describe a proposal capable of defining hybrid models to classify histological images from breast, colorectal and liver tissues, through the association of deep features by transfer learning, selection by ranking, and ensemble classifier. The features were defined by using layers from the AlexNet and ResNet-50 architectures. The attributes were organized into subsets of the most relevant features and submitted to a k-fold cross-validation process. The main hybrid models were defined with deep features from the ResNet-50 network, using the activation_48_relu and avg_pool layers. The best results were accuracy values of 98.00% and 99.32%, using a maximum of 35 deep features. The model was able to reduce the total of deep features by up to 99.86% to obtain the best accuracy values. Our proposal with detailed information regarding the methods, features and best combinations are relevant contributions for the community interested in the study of machine learning techniques for pattern recognition.

Keywords: *Deep features. Transfer learning. Hybrid models. Histological images. Pattern recognition.*

Lista de Figuras

2.1	Ilustração de uma convolução 2D.	23
2.2	Ilustração de uma operação de convolução e <i>pooling</i> sobre uma imagem.	25
2.3	Arquitetura de uma CNN para tarefa de reconhecimento de dígitos escritos a mão.	27
2.4	Arquitetura AlexNet: convolução, <i>max pooling</i> , LRN, FC e <i>softmax</i>	28
2.5	Estrutura de um bloco residual do modelo ResNet-50.	29
2.6	Arquitetura ResNet-50.	30
2.7	Ilustração do comportamento de um comitê de três classificadores em um conjunto de dados descrito por dois atributos.	35
4.1	Visão geral ilustrativa da estratégia proposta para classificar imagens histológicas.	48
4.2	Exemplos de imagens histológicas H&E: UCSB (GELASCA et al., 2008), benigno (4.2a) e maligno (4.2b); CR (SIRINUKUNWATTANA et al., 2017), benigno (4.2c) e maligno (4.2d); LG (AGEMAP, 2020), macho (4.2e) e fêmea (4.2f).	50
4.3	Ilustração da organização das <i>deep features</i> obtidas de uma camada em formato de vetor de atributos.	52
4.4	Ilustração do processo de seleção e classificação de atributos proposto em nosso modelo.	54

5.1	Comportamentos das principais associações a partir das <i>deep features</i> da rede ResNet-50: UCSB com a camada <i>activation_48_relu</i> ; CR com a camada <i>avg_pool</i> ; e LG com a camada <i>avg_pool</i>	56
5.2	Comportamentos das principais associações a partir das <i>deep features</i> da rede AlexNet: UCSB com a camada <i>relu2</i> ; CR com a camada <i>relu3</i> ; e LG com a camada <i>pool5</i>	57

Lista de Tabelas

3.1	Sumarização dos trabalhos relacionados, com informações sobre métodos de utilizados e desempenho obtido.	46
4.1	Detalhes sobre os conjuntos de imagens histológicas utilizados em nossa proposta.	49
4.2	Modelos CNN pré-treinados que foram considerados em nossa proposta.	51
4.3	Informações sobre as camadas das arquiteturas AlexNet e ResNet-50 investigadas em nossa proposta.	52
5.1	Combinações mais relevantes obtidas por nosso modelo para classificar as diferentes bases de imagens histológicas, com informações sobre a rede, nome da camada e o número de atributos que produziu maior taxa de acurácia.	58
5.2	Valores de áreas sob as curvas ROC (AUC) e de taxa de falsos positivos (<i>FP Rate</i>) das melhores associações combinações.	59
5.3	Taxas de acurácias (%) fornecidas pelas arquiteturas AlexNet e ResNet-50 explorando os conjuntos UCSB, CR e LG sem aumento de dados. .	60
5.4	Taxas de acurácias (%) fornecidas pelo modelo AlexNet explorando as bases UCSB, CR e LG com aumento de dados.	61
5.5	Taxas de acurácias (%) fornecidas pelas arquiteturas ResNet-50 explorando as bases UCSB, CR e LG com aumento de dados.	61

5.6	Taxas de acurácias (%) obtidas com os classificadores regularizados e explorando os mesmos conjuntos de <i>deep features</i> para avaliar o método proposto.	61
5.7	Classificação média considerando as melhores associações para conjuntos de dados UCSB, CR e LG.	63
5.8	Taxas de acurácia (%) fornecidas por diferentes abordagens para classificação de imagens histológicas da mama (UCSB).	64
5.9	Taxas de acurácia (%) definidas por diferentes abordagens para classificação de imagens histológicas colorretais (CR).	64
5.10	Valores de acurácia (%) obtidos em diferentes abordagens para classificação de gênero a partir de imagens de tecidos hepáticos (LG).	65

Sumário

1	Introdução	14
1.1	Motivação e justificativas	16
1.2	Objetivos	19
1.3	Organização da dissertação	19
2	Fundamentação teórica	20
2.1	<i>Deep learning</i>	20
2.2	Redes neurais convolucionais - CNN	22
2.2.1	<i>Pooling</i>	25
2.2.2	<i>Dropout</i>	25
2.2.3	Arquitetura de uma CNN	26
2.2.4	AlexNet	27
2.2.5	ResNet-50	28
2.3	Transferência de aprendizado	30
2.4	Seleção de atributos	31
2.4.1	<i>Relief</i>	32
2.5	Classificadores	34
2.5.1	Comitê de classificadores	35
2.6	Histologia digital	37
3	Trabalhos relacionados	38
3.1	Classificação de imagens utilizando CNN	38

3.2	CNN - Modelos híbridos	40
3.2.1	Transferência de aprendizado profundo	42
3.3	Considerações	44
4	Materiais e métodos	47
4.1	Pacotes de software e ambiente para os experimentos	47
4.2	Bases de imagens	48
4.3	Arquiteturas CNN e seleção de camadas: explorando transferência de aprendizado	49
4.4	Estratégia para selecionar e investigar <i>deep features</i>	52
4.5	Definição de um comitê de classificadores para a análise do poder dis- criminativo	54
5	Resultados e discussão	56
5.1	Comparações com técnicas de classificação e reconhecimento de padrões	59
5.2	Comparação de desempenho com a literatura: uma visão geral ilustrativa	63
6	Conclusão	66
	Referências	69

Capítulo 1

Introdução

Na área de processamento de imagens e visão computacional, a aplicação de técnicas para extração de atributos exige um tratamento especial para o processamento de dados naturais em sua forma bruta (YANG; KPALMA; RONSIN, 2008). Com isso, por décadas, a construção de um sistema de aprendizado de máquina exigiu engenharia cuidadosa ao projetar um descritor de características para transformar dados brutos, como os valores de *pixel* de uma imagem, em um vetor de valores com informações representativas o suficiente para detectar ou classificar padrões (GOODFELLOW; BENGIO; COURVILLE, 2016). Parte das dificuldades encontradas nesse processo foram minimizadas com técnicas baseadas no conceito de aprendizado profundo (do inglês *deep learning*) (LECUN; BENGIO; HINTON, 2015). O desenvolvimento de *deep learning* se concentra em fornecer conclusões a partir de dados brutos, sem a necessidade da etapa intermediária de engenharia de características (GAD, 2018).

Deep learning pode ser também entendido como modelos computacionais compostos por várias camadas de processamento com o objetivo de proporcionar uma representação dos dados em vários níveis de abstração. Esses modelos são conhecidos como redes neurais artificiais (RNA). As RNAs podem ser utilizadas para descobrir estruturas complexas em grandes conjuntos de dados. Apesar de exploradas desde 1970, o custo de treinamento das redes foi um dos principais limites impostos para sua utilização (LEE et al., 2017). Porém, com a crescente capacidade de processamento dos computadores, a aplicação de um tipo de específico de RNA, conhecido como *convolutional neural network* (CNN), permitiu avanços significativos em processamento de imagens, vídeos, áudios, descoberta de medicamentos e estruturas genéticas (LECUN; BENGIO; HINTON, 2015). De maneira complementar, a partir das propostas apresentadas por Hinton (2007) e Le (2013), o treinamento de redes neurais convolucionais tornou-se viável, pois a principal estratégia apresentada para a redu-

ção das taxas de erros de classificação, especialmente em problemas complexos, foi aumentar significativamente o número de camadas na arquitetura da rede.

Neste contexto, aumentar a profundidade da rede simplesmente empilhando camadas pode não ser uma estratégia interessante, uma vez que torna-se difícil treinar redes muito profundas em razão do problema do gradiente de fuga (HOCHREITER, 1998), que é a maneira como o gradiente é propagado entre as camadas. A multiplicação repetida pode tornar o gradiente extremamente pequeno. Como resultado, o desempenho da rede pode tornar-se saturado ou, até mesmo, degradado (ZHENG; YANG; TIAN, 2017). Por essa razão, a pesquisa com redes com um número relativamente pequeno no número de camadas mostra-se como uma tema desafiador e relevante para a área de aprendizado de máquina. Por exemplo, os trabalhos apresentados por Niu e Suen (2012), Elleuch, Maalej e Kherallah (2016), Yu e Dong (2018) superaram esse problema combinando redes convolucionais com poucas camadas, geralmente menos que 10, com métodos de seleção de atributos e classificadores externos à rede. Por outro lado, um problema que pode surgir em redes pouco profundas é a ocorrência de *overfitting*. Para evitar esta situação, Elleuch, Maalej e Kherallah (2016) propuseram um método híbrido CNN-SVM com aplicação de *droupout* para superar essa restrição.

É importante destacar também que redes muito profundas (do inglês *very deep networks*), tais como arquiteturas *Deep Residual Network* ou ResNet são outros modelos utilizados para minimizar os problemas apresentados previamente (HE et al., 2016a). Essa categoria possibilitou treinar de forma eficiente redes com centenas ou até milhares de camadas. O modelo ResNet inovou ao propor estruturas denominadas *identity shortcut connection*, conhecidas como blocos residuais. Por meio dessa técnica, ResNet permitiu o treinamento de redes profundas sem incorrer no problema do gradiente de fuga, uma vez que essas estruturas definem que uma ou mais camadas sejam puladas sem aumentar a taxa de erro de treinamento da rede. Redes profundas podem conter milhões de parâmetros treináveis, o que inviabiliza sua aplicação em algumas áreas, seja pela falta de recursos computacionais para o treinamento ou pelo número de amostras disponíveis. Entretanto, a partir de uma rede pré-treinada em uma base de imagens robusta, como o ImageNet (DENG et al., 2009), que possui atualmente mais de 14 milhões de imagens manualmente anotadas, distribuídas em mais de 20.000 categorias, tornou-se viável a aplicação de técnicas de transferência de aprendizado profundo (do inglês *deep transfer learning*) (TAN et al., 2018).

Na transferência de aprendizado profundo, uma das estratégias é a utilização dos mapas de atributos gerados pela CNN, também conhecidos como *deep features*, para a composição de vetores de atributos de alto nível (SANTOS; PONTI, 2019; COC-

CIA, 2020). A vantagem do uso deste tipo de composição está em evitar a ocorrência de *overfitting*, por meio da seleção de atributos (XIE et al., 2016), principalmente em conjuntos de imagens reduzidos. Essa prática tornou-se comum em diversas aplicações de CNNs, explorando, por exemplo, a base de imagens ImageNet (DENG et al., 2009). Assim, é possível treinar uma arquitetura de tal forma que determinadas camadas sejam capazes de reconhecer padrões específicos. As *Deep features* geradas por essas camadas especializadas podem ser extraídas e fornecer atributos de alto nível para modelos de classificação baseados em *deep learning* (MCALLISTER et al., 2018; KHAN et al., 2019; NOVITASARI et al., 2020). Neste contexto, dentre as estratégias comumente utilizadas, destacam-se a utilização de classificadores individuais para categorizar as *deep features* (NIU; SUEN, 2012; ABRAHAM; NAIR, 2020; NOVITASARI et al., 2020) e o emprego de comitê de classificadores (NANNI; LUMINI; GHIDONI, 2018; NANNI et al., 2019; ROBERTO et al., 2021). É importante destacar que o uso de comitês está entre as abordagens mais bem-sucedidas em aplicações de aprendizado de máquina, especialmente para produzir previsões mais seguras em comparação com o uso de um único classificador (SAGI; ROKACH, 2018; KASSANI; KASSANI, 2019).

1.1 Motivação e justificativas

De acordo com Zhang et al. (2017), a acurácia fornecida por uma CNN tende a ser maior com o aumento do número de camadas. Entretanto, se o problema do gradiente não for tratado, em algum momento, as camadas extras podem prejudicar o desempenho da rede. A arquitetura AlexNet, proposta por Krizhevsky, Sutskever e Hinton (2012), com um total de oito camadas, conquistou o primeiro lugar na competição anual ILSVRC¹ em 2012. Em 2015, a arquitetura vencedora foi a ResNet (HE et al., 2016a; HE et al., 2016b) com a utilização de até 152 camadas. Para reduzir o problema de *overfitting* da rede, a AlexNet utiliza a função de ativação ReLU (NAIR; HINTON, 2010) em substituição a sigmóide. Além disso, a solução encontrada pelos autores para o problema do gradiente de fuga foi aplicar o esquema *local response normalization* (LRN) antes de cada função ReLU. Por outro lado, a arquitetura ResNet utilizou o modelo de blocos residuais e, em uma de suas variações, foi a primeira CNN que superou a barreira de redes com mais de 1.000 camadas (ZHANG et al., 2017).

Neste contexto, variações na arquitetura da CNN ou a exploração de seu poder de abstração, por meio da extração de *deep features* com um classificador externo à rede,

¹ILSVRC: ImageNet Large Scale Visual Recognition Challenge

são estratégias que podem resultar em modelos eficientes e computacionalmente acessíveis (VOGADO et al., 2017; SANTOS; PONTI, 2018; DABEER; KHAN; ISLAM, 2019). Estudos presentes na literatura evidenciam que modelos híbridos, uso de seleção de atributos a partir de CNN com classificadores externos à rede, podem apresentar resultados iguais ou até melhores do que os fornecidos por um modelo específico de CNN (FU et al., 2016; XUE et al., 2016; LI et al., 2017; COCCIA, 2020). Nas abordagens híbridas, principalmente quando os classificadores são utilizados para categorizar *deep features*, uma etapa importante é a escolha do método de seleção de atributos. Os modelos disponíveis na literatura com foco em transferência de aprendizado profundo utilizam métodos reconhecidos para realizar ranqueamentos, como *Relief* e PCA, ou seleções por correlação, tais como: *deep features* por transferência de aprendizado com CFS para classificação sensoriameto remoto (YU; DONG, 2018); *deep features* por transferência de aprendizado com *Relief*, a fim de classificar imagens de ultrassom (LI et al., 2017); *deep features* fornecidas por arquiteturas ResNet e GoogleNet com PCA e *Relief*, visando a classificação de imagens radiográficas de Covid-19 (NOVI-TASARI et al., 2020).

Apesar de não existir claramente uma superioridade de um método sobre outro, a escolha de uma estratégia deve ser feita a partir da análise do domínio, número de amostras, complexidade e desbalanceamento dos dados a serem selecionados (BOLÓN-CANEDO et al., 2014). Por exemplo, o algoritmo CFS pode pesquisar exaustivamente o espaço de subconjuntos de atributos para encontrar o subconjunto mais descritivo. Entretanto, em conjuntos de atributos de alta dimensionalidades, seu uso torna-se impraticável devido a sua complexidade exponencial (ZENG et al., 2015). Para contornar este tipo de problema, uma solução é a escolha de um algoritmo com complexidade computacional menor e que permita controle sobre o número de características que devem ser obtidas (YU; DONG, 2018). Alguns modelos de seleção baseados em ranqueamento, como o algoritmo *Relief*, permitem identificar as características mais relevantes no espaço de atributos e um controle simples do número mais apropriado de atributos para a etapa de classificação (ZENG et al., 2015).

Outro aspecto importante presente no uso de transferência de aprendizado é a escolha das *deep features* que devem ser extraídas da rede neural. Por exemplo, Santos e Ponti (2018) investigaram a transferência de aprendizado utilizando *deep features* fornecidas pelas sete últimas camadas de três arquiteturas CNN (MobileNet, VGG-19 e ResNet-50). Entretanto, os autores não aplicaram nenhuma métrica para avaliar a escolha das camadas que poderiam gerar atributos de melhor qualidade. Na abordagem de Santos e Ponti (2019), os autores extraíram e concatenaram atributos obtidos a

partir de quatro camadas da ResNet-50, sendo três camadas iniciais e a penúltima camada. As motivações para essas escolhas são que as camadas iniciais ResNet-50 são responsáveis por identificar padrões locais, tais como cor, borda e forma, enquanto que camadas mais profundas generalizam padrões globais como textura e semântica.

Mesmo com os avanços indicados previamente, é possível verificar que Santos e Ponti (2019) não investigaram se o modelo poderia ser aplicado com sucesso em outras arquiteturas de redes, com uma quantidade de camadas maior ou menor, nem mesmo se apenas uma das camadas da arquitetura poderia fornecer atributos de alto nível para melhorar o desempenho de classificação. Outra constatação é que, mesmo com o uso de transferência de aprendizado profundo em imagens histológicas coradas com Hematoxilina-Eosina (H&E), tais como detecção de tumores de mama (CHOUGRAD; ZOUAKI; ALHEYANE, 2020), tumores colorretais (JAYACHANDRAN; GHOSH, 2020) e classificação de tecido hepático (ROBERTO et al., 2021), a investigação do poder discriminativo de modelos híbridos, explorando *deep features* por transferência de aprendizado, diferentes camadas de CNN e comitê de classificadores, ainda é um tema pouco explorado com o propósito de conhecer as associações mais apropriadas para o diagnóstico dessas patologias.

Além do exposto previamente, o estudo de tumores de mama, colorretais e tecidos hepáticos também justifica-se pelas taxas de incidência, mortalidade ou relevância científica. Por exemplo, de acordo com o Instituto Nacional do Câncer, o câncer de mama foi responsável por aproximadamente 2,3 milhões de novos casos em 2020, o que representa 24,5% dos casos de câncer em mulheres. É também a causa mais frequente de morte por câncer nessa população, com 684.996 óbitos estimados para o ano de 2021 (15,5% dos óbitos por câncer em mulheres) (SILVA, 2020). O câncer de cólon e reto, também conhecido como câncer colorretal, abrange os tumores que se iniciam na parte do intestino grosso (chamada cólon) e no reto (final do intestino, imediatamente antes do ânus). Esta doença é o terceiro tipo de câncer mais comum em homens e o segundo mais comum em mulheres (SIEGEL; MILLER; JEMAL, 2020). No Brasil, foram estimados 36.360 novos casos de câncer colorretal para o biênio 2018-2019 e 16.696 óbitos por ano (SILVA, 2020; GUIMARÃES et al., 2021). Em 2020, a estimativa de ocorrência foi de 78.300 novos casos para homens e 69.650 novos casos para mulheres nos Estados Unidos (SIEGEL; MILLER; JEMAL, 2020). O diagnóstico precoce desses tipos de cânceres é fundamental para o tratamento adequado e a redução no número de óbitos. Portanto, o desenvolvimento de abordagens para classificação histopatológica contribui significativamente para a detecção e estudo dessas doenças.

Mais ainda, um modelo robusto para classificação de imagens histológicas pode ser explorado em cenários além da histopatologia humana. Por exemplo, o AGEMAP (*Atlas of Gene Expression in Mouse Aging Project*) (ZAHN et al., 2007) é composto por um conjunto de imagens histológicas H&E, coletadas a partir de lâminas de tecido hepático de camundongos. O projeto tem por objetivo estudar as mudanças na expressão gênica em função da idade em camundongos. Os modelos de aprendizado de máquina podem ser utilizados para atingir esse objetivo. Embora essa combinação já tenha sido explorada na literatura (ROBERTO et al., 2021), a proposta de modelos generalistas, que forneçam altas taxas de acurácia em diferentes tipos de imagem H&E, também mostra-se como um importante desafio na pesquisa de histopatologia automatizada.

1.2 Objetivos

O objetivo principal deste trabalho foi apresentar uma estratégia para classificação e reconhecimento de padrões em imagens histológicas H&E. Para tanto, os objetivos específicos foram:

1. Obter *deep features* de imagens histológicas, via estratégia de transferência de aprendizado, de diferentes camadas das redes neurais AlexNet e ResNet-50;
2. Analisar as *deep features* mais relevantes para o processo de classificação dos grupos de imagens sob investigação, aplicando o algoritmo de ranqueamento *ReliefF*;
3. Analisar os desempenhos das *deep features* selecionadas a partir de um comitê de classificadores (*ensemble classifier*);
4. Identificar as associações mais apropriadas para classificar e reconhecer padrões em imagens histológicas de tecidos mamário, colorretal e hepático.

1.3 Organização da dissertação

No Capítulo 2 é apresentada a fundamentação teórica necessária para o desenvolvimento do modelo. Trabalhos relevantes e recentes relacionados ao tema estão no Capítulo 3. No Capítulo 4 é descrita a metodologia proposta neste trabalho. No Capítulo 5 os resultados são apresentados e discutidos. Por fim, no Capítulo 6, são apresentadas as conclusões, contribuições obtidas e uma perspectiva sobre os trabalhos futuros.

Capítulo 2

Fundamentação teórica

2.1 *Deep learning*

Técnicas de *deep learning* têm sido aplicadas a uma ampla variedade de problemas nos últimos anos, mais proeminentemente em visão computacional, processamento de linguagem natural e análise de áudio digital (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; AREL et al., 2010). Em várias dessas aplicações, os algoritmos baseados em *deep learning* ultrapassaram o desempenho anterior de algoritmos considerados estado da arte (LAI, 2015). No centro de todos os algoritmos de *deep learning* está a ideia de utilizar camadas hierárquicas de abstração aprendidas, para realizar eficientemente tarefas de alto nível (AREL et al., 2010).

É conhecido desde o início da década de 1990 que com apenas uma camada oculta, localizadas entre a camada de entrada e a camada de saída, pode-se resolver qualquer função contínua, no domínio do espaço, com qualquer precisão a partir da premissa de poder computacional, memória e dados para treinamento infinitos. Teoricamente, não haveria razão para ultrapassar duas camadas ocultas na rede. Entretanto, a maioria das aplicações modernas de *deep learning* usam mais de duas camadas ocultas para permitir que a rede modele a não linearidade no domínio do espaço (GAD, 2018).

A adição de mais camadas ocultas na arquitetura da rede torna o treinamento mais eficiente, sendo esse o principal benefício em adotar essa estratégia. Com *deep learning* é possível encontrar aproximações para funções complexas utilizando uma quantidade total de nós menor com um grau menor de liberdade na escolha dos parâmetros da rede. Outro benefício importante é que a divisão da rede em várias camadas tende a necessitar de um conjunto de dados de treinamento menor. Os algoritmos de treinamento de redes neurais se dividem em duas categorias principais, *gradient-based* e *nongradient-based*. Os métodos *gradient-based* têm sido mais comumente utilizados

nos últimos anos (LAI, 2015).

Além dos benefícios computacionais, um modelo com menor grau de liberdade (número total de parâmetros menor), requer um conjunto de dados menor para treinamento, e o tamanho do conjunto de treinamento é frequentemente um fator limitante no treinamento da rede neural. Uma rede com muitas camadas reduz a quantidade de trabalho redundante executado, pois a quantidade de parâmetros em cada camada tende a ser menor. Essa é a razão para que uma rede profunda seja mais eficiente que uma rede superficial (LAI, 2015).

Até por volta dos anos 2000, os modelos de *deep learning* foram consideradas difíceis de se treinar, devido a limitação de recursos computacionais disponível. Hinton (2009) demonstrou uma arquitetura de rede neural chamada *deep belief network* que podia ser treinada de forma não supervisionada eficientemente utilizando uma estratégia gulosa chamada *greedy layer-wise pretraining* (LAI, 2015). Até então, apenas redes convolucionais ou redes cuja profundidade resultava de recorrências eram consideradas viáveis para treinar. A estratégia proposta por Hinton (2009) poderia ser utilizada para encontrar bons parâmetros de inicialização para todas as camadas de um algoritmo de *deep learning* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Ba e Caruana (2014) mostraram que uma rede superficial pode ser treinada para ter um desempenho muito melhor do que uma rede superficial similar, treinada diretamente no conjunto de treinamento. O algoritmo proposto primeiro treina uma *deep learning* (ou um conjunto de *deep learning*) no conjunto de treinamento original e então, usa-a para fornecer rótulos estendidos para todas as entradas no conjunto de treinamento. Finalmente, uma rede superficial pode ser treinada para prever as probabilidades logarítmicas, em vez do rótulo de classe única original. Isso torna o treinamento muito mais fácil para as redes superficiais, porque os rótulos de probabilidades logarítmicas de várias classes fornecem muito mais informações do que um rótulo de classe única (LAI, 2015).

A utilização cada vez maior de *deep learning* tem sido impulsionada não só pelo desenvolvimento de algoritmos mais eficientes, como pela disponibilidade de computadores mais rápidos, com maior quantidade de memória e pela disponibilidade de conjunto de dados para treinamento maiores. Redes maiores são capazes de obter maior precisão em tarefas mais complexas (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2 Redes neurais convolucionais - CNN

Redes neurais convolucionais (do inglês *convolutional neural network*) são um tipo especializado de rede neural para processamento de dados com uma topologia conhecida semelhante a um vetor ou uma matriz, como por exemplo dados de séries temporais, que podem ser arranjados na forma de um vetor, tomando amostras em intervalos de tempo regulares, ou dados de imagem (GOODFELLOW; BENGIO; COURVILLE, 2016). Isso inclui imagens 2D e 3D e vídeos 2D, que podem ser representados como uma matriz de *pixels*. CNNs tradicionais são compostas de três tipos de camadas: camada convolucional; camada de *pooling*; e, camadas totalmente conectadas (do inglês *fully connected*) (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Camadas totalmente conectadas são frequentemente utilizadas como camadas finais para codificar informações dependentes de posição e padrões globais. O nome “rede neural convolucional” indica a presença de uma operação matemática chamada convolução (LAI, 2015). Convolução é uma operação entre duas funções f e g , que resulta em uma terceira função. A convolução é utilizada em qualquer dispositivo físico ou procedimento computacional que executa processamento para suavização ou nitidez de imagens. CNNs são uma variação das RNAs clássicas com a adição de uma operação de convolução em pelo menos uma de suas camadas aplicado a funções bidimensionais, como imagens. Também é útil para detecção de bordas, detecção de atributos, detecção de movimento, correspondência de imagens e outras tarefas (GETREUER, 2013).

Formalmente, para funções $f(x)$ e $g(x)$ no domínio contínuo da variável x , o produto de uma convolução é definido como:

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) \cdot g(x - \tau) d\tau. \quad (2.1)$$

em que, $f(x)$ contém os dados de entrada, $g(x)$ corresponde ao *kernel*, $*$ indica o produto da convolução e \cdot indica uma multiplicação ordinária.

Para funções com duas variáveis x e y (por exemplo, imagens) a definição no domínio contínuo é:

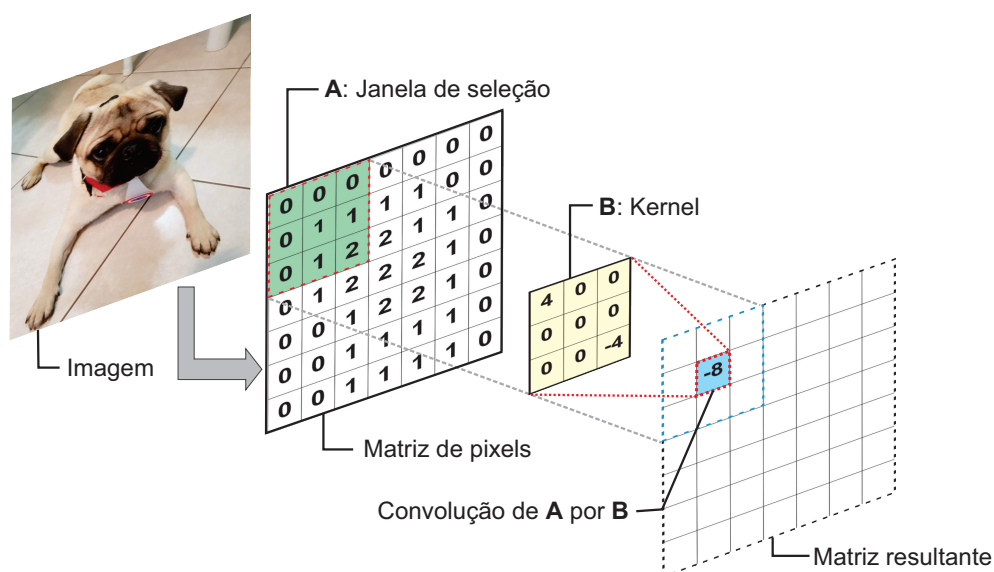
$$f(x,y) * g(x,y) = \int_{\tau_1=-\infty}^{\infty} \int_{\tau_2=-\infty}^{\infty} f(\tau_1, \tau_2) \cdot g(x - \tau_1, y - \tau_2) d\tau_1 d\tau_2. \quad (2.2)$$

Entretanto, para aplicações em algoritmos utilizamos funções discretas das variáveis x e y , ou seja, $f[x,y] * g[x,y]$. A definição é:

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1,y-n_2]. \quad (2.3)$$

Em CNNs, a terminologia adotada para $f[x,y]$ é o dado de entrada, e $g[x,y]$ o *kernel* ou núcleo da convolução. O resultado da convolução é normalmente referido como mapas de atributos (do inglês *feature maps*). Um exemplo de convolução de uma matriz 2D por um *kernel* 2D é mostrada na Figura 2.1.

Figura 2.1: Ilustração de uma convolução 2D.



Fonte: Elaborado pelo autor.

A Figura 2.1 ilustra o processo de convolução de uma matriz de *pixels* obtida a partir de uma imagem. Considerando a Equação 2.3 nesse exemplo, a janela de seleção destacada em verde corresponde à função $f[x,y]$ enquanto $g[x,y]$ define o *kernel* de tamanho 3×3 . O produto escalar, resultado da operação de convolução, é destacado em azul na matriz resultante. O processo de convolução de uma imagem é iterativo, dessa forma, a matriz de *pixels* é percorrida até o fim pela janela de seleção, comumente por meio da técnica *sliding window*, sem alterações no *kernel*. A matriz resultante desse processo possui dimensão inferior à original.

Convolução aproveita três ideias importantes que ajudam a melhorar um sistema de aprendizado de máquina: interações esparsas, compartilhamento de parâmetros e representações equivariantes. Normalmente, redes convolucionais possuem interações esparsas, também chamadas de conectividade esparsas ou pesos esparsos. Isso é possível tomando um *kernel* menor que a entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

Ao processar uma imagem, o dado de entrada pode ter milhares ou milhões de *pixels*. Entretanto, é possível detectar atributos pequenos e significativos, como bordas, por exemplo, utilizando *kernels* que ocupam apenas dezenas ou centenas de *pixels*. Isso significa que precisamos armazenar menos parâmetros, o que reduz a quantidade de memória necessária para execução e melhora sua eficiência estatística. O número de operações para calcular a saída também é reduzido. Se houver m entradas e n saídas, então a multiplicação das matrizes requer $m \times n$ parâmetros, e, na prática, o algoritmo tem complexidade $O(m \times n)$ de tempo de execução. Ao limitar o número de conexões que cada saída pode ter para k , a abordagem requer apenas $O(k \times n)$ parâmetros. Em aplicações práticas é possível obter ganhos de desempenho em tarefas de aprendizado de máquina, desde que k seja mantido em ordens de grandeza menores que m (GOODFELLOW; BENGIO; COURVILLE, 2016).

O compartilhamento de parâmetros permite utilizar os mesmos parâmetros de uma função em mais de um modelo de rede. Em redes neurais tradicionais a matriz de pesos é usada apenas uma vez ao calcular a saída de uma camada. Em uma CNN, cada elemento do *kernel* é utilizado em todas as posições da entrada. O compartilhamento de parâmetros utilizado em operações de convolução significa que, em vez de aprender um conjunto separado de parâmetros para cada local, é aprendido apenas um conjunto. Isso não afeta o tempo de execução de *forward propagation*, dado por $O(k \times n)$. Entretanto, esta estratégia reduz os requisitos de armazenamento do modelo para k parâmetros (GOODFELLOW; BENGIO; COURVILLE, 2016).

Na convolução, essa forma particular de compartilhamento de parâmetros faz com que a camada tenha uma propriedade chamada equivariância. Dizer que uma função é equivariante significa que, se a entrada for alterada, a saída é alterada da mesma maneira. Dessa forma, uma função $f(x)$ é equivariante para uma função $g(x)$ se $f(g(x)) = g(f(x))$.

CNNs têm um papel importante na história do *deep learning*. Elas são um exemplo chave de aplicação bem-sucedida de *insights* obtidos pelo estudo do cérebro em aplicações de aprendizado de máquina. Foram também um dos primeiros modelos a mostrar bom desempenho muito antes de outros modelos de *deep learning* serem considerados viáveis. As CNNs foram as primeiras redes neurais a resolver importantes problemas de classificação de imagens e permanecem na vanguarda de problemas envolvendo reconhecimento de objetos em imagens.

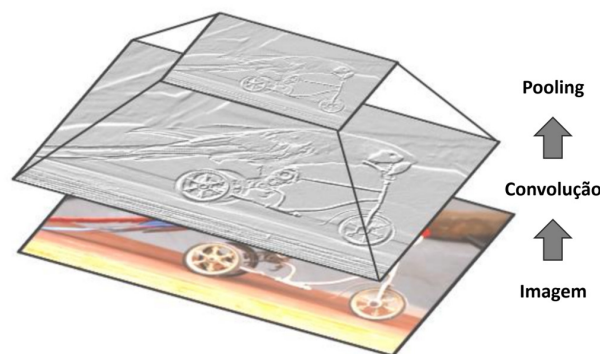
2.2.1 Pooling

Uma típica camada de uma CNN consiste em três estágios. No primeiro estágio a camada executa várias convoluções em paralelo para produzir um conjunto de ativações lineares. No segundo estágio, cada ativação linear é submetida a uma função de ativação não linear. No terceiro estágio, uma função de *pooling*, também chamada de *subsampling*, é utilizada para modificar a saída para a próxima camada.

A função do *pooling* é substituir o resultado da saída por um conjunto menor contendo estatísticas resumidas do conteúdo original, reduzindo a dimensionalidade do dado. Por exemplo, a operação *max pooling* retorna o máximo de uma saída dentro de uma vizinhança. Outra função de *pooling*, *average pooling*, consiste em utilizar o cálculo da média de uma vizinhança. ResNet é um exemplo de rede que faz uso de *average pooling* em sua arquitetura. Em todos os casos, *pooling* permite manter uma correta representação da saída mesmo que existam pequenas invariâncias rotacionais e translacionais na entrada (BOUREAU; PONCE; LECUN, 2010).

Como a utilização de *pooling* reduz o tamanho de saída de uma camada, o número de cálculos da camada seguinte será menor, isso contribui para um melhor desempenho da rede. Uma representação gráfica do comportamento da aplicação das técnicas de convolução e *pooling* sobre uma imagem é mostrada na Figura 2.2.

Figura 2.2: Ilustração de uma operação de convolução e *pooling* sobre uma imagem.



Fonte: Adaptado de Alom et al. (2018).

2.2.2 Dropout

Hinton et al. (2012) propuseram uma técnica que melhora o desempenho da rede nos casos em que há dados de treinamento limitados. A ideia deriva da observação de que, quando o conjunto de treinamento é pequeno, haverá muitos modelos possíveis que terão um bom desempenho no conjunto de treinamento, mas apenas alguns

terão um bom desempenho no conjunto de testes. A solução tradicional para isso é treinar um conjunto de redes a partir de diferentes inicializações e/ou subconjuntos do conjunto de treinamento, depois combinar suas saídas para produzir a saída final.

Essa abordagem funciona bem na melhoria do desempenho do modelo, mas muitas vezes não é computacionalmente viável em *deep learning*, em que o treinamento de uma única rede pode levar muitas horas ou dias, mesmo utilizando uma GPU considerada moderna. Muitas redes neurais treinadas também são utilizadas para aplicações em tempo real e o uso de um conjunto de treinamentos tornaria o modelo demasiadamente lento para aplicação (LAI, 2015).

A proposta de Hinton et al. (2012) chamada *dropout* fornece um método computacionalmente barato, porém, poderoso para regularizar uma ampla família de modelos de redes neurais, incluindo redes convolucionais. O algoritmo consiste em atribuir aleatoriamente com uma probabilidade de 0,5 a chance de cada nó ser multiplicado por zero durante a execução do *back propagation*, dessa forma, esse nó não participa dessa etapa. A aplicação desta técnica é muito comum na etapa de treinamento de redes neurais de grande porte, diminuindo o custo de processamento e armazenamento na etapa de treinamento com um efeito positivo de reduzir o *overfitting* no treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016; HINTON et al., 2012).

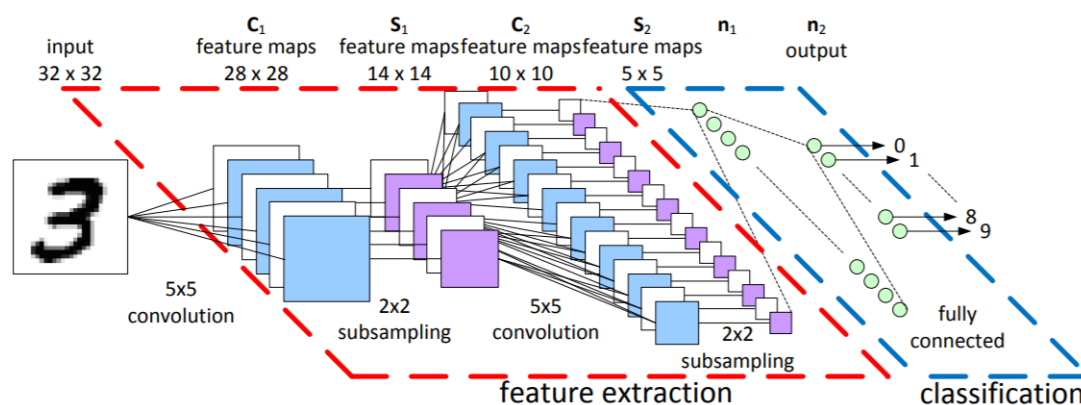
2.2.3 Arquitetura de uma CNN

As CNNs foram algumas das primeiras *deep learning* em funcionamento treinadas com *back propagation*. Não está totalmente claro o motivo das CNNs terem sido bem-sucedidas enquanto outras redes falharam (LAI, 2015). Um dos possíveis motivos se deve ao fato das CNNs serem computacionalmente mais eficientes do que redes totalmente conectadas, por isso, foi mais fácil executar vários experimentos com elas e ajustar sua implementação (GOODFELLOW; BENGIO; COURVILLE, 2016; LAI, 2015).

Uma arquitetura de exemplo de uma CNN tradicional é mostrada na Figura 2.3. Esta rede é utilizada para reconhecimento de dígitos manuscritos. As primeiras camadas da rede C_1 até S_2 funcionam como um extrator de atributos treinável. Todas as camadas de rede contêm modelos de neurônios como em uma rede clássica *multi-layer perceptron* (MLP). As camadas de extração de atributos C_1 até S_2 possuem restrições específicas, como conectividade local e compartilhamento de pesos. Com essas restrições, as primeiras camadas são capazes de extrair atributos invariantes de posição a partir de formas bidimensionais. As camadas de classificação n_1 e n_2 na saída são MLPs totalmente conectadas. Essas camadas usam os atributos locais extraídos para

executar a classificação da imagem de entrada (PEEMEN; MESMAN; CORPORAAL, 2011).

Figura 2.3: Arquitetura de uma CNN para tarefa de reconhecimento de dígitos escritos a mão.



Fonte: Peemen, Mesman e Corporaal (2011).

As camadas de convolução são sucedidas por uma camada de *pooling* para executar uma operação de redução dos dados resultantes da camada convolucional. Neste exemplo, essa operação de redução de dados é feita pela média local em uma janela predefinida sem sobreposição. O tamanho da janela de média é descrito pelo fator de sub-amostragem (*subsampling*) S (PEEMEN; MESMAN; CORPORAAL, 2011).

Na Figura 2.3, a classificação da imagem de entrada da CNN é realizada nas camadas de saída, como n_1 e n_2 . Nessas camadas, todos os neurônios têm um conjunto único de pesos, o que permite que eles detectem atributos complexos e realizem a classificação. Uma importante propriedade da arquitetura de uma CNN é que todos os valores de pesos e vieses sinápticos podem ser treinados de modo eficiente por meio do algoritmo *back propagation* (PEEMEN; MESMAN; CORPORAAL, 2011).

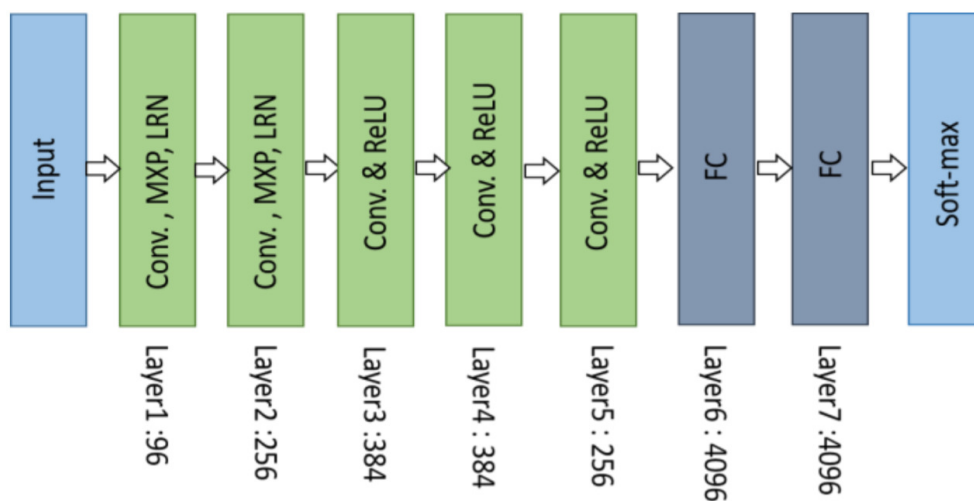
2.2.4 AlexNet

Em Krizhevsky, Sutskever e Hinton (2012), os autores propuseram um modelo de CNN profundo e amplo que venceu o desafio mais difícil do ImageNet para reconhecimento de objetos visuais chamado *ImageNet Large Scale Visual Recognition Challenge* - ILSVRC, em 2012. AlexNet alcançou alta precisão de reconhecimento de todas as abordagens tradicionais de aprendizado de máquina e visão computacional, com uma taxa de erro médio de 25% na classificação de objetos. Foi um avanço significativo no campo de aprendizado de máquina e visão computacional para tarefas de reconhecimento e classificação visual. É o ponto da história em que o interesse por

deep learning aumentou rapidamente (ALOM et al., 2018). Apesar do progresso nos anos seguintes com outras arquiteturas, AlexNet estabeleceu-se como um *benchmark* para a comparação com outros modelos.

A arquitetura da AlexNet é apresentada na Figura 2.4. A primeira camada convolucional realiza convolução e *max pooling* com *local response normalization*, onde são utilizados 96 filtros receptivos diferentes, com tamanho 11×11 . As operações de *max pooling* são realizadas com filtros 3×3 com um tamanho de passo igual a 2. As mesmas operações são realizadas na segunda camada com filtros 5×5 . Filtros 3×3 são usados na terceira, quarta e quinta camadas convolucionais com mapas de atributos 384, 384 e 256, respectivamente. Duas camadas totalmente conectadas são usadas com *dropout*, seguido por uma camada *softmax* ao final.

Figura 2.4: Arquitetura AlexNet: convolução, *max pooling*, LRN, FC e *softmax*.



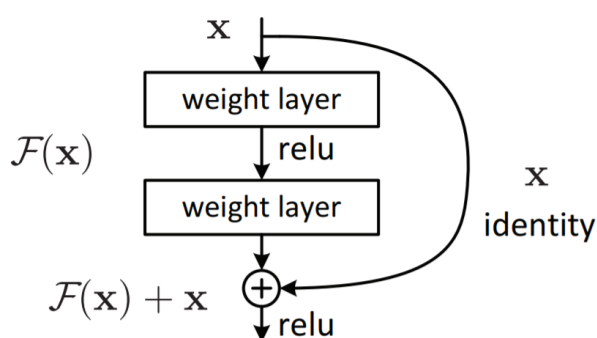
Fonte: Alom et al. (2018).

2.2.5 ResNet-50

Na pesquisa em *deep learning*, existe um consenso que quanto mais profundo o modelo, melhor desempenho ele pode alcançar no desafio ImageNet. Desde a vitória da rede AlexNet na competição ILSVRC-2012, a arquitetura *Deep Residual Network* ou ResNet (HE et al., 2016a) tem sido avaliada como um dos trabalhos mais inovadores em visão computacional. A ResNet torna possível treinar redes com centenas ou até milhares de camadas. Por meio da utilização de blocos residuais, ResNet permite o treinamento de redes muito profundas sem incorrer no problema do gradiente de fuga.

Antes da ResNet, outras técnicas foram utilizadas para lidar com o problema do gradiente de fuga, por exemplo, ao acrescentar uma perda auxiliar na camada intermediária como supervisão extra, entretanto, nenhuma técnica até então parecia resolver definitivamente o problema (SZEGEDY et al., 2015). A utilização de blocos residuais permite que ResNet seja capaz de pular uma ou mais camadas na excussão. A estrutura de um bloco residual pode ser observada na Figura 2.5. O principal benefício de treinar uma rede residual é que mesmo em redes muito profundas, a taxa de erro no treinamento não aumenta.

Figura 2.5: Estrutura de um bloco residual do modelo ResNet-50.

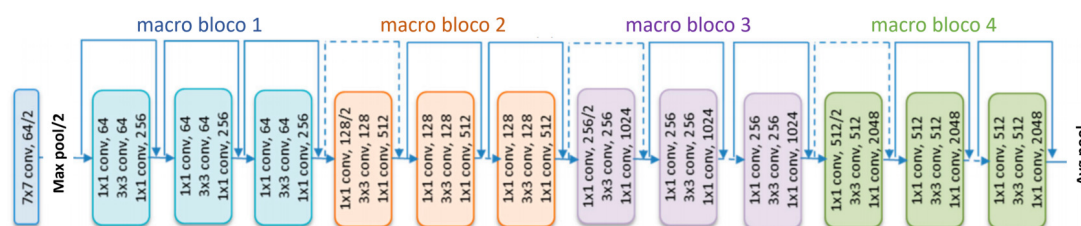


Fonte: He et al. (2016a).

Na Figura 2.5, dada uma entrada x , o mapeamento residual $\mathcal{F}(x)$ é denotado por $\mathcal{F}(x) + x$, ou seja, esse mapeamento soma a entrada de um bloco residual (x identity) à saída resultante desse mesmo bloco. Essas conexões de atalho não adicionam nenhum parâmetro extra na complexidade computacional do modelo.

ResNet possui uma quantidade de camadas de convolução empilhadas maior em comparação ao modelo AlexNet. Os autores da arquitetura ResNet, He et al. (2016a), afirmam que os atributos podem ser melhor representados de acordo com o aumento do número de camadas empilhadas, por isso, essa maior profundidade. Outra característica da ResNet é possuir apenas uma camada de *max pooling* logo após a execução da primeira convolução e uma camada *average pooling* na antepenúltima camada, mais precisamente antes da camada totalmente conectada. A Figura 2.6 ilustra a arquitetura de uma rede ResNet-50.

Figura 2.6: Arquitetura ResNet-50.



Fonte: Adaptado de Rahhal et al. (2018).

Na Figura 2.6, o macro bloco 1 contém três blocos residuais, totalizando nove camadas de convolução com a mesma resolução espacial e três conexões de atalho (uma em cada bloco residual). A saída do último bloco residual do macro bloco 1 é conectada, por meio da conexão de atalho (ilustrada pelas linhas pontilhadas da Figura 2.6) à saída do primeiro bloco residual do macro bloco 2. Esse padrão de conexão entre os macro blocos permanece o mesmo para todos eles. O macro bloco 2 possui quatro blocos residuais, com *kernels* $1 \times 1 \times 128$, $3 \times 3 \times 128$ e $1 \times 1 \times 512$, totalizando 12 camadas de convolução.

O macro bloco 3, por sua vez, possui seis blocos residuais, com *kernels* $1 \times 1 \times 256$, $3 \times 3 \times 256$ e $1 \times 1 \times 1024$, com o total de 18 camadas de convolução. O macro bloco 4 possui a mesma quantidade de blocos residuais que o macro bloco 1, entretanto a quantidade de filtros é diferente, sendo ela 512, 512 e 2.048. Após o último macro bloco, uma camada de *average pooling* (*avg pool*) realiza um *subsampling* no mapa de atributos. Por fim, a última camada do modelo ResNet é uma camada totalmente conectada, com um vetor de atributos $1 \times 1 \times 1.000$ (não exibida na Figura 2.6), onde 1.000 refere-se as classes presentes na base de dados ImageNet (DENG et al., 2009). A função de classificação *softmax* é aplicada nessa camada totalmente conectada, gerando as probabilidades referentes a cada uma das 1.000 classes originalmente previstas na arquitetura.

2.3 Transferência de aprendizado

Transferência de aprendizado (do inglês *transfer learning*) é um método de aprendizado de máquina em que um modelo desenvolvido para uma tarefa é reutilizado como ponto de partida para uma segunda tarefa. Embora a maioria dos algoritmos de aprendizado de máquina seja projetada para lidar com tarefas únicas, o desenvolvimento de algoritmos que facilitam a transferência de aprendizado é um tópico de interesse contínuo na pesquisa em aprendizado de máquina (TORREY; SHAVLIK, 2010). Em alguns

domínios, é relativamente difícil construir conjuntos de dados anotados em larga escala devido aos custos de aquisição e anotação dos dados. A transferência de aprendizado relaxa a hipótese de que os dados de treinamento devem ser independentes e distribuídos de forma idêntica com os dados de teste e validação. Essa característica motiva a utilização de transferência de aprendizado para resolver problemas onde dados de treinamento suficientes não estão disponíveis.

Como uma nova plataforma de classificação, o aprendizado profundo tem se mostrado uma abordagem poderosa e popular, onde modelos pré-treinados de várias fontes são utilizados como ponto de partida para tarefas de visão computacional. Dado que os recursos computacionais necessários para desenvolver e treinar grandes modelos de redes neurais podem ser inacessíveis, pesquisadores tem aplicado aprendizado profundo com sucesso em diversos domínios (TAN et al., 2018).

A rigor, de acordo com Tan et al. (2018), para definir transferência de aprendizado e transferência de aprendizado profundo é necessário fornecermos as definições de um domínio e uma tarefa. Um domínio pode ser representado por $\mathcal{D} = \{\mathcal{X}, \mathcal{P}(X)\}$, que contém duas partes: o espaço de características \mathcal{X} e a distribuição de probabilidade da aresta $\mathcal{P}(X)$, onde $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Uma tarefa pode ser representada por $\mathcal{T} = \{y, f(x)\}$, que consiste em duas partes: espaço de rótulos y e a função alvo de previsão $f(x)$. $f(x)$ também pode ser considerado como uma função de probabilidade condicional $P(y|x)$. Então, a transferência de aprendizado pode ser formalmente definida das seguintes formas:

Definição 1 (transferência de aprendizado): Dada uma tarefa de aprendizado \mathcal{T}_t baseada em \mathcal{D}_t . A transferência de aprendizado visa melhorar o desempenho da função preditiva $f_{\mathcal{T}}(\cdot)$ para a tarefa de aprendizagem \mathcal{T}_t ao descobrir e transferir conhecimento latente de \mathcal{D}_s e \mathcal{T}_s , onde $\mathcal{D}_s \neq \mathcal{D}_t$ e/ou $\mathcal{T}_s \neq \mathcal{T}_t$. Além disso, na maioria dos casos, o tamanho de \mathcal{D}_s é muito maior que o tamanho de \mathcal{D}_t , $N_s \gg N_t$.

Definição 2 (transferência de aprendizado profundo): Dada uma tarefa de aprendizado de transferência definida por $\langle \mathcal{D}_s, \mathcal{T}_s, \mathcal{D}_t, \mathcal{T}_t, f_{\mathcal{T}}(\cdot) \rangle$. É uma tarefa de transferência de aprendizado profundo, em que $f_{\mathcal{T}}(\cdot)$ é uma função não linear que reflete uma rede neural profunda.

2.4 Seleção de atributos

Deep features extraídas a partir de uma ou mais camadas de uma CNN podem ser utilizadas para compor um vetor de atributos, como apresentado em Elleuch, Maalej e Kherallah (2016). Entretanto, uma vez que os mapas de atributos de uma CNN

são constituídos por milhares de características, a atenção a quantidade de atributos utilizados é fundamental para evitar *overfitting* na classificação e, assim, garantir a viabilidade do modelo. Muitas vezes, técnicas de seleção de atributos são aplicadas para reduzir a dimensionalidade do conjunto de atributos escolhidos. A seleção de atributos pode ser entendida como o método utilizado para encontrar o subconjunto de atributos de um determinado tamanho que leva a maior generalização possível ou equivalente, a um risco mínimo (LAL et al., 2006).

O termo dimensionalidade é atribuído ao número de atributos de uma representação de padrões, ou seja, a dimensão do espaço de características. As duas principais razões para que a dimensionalidade seja a menor possível são: recursos computacionais disponíveis e precisão do classificador. Quando o espaço de características contém somente atributos de alto nível, a etapa de classificação será mais rápida e ocupará menos memória. Além disso, quando o conjunto de exemplos de treinamento não é grande, um espaço de características pequeno pode evitar a maldição da dimensionalidade e propiciar pequenas taxas de erro ao classificador. Além da necessidade de utilizar a menor dimensionalidade possível, há outro fator analisado pelo teorema do “patinho feito” proposto por Watanabe (1985), que diz que é possível fazer dois padrões arbitrários ficarem similares se esses forem codificados com um número suficientemente grande de atributos similares. Isso enfatiza a necessidade de uma escolha cuidadosa dos atributos.

A redução da dimensionalidade é uma importante etapa no processamento de dados de alta dimensão (BARTL; REZANKOVA; SOBISEK, 2011; ZHANG; JIANG; YE, 2010). Os métodos para seleção de atributos são agrupados em três categorias principais: filtro, *wrapper* e *embedded*. Atualmente, na aplicação com transferência de aprendizado profundo, a categoria filtro destaca-se como a mais comumente encontrada na literatura, principalmente via *Relief* e PCA (LI et al., 2017; YU; DONG, 2018; COCCIA, 2020). O algoritmo *Relief* se destaca por ser considerado um algoritmo simples, eficaz e tolerante a ruídos (KIRA; RENDELL, 1992; URBANOWICZ et al., 2018).

2.4.1 *Relief*

A seleção de atributos desempenha um papel crítico para a extração de conhecimento. O uso de estratégias para selecionar atributos foi impulsionado pelo aumento do espaço de características sob investigação. Os *relief-based algorithms* (RBAs) constituem uma família de algoritmos para selecionar atributos. Esta família pertence a categoria de filtros (URBANOWICZ et al., 2018). Neste contexto, *Relief* é um algo-

ritmo da categoria de filtros e está fundamentado no aprendizado baseado em instância, calculando a estatística de qualidade ou relevância de cada instância. A estatística de uma característica é referida como $W[A]$, peso da característica A ou como *score*, que pode variar de -1 (pior) a $+1$ (melhor) (URBANOWICZ et al., 2018).

No pseudocódigo ilustrado no Algoritmo 1, o método *Relief* avança por m instâncias aleatórias de treinamento (R_i), selecionadas sem substituições, em que m é definido pelo usuário. A cada ciclo, R_i é a instância alvo e o vetor de *score* de características W é atualizado, baseado no valor das diferenças das características das instâncias alvo e da vizinhança. Assim, a cada ciclo, a distância entre a instância alvo e todas as outras instâncias é calculado. A abordagem *Relief* identifica as duas instâncias vizinhas mais próximas do alvo; uma da mesma classe, denominada *nearest hit* (H) e outra da classe oposta, denominada *nearest miss* (M). A última etapa do ciclo atualiza o peso de uma característica A em W , caso o valor da característica se diferencie da instância alvo R_i e também de *nearest hit* (H) ou *nearest miss* (M). A lógica para atualizar os pesos consiste no fato que um bom atributo deve ter o mesmo valor para instâncias a partir da mesma classe (subtraindo a diferença $diff(A, R_i, H)$) e deve diferenciar entre as instâncias de diferentes classes (adicionando a diferença $diff(A, R_i, M)$) (KIRA; RENDELL, 1992).

Algoritmo 1 Proposta original do método de seleção *Relief*

- 1: $n \leftarrow$ número de instâncias de treinamento
 - 2: $a \leftarrow$ número de características
 - 3: Parâmetro: $m \leftarrow$ número de instâncias de treinamento aleatórias de n utilizadas para atualizar W
 - 4: Inicializar todos os pesos $W[A] = 0,0$
 - 5: **para** $i = 1$ até m **faça**
 - 6: Aleatoriamente selecionar uma instância “alvo” R_i
 - 7: Encontra um *nearest hit* “H” e *nearest miss* “M” (instâncias)
 - 8: **para** $A=1$ até a **faça**
 - 9: $W[A] = W[A] - diff(A, R_i, H)/m + diff(A, R_i, M)/m$
 - 10: **fim para**
 - 11: **fim para**
 - 12: **retorna** o vetor W de *scores* que estimam a qualidade das características
-

ReliefF (ROBNIK-ŠIKONJA; KONONENKO, 1997) é o RBA mais conhecido e utilizado atualmente (URBANOWICZ et al., 2018). Trata-se de uma expansão do algoritmo *Relief* para possibilitar análises de duas ou mais classes e grandes conjuntos de dados. Notavelmente, o “F” em *ReliefF* refere-se à sexta variante do algoritmo (de A a F). Este método considera uma aproximação de probabilidades mais confiável, uma vez que analisa a contribuição média de características vizinhas. O algoritmo

ReliefF consiste em um processo iterativo para definir as instâncias de dados mais próximas da mesma classe e de outras classes. A saída do algoritmo *ReliefF* consiste em um vetor de atributos com os pesos, ranqueando os atributos em ordem de importância (URBANOWICZ et al., 2018).

2.5 Classificadores

Em acordo com Han, Pei e Kamber (2011), classificadores para construção de modelos de classificação devem ser selecionados de acordo com critérios tais como, precisão do modelo, velocidade, robustez, escalabilidade e interpretabilidade. As taxas de acerto na etapa de classificação para um conjunto de dados podem variar de acordo com a escolha do algoritmo para esta finalidade. Uma estratégia para determinação dos algoritmos adequados na etapa de classificação é o teste do vetor de atributos em diferentes categorias de classificadores, observando métricas de desempenho tais como, área sob a curva ROC e/ou acurácia (TAN; STEINBACH; KUMAR, 2009).

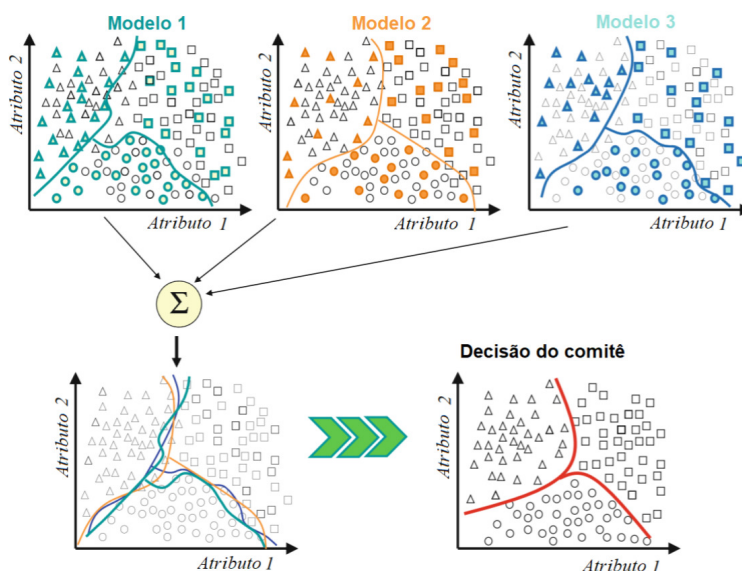
Duas abordagens básicas são utilizadas para as técnicas de classificação: supervisionada e não supervisionada. As etapas descritas por Conci, Azevedo e Leta (2008) para classificação supervisionadas podem ser resumidas em: (1) escolha de um conjunto de treinamento; (2) escolha dos parâmetros utilizados na discriminação; (3) obtenção de uma função discriminante; (4) eliminação de parâmetros irrelevantes; (5) realização de testes com elementos fora do conjunto de treinamento; e (6) atualização do dados ou realimentação. A classificação não-supervisionada é utilizada para contextos em que não há informações sobre os rótulos ou classes. Dessa forma, tenta-se estabelecer *clusters* ou agrupamentos naturais no espaço de características, a partir da medida de similaridade dos dados (RIBEIRO et al., 2019).

Segundo Tan, Steinbach e Kumar (2009), os classificadores supervisionados podem ser agrupados pelas distintas abordagem de execução da tarefa, tais como, vizinhos mais próximos, funções ou regras, bayesianos, árvores de decisão e vetores de suporte. Uma vez estabelecidas as classes onde o conjunto de atributos apresenta os melhores resultados, outros algoritmos da mesma categoria poderão ser testados e ter suas métricas avaliadas a fim de refinar o processo de classificação. Alguns exemplos de algoritmos das classes citadas são: *KStar* (K^*) (CLEARY; TRIGG, 1995), *Logistic Discrimination* (CESSIE; HOUWELINGEN, 1992), *Naive Bayes* (JOHN; LANGLEY, 2013), *Random Forest* (BREIMAN, 2001) e *Support Vector Machine* (SVM) (ALPAYDIN, 2009).

2.5.1 Comitê de classificadores

Um comitê de classificadores (do inglês *classifier ensemble*) deve ser entendido como um sistema de classificação, composto por múltiplos classificadores, os quais são treinados individualmente. Assim, as decisões são combinadas a partir de uma regra (SAGI; ROKACH, 2018). Dessa forma, os resultados produzidos por dois ou mais classificadores são unidos e, a partir dessa união, uma decisão é tomada com base no comportamento de todos os classificadores empregados. Essa estratégia permite melhorar as previsões, tornando-as mais confiáveis. Uma ilustração da aplicação dessa técnica é mostrada da Figura 2.7.

Figura 2.7: Ilustração do comportamento de um comitê de três classificadores em um conjunto de dados descrito por dois atributos.



Fonte: Adaptado de Polikar (2006).

Entre os principais fatores que motivam o uso de comitês de classificadores destacam-se a capacidade de evitar *overfitting* e de adaptar-se melhor à investigação de conjuntos com classes desbalanceadas (SAGI; ROKACH, 2018). Adicionalmente, Polikar (2006) agrupa as razões teóricas e práticas que motivam a utilização de comitês em quatro categorias principais:

1. Razões estatísticas: Um conjunto de classificadores com desempenhos de treinamento semelhantes pode ter desempenhos de generalização diferentes. Particularmente, nos casos em que o conjunto de dados de teste utilizados para determinar o desempenho de generalização não for suficientemente representativo dos dados. Combinar as saídas de vários classificadores pela média pode reduzir o risco de uma seleção incorreta de um classificador de baixo desempenho;

2. Grandes volumes de dados: Em certas aplicações, a quantidade de dados a ser analisada pode ser muito grande para ser tratada com eficácia por um único classificador. Particionar os dados em subconjuntos menores, treinar diferentes classificadores com diferentes partições de dados e combinar suas saídas por meio de uma regra de combinação inteligente costuma ser uma abordagem mais eficiente;
3. Poucos dados: Na ausência de dados de treinamento adequados, técnicas de reamostragem podem ser utilizadas para desenhar subconjuntos aleatórios sobrepostos dos dados disponíveis, cada um dos quais pode ser usado para treinar um classificador diferente, criando o conjunto mais representativo;
4. Dividir para conquistar: Independentemente da quantidade de dados disponíveis, certos problemas são muito difíceis para um determinado classificador resolver. Por exemplo, um classificador linear não é capaz de classificar dados não lineares complexos, entretanto, uma combinação adequada de vários classificadores lineares pode ser utilizada para modelar esse limite não linear.

As regras de combinação comumente utilizadas no uso de comitês de classificadores podem ser diferenciadas entre métodos de Bayes ingênuos e métodos de votação (ZHANG; ZHOU, 2011). A regra da soma, a regra do produto, a regra do máximo, a regra do mínimo e a regra da média são exemplos de abordagens Bayes ingênuos. Por sua vez, os principais métodos de votação são o voto simples e o voto majoritário. As definições e equações de cada uma das técnicas descritas previamente podem ser consultadas nos trabalhos de Kittler et al. (1998) e Li e Wang (2009). É importante destacar também que na investigação conduzida por Kittler e Alkoot (2003), os autores fazem uma análise sistemática em conjuntos de dados sintéticos e experimentos com dados reais entre a regra da soma e votação. O estudo conclui que, para problemas envolvendo duas classes, ainda que regra de decisão de Bayes incorra em erros nas estimativas das probabilidades de cada classe, a combinação pela regra da soma supera o método de votação. Mais ainda, a composição de um robusto comitê de classificadores por meio de uma regra de decisão eficiente pode gerar resultados relevantes em aplicações de visão computacional no campo de histopatologia em classificações por meio de CNN (NANNI; LUMINI; GHIDONI, 2018).

2.6 Histologia digital

A coloração Hematoxilina-Eosina, comumente conhecida como coloração H&E, é uma das principais técnicas para tingir tecidos e viabilizar análises por meio da microscopia óptica (TITFORD, 2005). Essa técnica é a mais amplamente utilizada em diagnósticos médicos e adotada como padrão ouro em histopatologia (TITFORD, 2005; SMITH, 2006; DAPSON; HOROBIN, 2009). Por exemplo, quando um patologista examina uma biópsia de tecido com suspeita de câncer, é provável que o tecido contido na lâmina tenha sido tingido com H&E (ROSAI, 2007).

A histologia digital envolve a aquisição, gerenciamento, compartilhamento e interpretação de informações de histologia, incluindo imagens e dados, em um ambiente digital. Os slides digitais são criados quando as lâminas de vidro são capturadas com um dispositivo de digitalização, para fornecer uma imagem digital de alta resolução e que pode ser visualizada na tela de um computador ou de um dispositivo móvel (BIOSYSTEMS, 2021). Ferramentas automatizadas para análises desse tipo de imagem também podem ser utilizadas para auxiliar na interpretação e quantificação da expressão de biomarcadores em seções dos tecidos (SRINIDHI; CIGA; MARTEL, 2020). Assim, a histopatologia digital oferece benefícios que não são facilmente alcançados apenas com lâminas de vidro. Entre os avanços enumerados por Biosystems (2021), destacam-se:

1. Análise aprimorada: A utilização de algoritmos de aprendizado de máquina podem tornar o processo de diagnóstico mais rápido e preciso. O armazenamento digital das imagens permite fácil acesso a casos anteriores, compartilhamento com outros especialistas e comparação lado a lado com imagens semelhantes, resultando em análises mais acuradas;
2. Redução de erros: Uma vez digitalizadas, os erros por quebra, degradação ou manipulação incorreta das lâminas de vidro deixam de existir. A rotulagem das lâminas digitais tendem a ser mais confiáveis que as feitas manualmente;
3. Melhor visualização: Equipamentos de digitalização permitem a captura da imagem a partir de vários ângulos, níveis de magnificação (*zoom*) e a anotação de regiões de interesse.

As principais abordagens em *deep learning* para classificação e segmentação de imagens histológicas utilizam bases de imagens de histológicas coradas com H&E (SRINIDHI; CIGA; MARTEL, 2020).

Capítulo 3

Trabalhos relacionados

Neste capítulo é apresentada uma seleção de artigos que realizam classificação de imagens por meio de *deep learning* em diferentes contextos e bases de imagens. Na seção 3.1, encontram-se trabalhos relacionados a classificação de imagem utilizando CNN, a seção 3.2 apresenta trabalhos relevantes na aplicação de modelos híbridos de CNN. Trabalhos recentes relacionados a transferência de aprendizado profundo que suportam a metodologia proposta nesse trabalho são apresentados na subseção 3.2.1. Algumas considerações sobre esse capítulo encontram-se na seção 3.3.

3.1 Classificação de imagens utilizando CNN

A partir do trabalho publicado por Krizhevsky, Sutskever e Hinton (2012), as estratégias de estado da arte para classificação de imagens tem aplicado técnicas de *deep learning* para classificação de imagens em diversas áreas, incluindo imagens médicas.

Em Litjens et al. (2016) os autores foram capazes de identificar metástase de câncer de mama a partir de amostras de linfonodos sentinela. Descobriu-se que todas as lâminas contendo micro e macro-metástases de câncer de mama puderam ser identificadas automaticamente, enquanto de 30 a 40% das lâminas que continham tecido normal e benigno puderam ser excluídas sem o uso de marcadores imuno-histoquímicos adicionais ou intervenção humana. O estudo utilizou imagens de lâminas digitalizadas coradas com H&E em que as metástases foram delineadas manualmente, essas imagens foram então utilizadas para treinar uma CNN para detectar áreas de câncer em conjuntos de dados de validação. Os autores afirmam que, durante o treinamento, a estrutura da rede foi continuamente aprimorada em relação ao número total de camadas e a quantidade máxima de filtros por camada a fim de obter o máximo desempenho no conjunto de validação, resultando em um total de oito camadas na arquitetura. Após

a etapa de treinamento, a CNN foi convertida em uma rede totalmente conectada fornecendo previsões da presença de câncer e metástases em conjuntos de teste. Para a análise quantitativa, foi utilizada a área sob a curva ROC para avaliar o desempenho. Observou-se que, a um custo de uma ou duas detecções de falso-positivo, 90% de todas as micro metástases e 93% das imagens de macro metástases individuais puderam ser identificadas. A AUC calculada foi 0,99.

No trabalho de Araújo et al. (2017) os autores propuseram um método para classificação de imagens histológicas de câncer de mama coradas com H&E utilizando uma CNN. A arquitetura da rede foi projetada com um total de 13 camadas para a classificação de quatro classes: tecido normal, lesão benigna, carcinoma invasivo (responsável por 80% dos cânceres de mama) e carcinoma *in situ*. Alternativamente uma classificação binária também é realizada pela rede: carcinoma ou não carcinoma. O conjunto de dados de treinamento é composto por 249 imagens balanceadas em quatro classes. As imagens foram normalizadas utilizando o método proposto por Macenko et al. (2009). A sensibilidade geral da CNN proposta para detecção de carcinoma invasivo foi de 74,1%. Atributos extraídos da CNN foram utilizados para treinar um classificador SVM em que foram alcançadas precisões de 77,8% para quatro classes e 83,3% para carcinoma versus não carcinoma. A sensibilidade do método proposto para a identificação do câncer foi de 95,6%. O desempenho do sistema proposto foi semelhante ou superior a outros trabalhos relacionados como Spanhol et al. (2016) e Cruz-Roa et al. (2014), mesmo utilizando um conjunto de dados de treinamento menor e uma maior quantidade de classes.

Dabeer, Khan e Islam (2019) propuseram uma CNN para classificação de imagens histológicas coradas com H&E para detecção de câncer de mama. Os autores elegeram a base de dados BreakHis¹, composta por 2.480 imagens benignas e 5.429 malignas, coletadas no Brasil por P&D Lab, de janeiro a dezembro de 2014. A rede é composta três camadas convolucionais e três camadas de *pooling*. O processo de classificação é realizado por meio de três camadas totalmente conectadas. Durante a etapa de validação a acurácia foi de 93,45%. Testes com imagens de diferentes resolução de imagem histopatológica foram executados no modelo e a rede mostrou-se insensível a resolução. Na etapa de testes, o método proposto forneceu uma acurácia de classificação de 99,86%. Os autores afirmam que esse método de baixo custo de implementação pode ser utilizado para detecção de câncer nos estágios iniciais, o que pode aumentar a taxa de sobrevivência entre pacientes com câncer de mama.

Os trabalhos citados anteriormente fornecem contribuições importantes para incre-

¹Disponível em: <<https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>>. Acessado em: 29/06/2021.

mento das taxas de precisão na detecção de câncer de mama utilizando CNNs em bases de imagens histológicas. Os autores preocuparam-se em reduzir a quantidade total de camadas das arquiteturas durante a fase de treinamento, remodelando suas estruturas a partir dos resultados obtidos no conjunto de validação. Treinar uma CNN do zero pode não ser uma alternativa viável em alguns cenários, seja pela escassez de recursos computacionais disponíveis ou pela falta de base de imagens rotuladas em quantidade suficiente para não incorrer em problemas de *overfitting* no treinamento. Outras abordagens baseadas em *deep learning* podem ser aplicadas com resultados satisfatórios, uma visão sobre essas abordagens que utilizam modelos híbridos é apresentada na seção a seguir.

3.2 CNN - Modelos híbridos

Neste trabalho, a terminologia modelos híbridos deve ser entendida como abordagens que submetem os vetores de atributos extraídos a partir de *deep features* de uma CNN a um processo de classificação de imagens. Em Niu e Suen (2012) os autores reproduziram um experimento semelhante a LeCun et al. (1989), entretanto, propondo um modelo híbrido entre uma CNN com cinco camadas e um classificador SVM para reconhecimento de dígitos escritos a mão. Nesse modelo, a CNN atua como um extrator de atributos treináveis e o SVM como classificador. Os experimentos foram realizados na base de dados de dígitos MNIST² com 60.000 amostras para treinamento e 10.000 para testes. Segundo os autores, as altas confiabilidades dos sistemas propostos foram alcançadas por meio de uma regra de rejeição, proposta por Cecotti e Belaïd (2005), onde uma amostra de teste é rejeitada quando a diferença entre os dois principais valores de probabilidade nas previsões classificadas é menor que um limite predefinido. As comparações com outros estudos na mesma base de dados indicam que essa fusão obteve melhores resultados: uma taxa de reconhecimento de 99,81% sem rejeição e, uma taxa de reconhecimento de 94,40% com rejeição de 5,60%. De acordo com os autores, o modelo híbrido proposto CNN-SVM supera cada classificador individualmente uma vez que o sistema híbrido compensa os limites dos classificadores CNN e SVM, incorporando méritos de ambos.

Elleuch, Maalej e Kherallah (2016) propuseram um modelo híbrido entre CNN-SVM com *droupout* para reconhecimento de caracteres árabes escritos a mão. No modelo proposto, o objetivo foi mesclar as capacidades de reconhecimento de uma CNN e com um classificador SVM. A CNN proposta é composta de cinco camadas

²Disponível em: <<https://github.com/zalandoresearch/fashion-mnist>>. Acessado em: 29/06/2021.

convolucionais e uma camada totalmente conectada, a camada final foi substituída por um SVM. *Dropout* foi utilizado na camada totalmente conectada a fim de reduzir o *overfitting* da rede. Os experimentos foram executados na base dados HACDB³, contendo 6.600 caracteres escritos a mão por 50 pessoas, sendo 5.280 imagens para treinamento e 1.320 imagens para teste. A base de dados IFN/ENIT⁴ também foi utilizada, contendo 26.459 palavras árabes escritas a mão por 411 pessoas, sendo que o conjunto foi dividido em quatro conjuntos para treinamento, reservando um total de 1.120 imagens para teste. Os autores compararam as taxas de erro da arquitetura proposta, do modelo 1, CNN-SVM (com *dropout*), com um modelo 2, CNN-SVM e um modelo 3, utilizando apenas CNN. Para a base dados HACDB, a taxa de erro para o modelo 1 foi 5,83%, para o modelo 2 a taxa de erro foi 6,59% enquanto que, para o modelo 3, a taxa de erro foi de 14,71%. Na base de dados IFN/ENIT a taxa de erro para o modelo 1 foi 7,05% e para o modelo 2 a taxa de erro foi de 7,32%. Não foi realizado o teste na base IFN/ENIT para o modelo 3.

No modelo PTL-CFS-CNN proposto por Yu e Dong (2018) para classificação de imagens de sensoriamento remoto, a seleção de atributos com base em correlação é adotada durante o pré-processamento das amostras, visando a redução dimensional dos atributos. Segundo os autores, essa abordagem aumenta a precisão e a eficiência da CNN, diminuindo o tempo de execução do algoritmo. A abordagem proposta utiliza o algoritmo PTL (*parameter transfer learning*) para obter a inicialização dos parâmetros de uma CNN, então, faz-se a seleção de atributos por meio do algoritmo CFS a fim de eliminar os atributos redundantes e ruídos do conjunto de dados original. Por fim, as imagens de sensoriamento são classificadas em uma CNN convencional. O método proposto de PTL-CFS + CNN foi comparado com CFS + CNN, PTL + CNN e CNN tradicional. Os autores afirmam que, comparado ao método utilizando uma CNN convencional e algoritmos de aprendizado de classificação superficial, o método proposto tem uma clara vantagem com uma acurácia média de 92,78%.

Nos trabalhos anteriores, é possível observar que a classificação de imagens por meio de *deep learning* pode ser aplicada com sucesso em imagens de diferentes contextos, também é possível verificar que a eficiência de classificação de vetores de atributos produzidos por uma CNN podem ser iguais ou superiores quando comparados a utilização de uma CNN com classificador interno. A proposta de Yu e Dong (2018) utiliza transferência de aprendizado para inicializar os parâmetros da rede, reduzindo assim seu tempo de treinamento quando comparada a uma rede que faz a inicialização de parâmetros de forma aleatória.

³Disponível em: <<http://dro.deakin.edu.au/view/DU:30092688>>. Acessado em: 29/06/2021.

⁴Disponível em: <<http://www.ifnenit.com/>>. Acessado em: 29/06/2021.

A transferência de aprendizado pode ser utilizada de forma ainda mais efetiva quando parâmetros de camadas inteiras são transportados entre arquiteturas, reduzindo ainda mais os recursos computacionais necessários ao treinamento. Normalmente, quando essa estratégia é utilizada, camadas iniciais de uma arquitetura são transferidas para outra e apenas as camadas remanescentes são treinadas no conjunto de imagens que se deseja classificar. Entretanto, quando uma rede é pré-treinada em uma base de imagens grande, com dezenas ou até centenas de classes, uma estratégia que também pode produzir bons resultados é a transferência de aprendizado profundo por meio da extração e classificação de *deep features*. Trabalhos recentes que utilizam essa abordagem são apresentados na próxima subseção.

3.2.1 Transferência de aprendizado profundo

O uso de redes convolucionais pré-treinadas para extração de atributos pode ser observado em vários estudos, entretanto, não existe atualmente um conjunto de dados de imagens médicas anotadas, por exemplo, em larga escala comparável ao ImageNet (DENG et al., 2009), pois a aquisição de dados é difícil e a anotação das imagens é cara. Ao contrário dos conjuntos de dados de imagens comumente utilizados em visão computacional, o ImageNet oferece atualmente um banco de dados muito abrangente com mais de 14 milhões de imagens naturais categorizadas de mais de 20.000 classes. Os modelos de CNN treinados nesse banco de dados atuam como a espinha dorsal para melhorar significativamente muitos problemas de detecção de objetos e segmentação de imagens utilizando outros conjuntos de dados (SHIN et al., 2016).

Santos e Ponti (2019) propuseram um modelo híbrido em que atributos específicos foram extraídos a partir de camadas de uma CNN, recombinaos e classificados. A arquitetura escolhida foi ResNet-50, pré-treinada na base de imagens ImageNet⁵. Nesta abordagem, as camadas iniciais da rede atuam como um extrator de atributos locais tais como, forma, borda e cor. A penúltima camada da rede é responsável por fornecer atributos globais, como textura e semântica. Então, a fusão entre os atributos locais e globais é submetida a um processo de redução de atributos, utilizando o algoritmo TCA (*transfer component analysis*). Finalmente, o conjunto de atributos é submetido a um classificador SVM. Quatro bases de imagens foram testadas: frutas (*Fruits-360*⁶), objetos (*Amazon and Webcam*⁷), lesões de pele (*HAM10000*⁸) e fotos

⁵Disponível em: <<http://image-net.org/download-images>>. Acessado em: 29/06/2021.

⁶Disponível em: <<https://github.com/Horea94/>>. Acessado em: 29/06/2021.

⁷Disponível em: <<https://github.com/jindongwang/transferlearning/blob/master/data/dataset.md>>. Acessado em: 29/06/2021.

⁸Disponível em: <https://github.com/ptschandi/HAM10000_dataset>. Acessado em: 29/06/2021.

(*Corel1000*⁹). Os autores afirmam que o objetivo da proposta não é superar métodos concorrentes descritos na literatura, mas sim, apresentar a importância da investigação adicional do poder discriminatório das camadas iniciais de uma CNN, que são frequentemente negligenciadas na concepção de novos métodos. Ainda assim, segundo os autores, o modelo proposto obteve bom desempenho, quando comparado à classificação utilizando apenas CNN, indicando que, metodologias que fazem uso de extração de atributos ao longo da arquitetura da rede, a fim de determinar as camadas mais discriminativas, podem fornecer modelos robustos e computacionalmente acessíveis.

O uso de redes convolucionais pré-treinadas para extração de atributos na área médica pode ser observado em outros estudos em imagens histológicas e radiográficas. Em Vogado et al. (2017) os autores analisaram imagens de amostras de esfregaço de sangue para detecção de leucemia por meio de *deep features* extraídas a partir de três diferentes arquiteturas CNN pré-treinadas, AlexNet, CaffeNet, VGG-F. Os resultados foram avaliados para cada rede separadamente, como também, a partir da construção de um vetor de atributos concatenando dados originários das três arquiteturas. A técnica utilizada para seleção de atributos foi PCA e três algoritmos foram utilizados na etapa de classificação, SVM, MLP e RF. A base testada, ALL-IDB¹⁰, é composta por dois conjuntos de dados, ALL-IDB1, que contém 108 imagens (59 imagens de amostras de tecido saudável e 49 imagens de tecido com câncer) e ALL-IDB2, que contém 260 imagens (sendo que, cada uma das imagens possui um linfoblasto ou um linfócito). De acordo com os autores, o modelo proposto aplicado a rede VGG-F obteve taxa de precisão de 100% além de requerer menos tempo de processamento que as demais combinações avaliadas nesse trabalho.

Santos e Ponti (2018) apresentaram um estudo detalhado sobre seleção de atributos extraído de três arquiteturas de CNN com diferentes quantidades de camadas, MobileNet, VGG-19 e ResNet-50, pré-treinadas na base ImageNet para detecção de câncer em imagens de lesões de pele. De acordo com os autores, os atributos foram extraídos de cada uma das sete últimas camadas de cada arquitetura incluindo camadas de convolução, *pooling*, totalmente conectadas e *softmax*. Para classificação, três classificadores foram eleitos, *Linear SVM*, RF e *AdaBoost*. A base de dados utilizada, PH2¹¹, possui 200 imagens dermoscópicas, sendo 40 imagens malignas (melanomas) e 160 imagens não malignas (80 *common nevi* e 80 *dysplastic nevi*). A técnica de seleção de atributos utilizada foi PCA. Na avaliação do modelo, o melhor espaço de atributos foi gerado por uma das camadas densas de MobileNet, com uma precisão

⁹Disponível em: <<https://sites.icmc.usp.br/moacir/data/>>. Acessado em: 29/06/2021.

¹⁰Disponível em: <<https://homes.di.unimi.it/scotti/all/>>. Acessado em: 29/06/2021.

¹¹Disponível em: <<https://www.fc.up.pt/addi/ph2%20database.html>>. Acessado em: 29/06/2021.

balanceada de 94%, indicando que, camadas mais profundas não necessariamente geram espaço de atributos melhores e que redes com mais camadas nem sempre geram atributos de melhor qualidade.

Em Novitasari et al. (2020) os autores utilizaram extração de *deep features* para categorizar pacientes com Covid-19 em imagens de raios X a partir da base Cohen et al. (2020) — disponível em *preprint*. A base utilizada contém 100 imagens separadas em três classes: normal, pneumonia e Covid-19. Metade das imagens correspondem a classe diagnóstico normal, as demais imagens correspondem a raios X de pacientes diagnosticados com síndrome de dificuldade respiratória aguda, Covid-19, síndrome respiratória aguda grave (SARS) e síndrome respiratória do Oriente Médio (MERS). *Deep features* foram extraídas a partir das arquiteturas GoogleNet, ResNet-18, ResNet-50 e ResNet-101. A seleção de atributos foi feita por meio de PCA e *Relief* e SVM foi utilizado na classificação. O modelo proposto atingiu acurácia de 97,33% utilizando a arquitetura ResNet na separação de três classes, e 100% na separação de duas classes, normal/Covid-19 e pneumonia/Covid-19, respectivamente.

Outro exemplo de modelo híbrido é descrito por Kumar e Sharma (2021). A proposta explorou os modelos CNN pré-treinados Xception e VGG16 com classificadores convencionais de aprendizado de máquina (*logistic regression*, SVM, e *decision tree*) para produzir as maiores taxas de acurácia envolvendo o câncer de mama, sem considerar o fator de magnificação das imagens histopatológicas H&E. A base de imagens utilizada foi BreakHis¹. As arquiteturas CNN foram utilizadas como extrator de atributos. Para reduzir *overfitting*, os autores utilizaram aumento artificial de dados. A arquitetura Xception produziu a melhor acurácia do modelo, 82,45%. Os autores sugerem que essa associação pode ser utilizada como um sistema automatizado para o diagnóstico precoce em imagens histopatológicas de câncer de mama.

3.3 Considerações

A partir dos trabalhos apresentados previamente, é possível verificar os benefícios do uso de modelos híbridos explorando a transferência de aprendizado, especialmente em contextos com número reduzido de amostras anotadas para treinamento. Enquanto bases compostas por frutas e objetos não proporcionam ganho expressivo em classificação (SANTOS; PONTI, 2019), os modelos híbridos aplicados em imagens histológicas forneceram resultados importantes (LITJENS et al., 2016; ARAÚJO

¹Disponível em: <<https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>>. Acessado em: 29/06/2021.

et al., 2017; VOGADO et al., 2017; DABEER; KHAN; ISLAM, 2019; KUMAR; SHARMA, 2021). Apesar de tais contribuições, destacamos que a investigação sobre o poder discriminativo de modelos híbridos, explorando *deep features* por transferência de aprendizado em diferentes camadas de CNN e comitê de classificadores ainda foi pouco explorado em imagens H&E representativas de câncer de mama, colorretal e tecido hepático, especialmente para definir as associações mais apropriadas para o diagnóstico dessas doenças. Uma visão geral dos trabalhos revisados nesse capítulo está na Tabela 3.1.

A extração de *deep features* resulta em um espaço de características com dezenas ou até centenas de milhares de atributos. Conforme observado por Li et al. (2017) e Yu e Dong (2018), a seleção de atributos baseados em ranqueamento, como o algoritmo *Relief*, é uma opção viável para essa abordagem por possuir complexidade computacional quadrática, controle simples do número de características mais apropriado para a etapa de classificação e pela eficácia em descartar atributos irrelevantes no espaço de características pesquisado (ZENG et al., 2015). Em Rehman et al. (2019) os autores utilizaram o algoritmo *Relief* para selecionar *deep features* extraídas de uma arquitetura AlexNet com o objetivo de verificar grau de parentesco a partir de um banco de imagens de rostos. Em Cohen et al. (2020), os autores extraíram *deep features* de imagens médicas a partir das arquiteturas GoogleNet e ResNet, a etapa de seleção foi feita por meio do algoritmo *Relief*, com o objetivo de classificar imagens de pneumonia e Covid-19. Os bons resultados obtidos pela aplicação de seleção de atributos por ranqueamento nesses trabalhos justificam a escolha dessa abordagem de seleção em um contexto de características similares.

Embora o trabalho de Dabeer, Khan e Islam (2019) tenha atingido acurácia de 99,86% na detecção de câncer de mama, assim como os demais autores aqui citados, cada trabalho explorou uma única base histológica. Não há como afirmar, pelas informações contidas nesses trabalhos, que os modelos aqui expostos produziram resultados similares quando testados em outras bases de tecido histológico. Por essa razão, a proposta de um modelo híbrido robusto, capaz de categorizar diferentes bases, por meio de extração de *deep features*, utilizando transferência de aprendizado e classificação via comitê, mostra-se com um desafio importante e ainda não disponível na literatura.

Tabela 3.1: Sumarização dos trabalhos relacionados, com informações sobre métodos de utilizados e desempenho obtido.

Trabalho	Imagem	Resultado	Abordagem
Niu e Suen (2012)	Escrita a mão	Acurácia	99,81% CNN 5 camadas + SVM
Litjens et al. (2016)	Histologia	AUC	0,99 CNN 8 camadas
Elleuch, Maalej e Kherallah (2016)	Escrita a mão	Tx. Erro	5,83% CNN + SVM dropout
		Tx. Erro	6,59% CNN + SVM
		Tx. Erro	14,71% CNN
Araújo et al. (2017)	Histologia	Acurácia	95,60% CNN 13 camadas + SVM
Vogado et al. (2017)	Histologia	Acurácia	100,00% AlexNet, CaffeNet, VGG-F + PCA
Yu e Dong (2018)	Sens. remoto	Acurácia	92,78% PTL-CFS + CNN 8 camadas
Santos e Ponti (2018)	Lesões de pele	Acurácia	94,00% MobileNet, VGG-19, ResNet-50 + PCA
Dabeer, Khan e Islam (2019)	Histologia	Acurácia	99,86% CNN 6 camadas
Santos e Ponti (2019)	Frutas	Acurácia	41,46% ResNet-50 + PCA
	Objetos	Acurácia	55,12% ResNet-50 + PCA
	Lesões de pele	Acurácia	87,50% ResNet-50 + PCA
	Fotos	Acurácia	96,17% ResNet-50 + PCA
	Radiologia	Acurácia	100,00% GoogleNet, ResNet + PCA, Relief
Novitasari et al. (2020)			
Kumar e Sharma (2021)	Histologia	Acurácia	82,45% Xception, VGG16 + SVM

Fonte: Elaborado pelo autor.

Capítulo 4

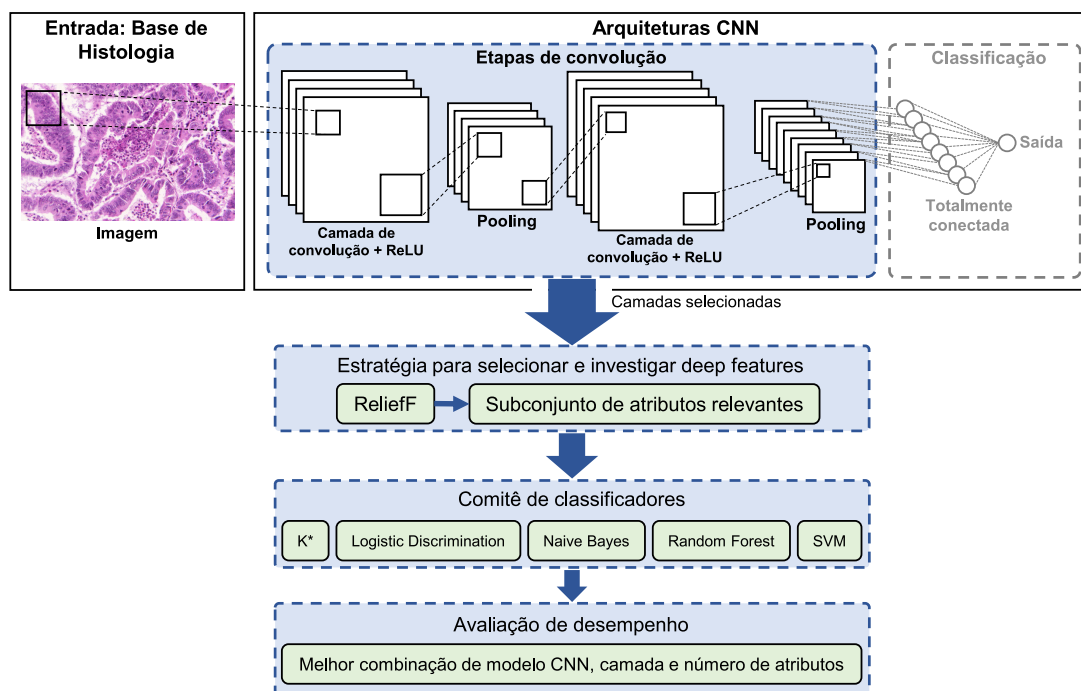
Materiais e métodos

Nosso modelo explora *deep features* obtidas de arquiteturas CNN, pré-treinadas no conjunto de dados ImageNet (DENG et al., 2009), e comitê de classificadores (*ensemble*) a fim de investigar diferentes categorias de imagens H&E. Para tanto, definimos conjuntos de camadas dos modelos CNN e coletamos as *deep features* fornecidas por cada camada, processo realizado durante a execução de cada rede. Em seguida, as *deep features* mais relevantes foram definidas a partir de uma análise sistemática por meio do uso combinado de uma estratégia de ranqueamento com avaliação via validação cruzada (*cross-validation*) *k-fold*. Por fim, propomos um comitê de classificadores de diferentes categorias a fim de identificar as melhores combinações. Os resultados obtidos foram associações (arquitetura, identificação da camada e número de atributos) para distinguir os diferentes padrões presentes em imagens histológicas H&E. Uma visão geral do modelo é ilustrada na Figura 4.1 e os detalhes são apresentados nas próximas seções.

4.1 Pacotes de software e ambiente para os experimentos

Neste trabalho, a abordagem para processamento de redes neurais e extração de *deep features* foram implementadas por meio das ferramentas *Deep Network Design* e *Transfer Learning Toolbox*, disponíveis no pacote MATLAB R2019a (MATLAB, 2019). Os nomes das camadas exploradas em nosso trabalho seguem as nomenclaturas definidas nesses pacotes. Os algoritmos empregados para selecionar e classificar os atributos são os disponíveis na plataforma Weka 3.6.15 (WITTEN; FRANK, 2002). Consideramos também descida do gradiente estocástico com otimizador de *momentum*

Figura 4.1: Visão geral ilustrativa da estratégia proposta para classificar imagens histológicas.



Fonte: Elaborado pelo autor.

usando os parâmetros padrão: taxa de *initial learning* de 0,0001; período de *learning rate drop* de 10; fator de *learning rate drop* de 0,1; valor de *L2 regularization* de 0,0001; tamanho do *mini-batch* de 4; e *split data* de 80% para treinamento e 20% para teste. Os experimentos foram realizados em uma CPU AMD Ryzen 5 3600X 6-Core a 3,79 GHz com 64 GB de RAM e uma NVIDIA GeForce RTX 1660 SUPER.

4.2 Bases de imagens

O modelo proposto foi aplicado em três diferentes bases públicas de imagens histológicas H&E: mama, colorretal e de tecido hepático. Para investigação do câncer de mama, utilizamos imagens fornecidas pelo *Center of Bio-image Informatics*, da University of California, Santa Barbara (UCSB) (GELASCA et al., 2008). A base consiste em 58 imagens de histológicas de tecido mamário (32 benignas e 26 malignas). A segunda base de imagens, câncer colorretal (CR), é fornecida por (SIRINUKUNWATTANA et al., 2017). Essa base consiste em 74 amostras benignas e 91 malignas, totalizando 165 imagens. A terceira base é nomeada como *liver gender* (LG), é disponibilizada pelo Atlas of Gene Expression in Mouse Ageing Project (AGEMAP) (AGEMAP, 2020). Essa base é constituída por imagens de tecido hepático de camun-

dongos separados como machos e fêmeas. Assim, essas duas classes representam o gênero da amostra coletada, totalizando 265 exemplos: macho com 150 imagens e fêmea com 115 amostras. Neste trabalho, as quantidades de imagens foram ajustadas a partir do menor número de exemplos disponíveis em cada grupo de cada base. Esse procedimento foi empregado para evitar que um grupo dominante pudesse afetar o resultado. Na Tabela 4.1 estão contidos os dados referentes às bases exploradas em nossos experimentos. Todas as bases investigadas possuem duas classes e são compostas exclusivamente por imagens histológicas coradas com H&E. Alguns exemplos dessas imagens estão na Figura 4.2.

Tabela 4.1: Detalhes sobre os conjuntos de imagens histológicas utilizados em nossa proposta.

Base	Imagem	Amostras	Resolução	Classes
UCSB	Tumores de mama	52	896 × 768	Benigno vs Maligno
CR	Tumores colorretais	148	Entre 567 × 430 e 775 × 522	Benigno vs Maligno
LG	Tecido hepático	230	417 × 312	Macho vs Fêmea

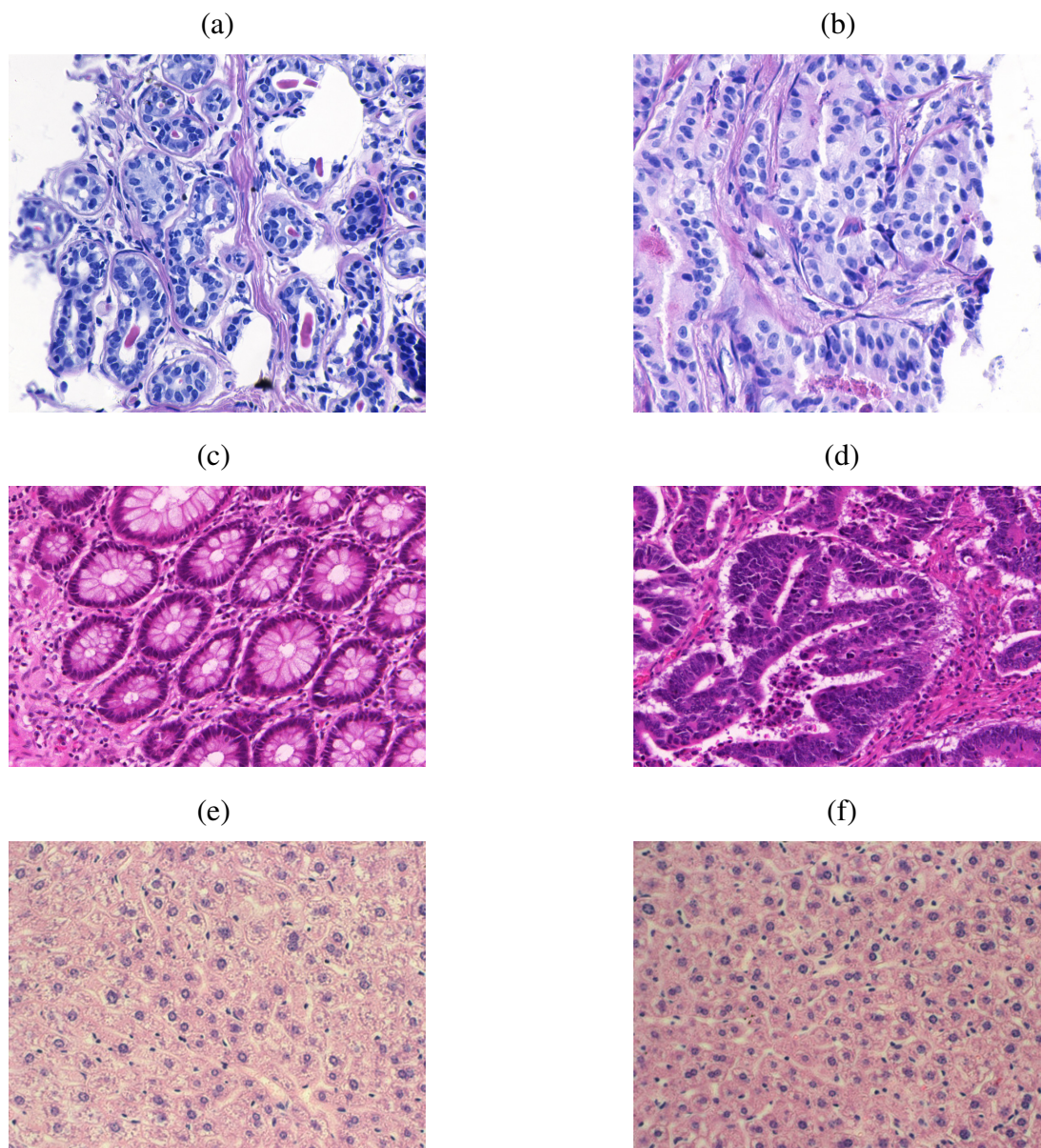
Fonte: Elaborado pelo autor.

4.3 Arquiteturas CNN e seleção de camadas: explorando transferência de aprendizado

Para analisar os contextos de imagens H&E, selecionamos as arquiteturas AlexNet e ResNet-50 pré-treinadas no conjunto de dados ImageNet (DENG et al., 2009) por fornecerem resultados relevantes no contexto de classificação de imagens histológicas (JIANG et al., 2019; MAHBOD et al., 2018; RAKHLIN et al., 2018; AHMAD; GHUFFAR; KHURSHID, 2019; TITORIYA; SACHDEVA, 2019). A exploração dessas arquiteturas por meio de transferência de aprendizado permitiu investigarmos as melhores associações para minimizar situações de *overfitting*. O modelo AlexNet consiste de cinco camadas convolucionais, três camadas de *pooling*, duas camadas totalmente conectadas e uma camada *softmax* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Essa arquitetura utilizou esquemas de regularização *dropout* e ReLU para reduzir *overfitting*, bem como LRN para minimizar o problema do gradiente de fuga (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Por outro lado, o modelo ResNet-50 consiste de quatro blocos, cada um com camadas de convolução e blocos residuais. O primeiro bloco possui nove camadas de convolução e três blocos residuais. O segundo bloco tem 12 camadas de convolução

Figura 4.2: Exemplos de imagens histológicas H&E: UCSB (GELASCA et al., 2008), benigno (4.2a) e maligno (4.2b); CR (SIRINUKUNWATTANA et al., 2017), benigno (4.2c) e maligno (4.2d); LG (AGEMAP, 2020), macho (4.2e) e fêmea (4.2f).



Fonte: Elaborado pelo autor.

e quatro blocos residuais. O terceiro bloco possui seis blocos residuais e 18 camadas de convolução. O quarto bloco possui a mesma quantidade de camadas de convolução e blocos que residuais que o bloco 1 (RAHHAL et al., 2018). Nessa arquitetura, as camadas receberam valores resultantes das função de ativação ReLU e os valores de entrada dessas funções. Assim, a arquitetura ResNet-50 utilizou *identity shortcut connection* contendo normalização de lotes para saltar grupos de camadas, permitindo mi-

nimizar questões envolvendo *overfitting* e o problema do gradiente de fuga (RAHHAL et al., 2018). Uma visão geral dos modelos CNN está na Tabela 4.2.

Tabela 4.2: Modelos CNN pré-treinados que foram considerados em nossa proposta.

Modelo	Camadas	Parâmetros	Input size
AlexNet	8	$6,2 \times 10^7$	$224 \times 224 \times 3$
ResNet-50	50	$2,6 \times 10^7$	$256 \times 256 \times 3$

Fonte: Elaborado pelo autor.

De acordo com os modelos CNN descritos previamente, *deep features* foram obtidas considerando a estratégia apresentada por Tan et al. (2018), porém, apenas das camadas escolhidas em nossa proposta e pré-treinadas no conjunto de dados ImageNet (DENG et al., 2009).

Da ResNet-50, escolhemos duas camadas iniciais e as últimas três camadas da rede. As camadas iniciais forneceram *deep features* responsáveis por quantificar forma, borda e cor da imagem. As camadas mais profundas foram utilizadas para identificar padrões globais, como textura e semântica (SANTOS; PONTI, 2019). Por exemplo, a camada *max_pooling2d_1* correspondeu ao *max pooling* (com tamanho de passo igual a 2×2) da primeira camada de convolução, que possui tamanho de núcleo de 7×7 e 64 diferentes filtros. A camada com mais atributos foi a *activation_4_relu*, correspondendo a função $\mathcal{F}(x) + x$ do primeiro bloco residual, que foi útil para avaliar a precisão de nosso modelo quando submetido à um conjunto de atributos densos. As camadas *activation_48_relu* e *activation_49_relu* pertencem ao segmento final da arquitetura ResNet-50, sendo parte do último bloco residual e a última camada de ativação ao logo da rede, respectivamente. Também, a partir do *average pooling* de tamanho 7×7 aplicado em *activation_49_relu*, escolhemos a última camada da arquitetura, *avg_pool*, que possui a menor quantidade de atributos desta arquitetura.

Em relação ao modelo de *deep features* via AlexNet, a investigação ocorreu sobre as cinco camadas convolucionais da rede, excluindo-se as camadas totalmente conectadas e *softmax*. É importante ressaltar que as quatro primeiras camadas de convolução foram selecionadas a partir da função de ativação ReLU, eliminando atributos com valores negativos. Além disso, a camada *pool5* correspondeu ao *max pooling* da última camada de convolução da AlexNet.

Os nomes das camadas e os valores correspondentes às *deep features* utilizadas em nossa proposta são mostrados na Tabela 4.3.

Tabela 4.3: Informações sobre as camadas das arquiteturas AlexNet e ResNet-50 investigadas em nossa proposta.

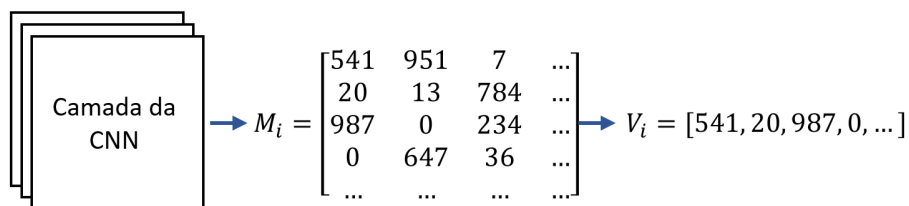
Arquitetura	Nome da Camada	Atributos
AlexNet	relu1	290.400
AlexNet	relu2	186.624
AlexNet	relu3	64.896
AlexNet	relu4	64.896
AlexNet	pool5	9.216
ResNet-50	max_pooling2d_1	193.600
ResNet-50	activation_4_relu	774.400
ResNet-50	activation_48_relu	25.088
ResNet-50	activation_49_relu	100.352
ResNet-50	avg_pool	2.048

Fonte: Elaborado pelo autor.

4.4 Estratégia para selecionar e investigar *deep features*

As camadas das arquiteturas CNN são representadas por matrizes n -dimensionais nomeadas $M_i[\dots]$, em que cada i define cada uma das cinco camadas de cada modelo CNN sob investigação. Em nosso trabalho, cada coluna de uma matriz M_i foi organizada sequencialmente em um vetor de atributos $V_i[\dots]$, de tal forma que o número de elementos em M_i é igual ao número de elementos em V_i . É importante destacar também que preservamos a ordem das *deep features* observadas em cada M_i , o que permite reconstruir cada matriz por meio dos valores de cada V_i e da dimensão correspondente a M_i . Uma ilustração dessa representação é exibida na Figura 4.3.

Figura 4.3: Ilustração da organização das *deep features* obtidas de uma camada em formato de vetor de atributos.



Fonte: Elaborado pelo autor.

Após a definição dos vetores V_i , utilizamos uma estratégia para selecionar as *deep features* mais relevantes. Nesse processo, cada conjunto V_i foi distribuído em subconjuntos S em função de m . Definimos m como o número de *deep features* em cada subconjunto, de acordo com a Equação 4.1. O número limitado a 100 *deep features* foi estabelecido baseado nos modelos descritos por Urbanowicz et al. (2018) e Ribeiro et al. (2019). Assim, cada subconjunto S_m é definido a partir dos m elementos melhor

ranqueados em cada V_i sob investigação. O ranqueamento foi definido por meio do algoritmo *ReliefF* (KONONENKO; ROBNIK-SIKONJA; POMPE, 1996). Nossa proposta consiste em um processo iterativo a fim de definir a camada e as *deep features* correspondentes mais relevantes.

$$m \in \{5 \leq m \leq 100, \frac{m}{5} \in \mathbb{N}\}. \quad (4.1)$$

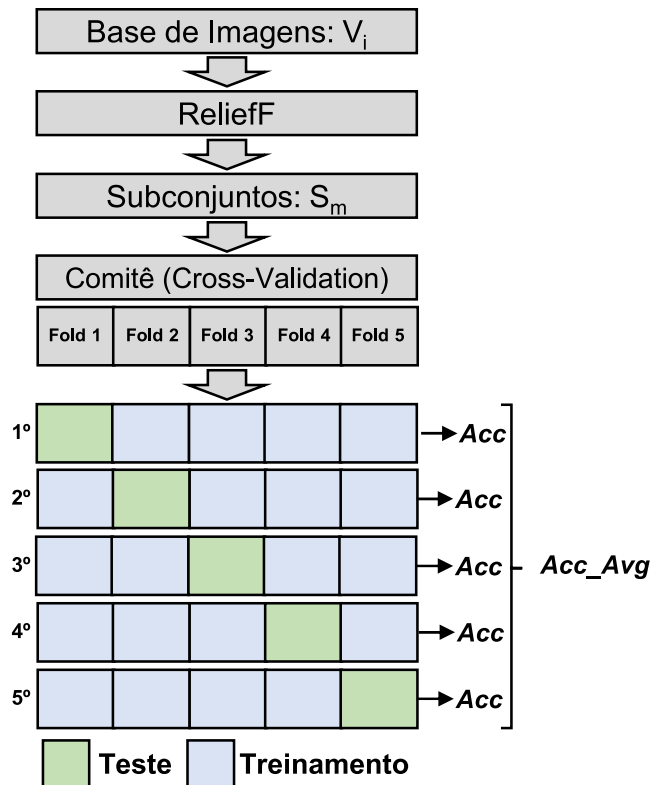
A análise de cada subconjunto S_m foi avaliada por meio da estratégia de validação cruzada (*cross-validation*) *k-fold*, a fim de verificar a capacidade de generalização de nossa proposta. Optamos pela utilização de $k = 5$ em todos os testes devido ao reduzido número de amostras disponíveis em cada base histológica. Por fim, aplicamos nosso método de comitê de classificadores (descrito na seção 4.5) para calcular a taxa de acurácia (Equação 4.2) em cada *k-fold*. A taxa de acurácia média em cada subconjunto S_m foi calculada por meio da Equação 4.3. Portanto, a melhor associação de V_i a um subconjunto correspondente S_m foi definida por meio da maior taxa de acurácia média (*Acc_Avg*) em cada base de imagens avaliada. Consequentemente, os resultados obtidos corresponderam às *deep features* mais relevantes, via transferência de aprendizado, para classificar e reconhecer padrões em imagens H&E. A Figura 4.4 ilustra as etapas de seleção de atributos e classificação utilizadas em nosso trabalho.

$$Acc_j = \frac{VP + VN}{VP + FP + VN + FN}, \quad (4.2)$$

em que: j refere-se ao número do *fold* correspondente à iteração do processo da validação cruzada; VP , taxa de verdadeiro positivo, define um resultado em que o modelo prevê corretamente o grupo positivo; VN , taxa de verdadeiro negativo, indica um resultado em que o modelo prevê corretamente o grupo negativo; FP , taxa de falso positivo, representa um resultado em que o modelo prevê incorretamente o grupo positivo; e, FN , taxa de falso negativo, define um resultado em que o modelo prevê incorretamente o grupo negativo.

$$Acc_Avg = \frac{1}{k} \sum_{j=1}^k Acc_j. \quad (4.3)$$

Figura 4.4: Ilustração do processo de seleção e classificação de atributos proposto em nosso modelo.



Fonte: Elaborado pelo autor.

4.5 Definição de um comitê de classificadores para a análise do poder discriminativo

A utilização de diferentes classificadores para reconhecimento de padrões é uma estratégia comumente aplicada na literatura, oferecendo análises de sucesso no contexto de imagens histológicas (SAGI; ROKACH, 2018; KASSANI; KASSANI, 2019), especialmente para conceder maior representatividade do problema sob investigação e minimizar a presença de *overfitting*. Para tanto, definimos um comitê de classificadores contendo cinco algoritmos de diferentes categorias. Os classificadores escolhidos foram K^* (CLEARY; TRIGG, 1995), *Logistic Discrimination* (CESSIE; HOUWELINGEN, 1992), *Naive Bayes* (JOHN; LANGLEY, 2013), *Random Forest* (BREIMAN, 2001) e SVM (ALPAYDIN, 2009). Nosso modelo combinou os resultados dos classificadores por meio da regra da soma, que pode ser resumida como a soma das probabilidades de previsão obtidas em cada classificador (KITTLER et al., 1998). Escolhemos essa regra considerando os bons resultados obtidos na classificação de ima-

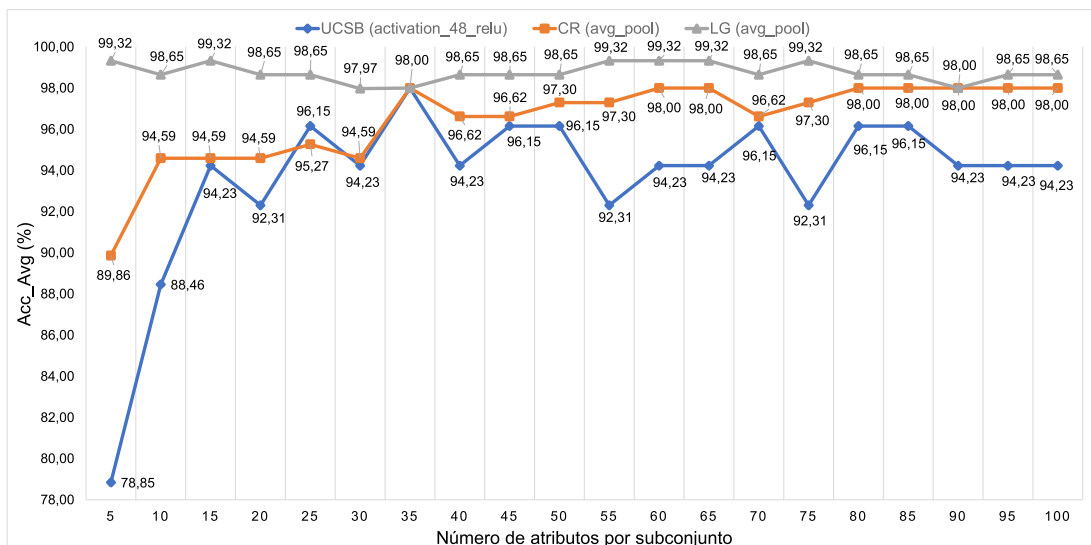
gens histológicas via estratégias de exploração de modelos CNN (NANNI; LUMINI; GHIDONI, 2018; ROBERTO et al., 2021). A decisão é fornecida pelo comitê e permite definir quais as associações mais relevantes para distinguir os grupos de imagens histológicas investigadas.

Capítulo 5

Resultados e discussão

Nosso modelo foi testado em três conjuntos de imagens histológicas, conforme descrição apresentada no Capítulo 4. As comparações foram: benigno versus maligno para as bases UCSB e CR; e como macho *versus* fêmea para a base LG. Nas Figuras 5.1 e 5.2 são apresentados os comportamentos das principais associações com os modelos CNN em função das taxas de acurácias e número de *deep features* em cada subconjunto avaliado.

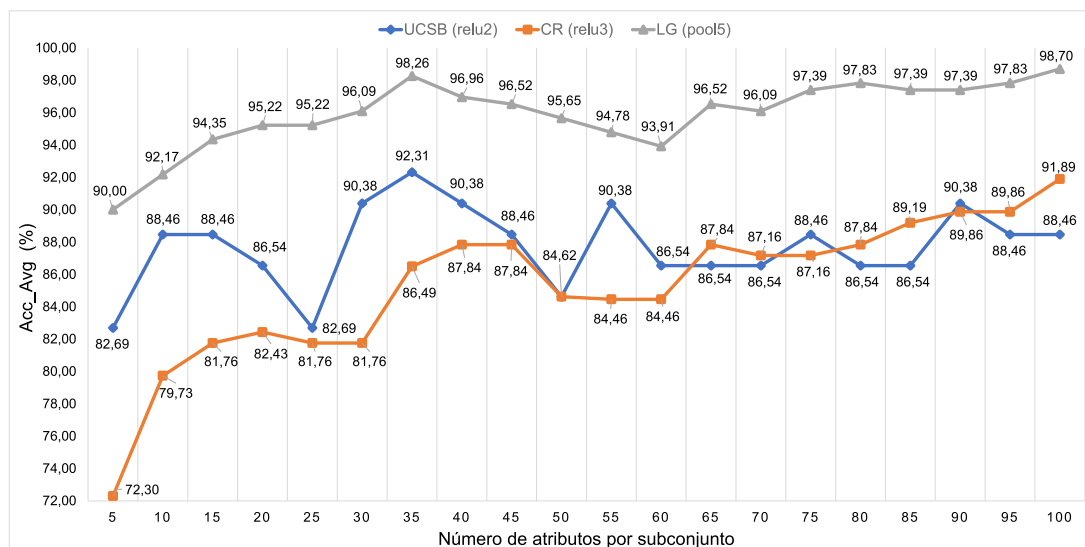
Figura 5.1: Comportamentos das principais associações a partir das *deep features* da rede ResNet-50: UCSB com a camada *activation_48_relu*; CR com a camada *avg_pool*; e LG com a camada *avg_pool*.



Fonte: Elaborado pelo autor.

Considerando os principais comportamentos mostrados nas Figuras 5.1 e 5.2, a estratégia proposta permitiu conhecer os modelos híbridos mais relevantes, explorando *deep features* por transferência de aprendizado, diferentes camadas de CNN e comitê

Figura 5.2: Comportamentos das principais associações a partir das *deep features* da rede AlexNet: UCSB com a camada *relu2*; CR com a camada *relu3*; e LG com a camada *pool5*.



Fonte: Elaborado pelo autor.

de classificadores. Por exemplo, o modelo explorando *deep features* da camada mais profunda da rede AlexNet (*pool5*) alcançou 98,70% de acurácia na base LG. Comportamento semelhante ocorreu com o modelo explorando a camada mais profunda da rede ResNet-50 (*avg_pool*), porém, a taxa de acurácia foi de 99,32%. Esta última combinação também foi a responsável por fornecer os atributos mais relevantes para o conjunto CR. O resultado foi uma taxa de acurácia de 98,00%. Esses comportamentos estão alinhados com as investigações disponíveis na literatura, de que camadas mais profundas tendem a produzir atributos de mais alto nível (MAJTNER; YILDIRIM-YAYILGAN; HARDEBERG, 2016; ESTEVA et al., 2017; SANTOS; PONTI, 2018; SANTOS; RIBEIRO; PONTI, 2019; SHI et al., 2019). Entretanto, uma contribuição importante do nosso estudo está em mostrar que uma combinação usando a camada *activation_48_relu* da ResNet-50 produziu acurácia máxima para classificar os grupos que definem a base UCSB. Mais ainda, modelos explorando as camadas *relu2* e *relu3*, pertencentes ao segmento intermediário da arquitetura AlexNet, também foram os responsáveis por fornecer as acurácias mais expressivas nas bases UCSB e CR, valores superiores a 91% de acurácia.

A partir das condições discutidas previamente, na Tabela 5.1 é apresentado um resumo das principais associações, com as indicações, em **negrito**, das camadas que produziram as *deep features* mais relevantes, os totais de atributos utilizados e as taxas de acurácias em cada base histológica. É possível constatar que os principais modelos

híbridos foram definidos com *deep features* da rede ResNet-50. A partir destes resultados, destacamos o número reduzido de características para obter as maiores taxas de acurácias em diferentes bases. Por exemplo, os percentuais de reduções nos totais de atributos foram de 99,86%, com o modelo híbrido definido a partir da camada *activation_48_relu* (UCSB), e 98,29% para o modelo usando os atributos da camada *avg_pool*, base CR. Esta última combinação também proporcionou uma redução de 99,76% no total de atributos para classificar as imagens da base LG. As principais associações utilizaram no máximo 35 *deep features*. Os desempenhos conquistados com um número reduzido de atributos definem mais uma contribuição da nossa proposta, pois estas condições permitem o uso de modelos baseados em estratégias de redes convolucionais em bases com um número reduzido de amostras e sem a presença de *overfitting*. É importante destacar que esta situação ocorreu pelas condições descritas previamente e também pela presença de um robusto comitê de classificadores, composto por cinco classificadores de diferentes categorias. Mais ainda, a metodologia proposta tornou viável a identificação das principais camadas e os atributos correspondentes para quantificar os padrões globais e locais, tais como forma, borda, semântica e cor, de diferentes conjuntos de imagens histológicas (SANTOS; PONTI, 2019).

Tabela 5.1: Combinações mais relevantes obtidas por nosso modelo para classificar as diferentes bases de imagens histológicas, com informações sobre a rede, nome da camada e o número de atributos que produziu maior taxa de acurácia.

Base	Rede	Camada	Atributos	Acc_Avg (%)
UCSB	AlexNet	relu2	35	92,31
	ResNet-50	activation_48_relu	35	98,00
CR	AlexNet	relu3	100	91,89
	ResNet-50	avg_pool	35	98,00
LG	AlexNet	pool5	100	98,70
	ResNet-50	avg_pool	5	99,32

Fonte: Elaborado pelo autor.

A partir dos experimentos realizados previamente, complementamos nossos resultados com os valores das métricas área sob a curva ROC (AUC) e taxa de falsos positivos (*FP Rate*), todas obtidas a partir das melhores associações em cada uma das bases de imagens histológicas. De acordo com Saito e Rehmsmeier (2015), a AUC é uma métrica útil para avaliar o desempenho de classificadores binários. No entanto, requer cuidado especial quando usada para aferir conjuntos de dados com classes desbalanceadas, situação diferente da explorada em nosso trabalho. Em complemento, a métrica *FP Rate* não é afetada pela distribuição de amostras nas classes, sendo essa também

uma métrica importante a ser observada. Os resultados detalhados são mostrados na Tabela 5.2.

Tabela 5.2: Valores de áreas sob as curvas ROC (AUC) e de taxa de falsos positivos (*FP Rate*) das melhores associações combinações.

Base	AUC	FP Rate
UCSB	0,981	0,019
CR	0,998	0,020
LG	0,998	0,024

Fonte: Elaborado pelo autor.

A partir desses resultados, observamos que os valores de AUC foram de 0,981 a 0,998, enquanto que a métrica *FP Rate* indicou 0,019 como a menor taxa e 0,024 como a maior. Com isso, pode-se concluir que os altos valores de acurácias e AUC, associados aos valores baixos de *FP Rate*, afastam a possibilidade de enviesamento dos resultados.

5.1 Comparações com técnicas de classificação e reconhecimento de padrões

Neste trabalho, aplicamos também técnicas reconhecidas de aprendizado de máquina a fim de observar os desempenhos das principais combinações. Os testes foram baseados em classificações fornecidas diretamente pelas redes AlexNet e ResNet-50, bem como via técnicas de classificação regularizadas, *Lasso* (*least absolute shrinkage and selection operator*) e *Ridge Regression* (NG, 2004; KOLTER; NG, 2009; SCHÖLKOPF et al., 2002). *Lasso* e *Ridge* são técnicas eficientes para reduzir a complexidade e o *overfitting* do modelo final de predição em conjuntos de atributos com alta dimensionalidade (FRIEDMAN; HASTIE; TIBSHIRANI, 2010), situação explorada em nosso trabalho. Os experimentos utilizando técnicas regularizadas foram executados no pacote Scikit-Learn 0.18.1 (PEDREGOSA et al., 2011).

Para fornecer um panorama dos resultados produzidos por nosso modelo com os modelos CNN originais, consideramos em nossos testes as mesmas imagens disponíveis nas bases CR, UCSB e LG. É importante destacar que os testes com classificador interno das CNNs também foram realizados com uma abordagem de aumento artificial de dados a fim de incrementar o número de amostras disponíveis e introduzir variabilidade em cada conjunto de imagens. Os dados aumentados contribuem para um conjunto mais abrangente de possíveis pontos de dados, minimizando assim o *overfitting*

entre o conjunto de treinamento e validação (SHORTEN; KHOSHGOFTAAR, 2019). Portanto, as estratégias que utilizamos foram: reflexão aleatória de 50% nos eixos x e y ; rotação aleatória em até 1 grau; e translação horizontal e vertical aleatórias em até 1 pixel. Esses valores foram empregados para minimizar a possível degradação das taxas de classificação devido ao fundo da imagem, conforme descrito em MathWorks (2021). O pacote *Transfer Learning Toolbox* foi aplicado para o aumento artificial de dados, disponível no MATLAB R2019a (MATLAB, 2019). Essas condições permitiram duplicar o total de amostras disponíveis para as etapas de treinamento e validação. Nesse tipo de experimento, foram consideradas as taxas de acurácia fornecidas diretamente pelos modelos CNN, obtidas por meio dos classificadores internos de cada modelo. As classificações foram repetidas três vezes para calcular as médias e desvios padrão em cada conjunto de imagens. Os resultados fornecidos pelos modelos CNN sem aumento de dados são apresentados na Tabela 5.3. Esses valores foram obtidos a partir da primeira época de cada rede para evitar *overfitting*. Da mesma forma, os desempenhos após a aplicação da estratégia de aumento de dados são mostrados nas Tabelas 5.4 e 5.5 para as redes AlexNet e ResNet-50, respectivamente. Os valores de acurácia foram definidos com o número de épocas de treinamento variando de 1 a 30, coletando os resultados em intervalos regulares a cada 10 épocas. Destacamos em negrito as taxas mais significativas.

Além disso, as técnicas regularizadas foram aplicadas considerando os mesmos conjuntos de *deep features* explorados em nossos experimentos: 186.624 (*relu2*); 64.896 (*relu3*); 9.216 (*pool5*); 25.088 (*activation_48_relu*); e 2.048 (*avg_pool*). Em relação às combinações de técnicas de classificação regularizadas, nossos testes foram realizados combinando os classificadores SVM e LD com *Lasso* e *Ridge* (NG, 2004; CESSIE; HOUWELINGEN, 1992; FAN et al., 2008). As taxas de acurácias por meio dessas combinações estão na Tabela 5.6 para as bases CR, UCSB e LG. As associações mais relevantes foram destacadas em negrito.

Tabela 5.3: Taxas de acurácias (%) fornecidas pelas arquiteturas AlexNet e ResNet-50 explorando os conjuntos UCSB, CR e LG sem aumento de dados.

Base	AlexNet	ResNet-50
UCSB	53,33 ± 0,05	60,00 ± 0,14
CR	57,78 ± 0,06	78,89 ± 0,03
LG	86,96 ± 0,05	88,41 ± 0,05

Fonte: Elaborado pelo autor.

Quando comparamos os desempenhos com os valores obtidos por outras técnicas (classificações por redes e técnicas regularizadas), nossa proposta apresentou soluções

Tabela 5.4: Taxas de acurácias (%) fornecidas pelo modelo AlexNet explorando as bases UCSB, CR e LG com aumento de dados.

Base	Época 1	Época 10	Época 20	Época 30
UCSB	60,00 ± 0,08	73,33 ± 0,12	76,67 ± 0,12	83,33 ± 0,12
CR	68,89 ± 0,15	80,78 ± 0,08	88,89 ± 0,02	95,55 ± 0,03
LG	73,92 ± 0,09	98,55 ± 0,01	95,53 ± 0,06	98,55 ± 0,01

Fonte: Elaborado pelo autor.

Tabela 5.5: Taxas de acurácias (%) fornecidas pelas arquiteturas ResNet-50 explorando as bases UCSB, CR e LG com aumento de dados.

Base	Época 1	Época 10	Época 20	Época 30
UCSB	53,33 ± 0,12	80,00 ± 0,01	86,67 ± 0,05	83,33 ± 0,12
CR	83,33 ± 0,08	95,56 ± 0,02	97,78 ± 0,02	98,89 ± 0,02
LG	89,85 ± 0,09	99,28 ± 0,01	97,83 ± 0,02	97,10 ± 0,01

Fonte: Elaborado pelo autor.

Tabela 5.6: Taxas de acurácias (%) obtidas com os classificadores regularizados e explorando os mesmos conjuntos de *deep features* para avaliar o método proposto.

Classificador	Base	Camada e Rede	Acurácia
SVM com Lasso	UCSB	relu2, AlexNet	86,50
	CR	avg_pool, ResNet-50	97,29
	LG	avg_pool, ResNet-50	98,64
SVM com Ridge	UCSB	activation_48_relu, ResNet-50	92,30
	CR	relu3, AlexNet	87,83
	LG	avg_pool, ResNet-50	97,82
LD com Lasso	UCSB	activation_48_relu, ResNet-50	75,00
	CR	avg_pool, ResNet-50	97,97
	LG	avg_pool, ResNet-50	90,31
LD com Ridge	UCSB	activation_48_relu, ResNet-50	94,23
	CR	avg_pool, ResNet-50	96,62
	LG	avg_pool, ResNet-50	97,39

Fonte: Elaborado pelo autor.

capazes de oferecer resultados competitivos em todas as condições aqui exploradas. É notável que as diferenças mais significativas (em todos as bases) foram em relação às taxas de acurácias fornecidas pelas redes aplicadas diretamente nas imagens sem aumento de dados. Nesses experimentos, nossa proposta forneceu os melhores resultados. Por exemplo, a classificação considerando a base UCSB indicou a diferença mais significativa: taxa de acurácia de 60,00% com ResNet-50 versus 98,00% com o modelo proposto. Em relação às classificações que exploram conjuntos de dados com

aumento de dados, nossos modelos forneceram as taxas mais altas nos conjuntos de dados UCSB e LG: a maior diferença (11,67%) também pode ser observada no conjunto de dados UCSB. Em relação à base LG, os resultados foram 99,32% (modelo proposto) contra 99,28 % (ResNet-50), uma diferença de 0,04%. Entretanto, nossa melhor combinação forneceu uma taxa de acurácia ligeiramente inferior (0,89% de diferença) quando a base CR é considerada. Nesses casos, nossa proposta tem a vantagem de apresentar soluções aceitáveis com um número reduzido de atributos e sem requerer aumento de dados para a etapa de classificação.

Para as técnicas apresentadas na Tabela 5.6, a melhor acurácia foi de 94,23% via LD com *Ridge* para a base UCSB. A melhor combinação fornecida por nosso método resultou em uma acurácia de 98,00%. Este taxa é 3,77% maior que os valores obtidos por meio das classificações baseadas em técnicas regularizadas. Em relação à base CR, o melhor resultado entre as técnicas comparadas foi via LD com *Lasso* (acurácia de 97,97%). Este valor é sutilmente inferior ao alcançado pela nossa proposta (98,00%). Quando a base LG é considerada, nossa melhor associação resultou em uma acurácia de 99,32%, superando a melhor taxa entre as técnicas comparadas: uma taxa de acurácia de 98,64% via SVM com *Lasso*.

Por fim, aplicamos o teste de Friedman para avaliar as classificações fornecidas pelas melhores associações, considerando uma visão geral de todos os conjuntos de dados (Tabelas 5.4, 5.5 e 5.6). Este teste é uma abordagem estatística não-paramétrica capaz de ranquear k associações de forma que a melhor solução adquira a 1^o posição (DEMŠAR, 2006). O teste de Friedman permite analisar a variância de medidas repetidas assumindo que os valores são pareados entre as amostras e independentes uns dos outros dentro das amostras. A análise depende da classificação dos valores emparelhados em cada conjunto de dados, do menor para o maior, com cada conjunto de valores sendo classificado separadamente. As classificações em cada amostra são então somadas, compondo o *Friedman's Score* (FRIEDMAN, 1937). Assim, o ranking médio é mostrado na Tabela 5.7 considerando as taxas de acurácias.

Pode-se observar a partir desses resultados que nossa proposta foi capaz de fornecer soluções que adquirem a primeira posição, mesmo considerando os desempenhos alcançados por técnicas importantes. Isso indica o potencial de nossa abordagem nas diferentes condições em que foi testada. Logo, acreditamos que os modelos híbridos foram capazes de otimizar o processo de transferência de aprendizagem, considerando as associações mais relevantes para a classificação de imagens histológicas de H&E comumente investigadas pela comunidade científica.

Outras comparações poderiam ser realizadas para verificar se esses resultados se

Tabela 5.7: Classificação média considerando as melhores associações para conjuntos de dados UCSB, CR e LG.

Classificação	Solução	Friedman's Score
1°	Solução proposta (modelos híbridos)	1,5
2°	Rede ResNet-50 com aumento de dados	2,17
3°	SVM com Lasso	4,0
4°	LD com Ridge	4,33
5°	SVM com Ridge	5,0
6°	Rede AlexNet com aumento de dados	5,33
7°	LD com Lasso	5,67

Fonte: Elaborado pelo autor.

mantêm em mais condições e configurações, ou mesmo para fazer ajustes e definir os limites de cada modelo. No entanto, acreditamos que os experimentos apresentados foram capazes de fornecer uma visão ilustrativa relevante sobre a nossa proposta quando comparados às abordagens consolidadas comumente exploradas na literatura para os processos de classificação e reconhecimento de padrões. Além disso, considerando que nossa proposta foi capaz de definir as principais associações, em detalhes quanto às camadas e quantidade de atributos, as informações aqui apresentadas podem contribuir para o desenvolvimento de modelos computacionais direcionados a auxiliar especialistas interessados na composição de técnicas de aprendizado de máquina para reconhecimento de padrões.

5.2 Comparação de desempenho com a literatura: uma visão geral ilustrativa

Diferentes técnicas têm sido exploradas para investigar padrões em imagens histológicas, tais como para as bases UCSB, CR e LG (ROBERTO et al., 2021; BOUZIANE et al., 2020; BIANCONI; KATHER; REYES-ALDASORO, 2020; AWAN et al., 2020, 2020; ROBERTO et al., 2019; NANNI et al., 2019). No entanto, nenhum deles explorou o modelo de associações apresentado em nosso trabalho. Assim, uma visão geral ilustrativa é importante para mostrar a qualidade do nosso estudo, indicada nas Tabelas 5.8, 5.9 e 5.10 para as bases UCSB, CR e LG, respectivamente.

Ao levar em consideração esta visão ilustrativa, é possível verificar que nossa proposta foi capaz de fornecer modelos com resultados dentre os relatados na literatura, mesmo sem explorar combinações com atributos *handcrafted* ou CNN *ensembles*. Nossa proposta indicou modelos híbridos ainda não disponíveis na literatura para classificar padrões em imagens histológicas H&E, explorando *deep features* por transfe-

Tabela 5.8: Taxas de acurácia (%) fornecidas por diferentes abordagens para classificação de imagens histológicas da mama (UCSB).

Método	Abordagem	Acurácia (%)
Método proposto	ResNet-50 (activation_48_relu), Relief e comitê (DL) com 35 deep features	98,00
Li et al. (2019)	RefineNet e Atrous DenseNet (DL)	97,63
Yu et al. (2019)	CNN, LBP, SURF, GLCM e outros atributos handcrafted (DL+HC)	96,67
Feng, Zhang e Yi (2018)	Stacked denoising autoencoder (DL)	94,41
Kausar et al. (2019)	Normalização de cor, Haar wavelet decomposition e CNN 16 camadas (DL)	91,00
Roberto et al. (2021)	ResNet-50, FD, LAC e PERC (DL+HC)	89,66
Bouziane et al. (2020)	Atributos morfológicos, de intensidade e textura (HC)	86,20
Papastergiou, Zacharaki e Megalooikonomou (2018)	Decomposição Espacial, tensoress (DL)	84,67
Araújo et al. (2017)	Normalização de cor, CNN 13 camadas e SVM (DL)	83,30
Saxena, Shukla e Gyanchandani (2020)	ResNet-50 e kernelized weighted (DL)	60,30

Fonte: Elaborado pelo autor.

Tabela 5.9: Taxas de acurácia (%) definidas por diferentes abordagens para classificação de imagens histológicas colorretais (CR).

Método	Abordagem	Acurácia (%)
Roberto et al. (2021)	ResNet-50, FD, LAC e PERC (DL+HC)	99,39
Método proposto	ResNet-50 (avg_pool), Relief e comitê (DL) com 35 deep features	98,00
Dabass, Vig e Vashisth (2018)	CNN 31 camadas (DL)	96,97
Tavolara et al. (2019)	GAN e U-Net (DL)	94,02
Sena et al. (2019)	CNN 12 camadas (DL)	93,28
Santos et al. (2018)	Sample entropy e fuzzy (HC)	91,39
Roberto et al. (2019)	PERC (HC)	90,90
BenTaieb e Hamarneh (2017)	U-Net e AlexNet (DL)	87,50
Awan et al. (2020)	Normalização de cor, U-Net e GoogLeNet (DL)	85,00
Bianconi, Kather e Reyes-Aldasoro (2020)	ResNet, VGG, atributos espaciais e espectrais (DL+HC)	94,37

Fonte: Elaborado pelo autor.

rência de aprendizado, diferentes camadas de CNN e um comitê de classificadores. Por exemplo, na literatura, algumas combinações foram com o uso de único classificador (ARAÚJO et al., 2017; BARDOU; ZHANG; AHMAD, 2018; GOLATKAR; ANAND; SETHI, 2018). De fato, essa técnica pode resultar em métricas de classificação mais altas, entretanto, é preciso avaliar situações em que o classificador ajustar-se totalmente aos dados de treinamento, principalmente em bases com poucas amostras.

Tabela 5.10: Valores de acurácia (%) obtidos em diferentes abordagens para classificação de gênero a partir de imagens de tecidos hepáticos (LG).

Método	Abordagem	Acurácia (%)
Nanni et al. (2019)	6 modelos CNN e descritores HC (DL+HC)	100,00
Roberto et al. (2021)	ResNet-50, FD, LAC e PERC (DL+HC)	99,62
Método proposto	Relief e comitê (DL) com 5 deep features	99,32
Andrearczyk e Whelan (2017)	Texture CNN (DL)	99,10
Watanabe, Kobayashi e Wada (2016)	Descritores GIST, PCA e LDA (HC)	93,70

Fonte: Elaborado pelo autor.

Os modelos obtidos em nossa proposta evitam essa situação por meio de uma análise sistemática de *deep features* com um comitê robusto de classificadores, mesmo que isso implique em algumas reduções nas taxas de classificações. Mais ainda, é importante destacar que a maioria dessas propostas leva a um modelo quase ideal, pois cada uma das estratégias mencionadas utiliza diferentes tipos de atributos e combinações que são capazes de quantificar várias propriedades presentes em imagens histológicas. Portanto, consideramos esses métodos mais complementares do que comparáveis, uma vez que as valiosas informações contidas em cada uma dessas referências podem contribuir significativamente com a comunidade interessada no desenvolvimento de modelos para classificações de padrões em imagens H&E.

Capítulo 6

Conclusão

Neste trabalho, apresentamos uma estratégia capaz de fornecer modelos híbridos para classificar imagens histológicas tingidas com H&E de tecidos mamários, colorretais e hepáticos. Nossa proposta foi testada para o diagnóstico de grupos benigno versus maligno (bases UCSB e CR) e reconhecimento de padrões em imagens de tecido hepático de camundongos separados nas classes de macho e fêmea (base LG). Os experimentos foram realizados por meio de diferentes combinações de subconjuntos contendo os atributos mais relevantes e avaliados via validação cruzada *k-fold* para fornecer soluções viáveis com um reduzido número de atributos. Os modelos híbridos foram as melhores associações de *deep features* via transferência de aprendizado com um comitê de classificadores.

Os resultados indicaram os melhores modelos híbridos como aqueles produzidos pelas *deep features* obtidas da rede ResNet-50. Por exemplo, o modelo híbrido com a camada mais profunda da rede AlexNet (*pool5*) obteve a maior taxa de classificação (98,70%) na base LG. Por sua vez, na mesma base, o modelo híbrido com a camada mais profunda da rede ResNet-50 (*avg_pool*) atingiu 99,32% de acurácia. Essa associação também forneceu os atributos mais relevantes para a base CR, com um valor de acurácia de 98,00%. Os modelos que exploram as camadas mais profundas de uma CNN são os mais comumente utilizados em abordagens disponíveis na literatura. No entanto, uma contribuição importante da nossa proposta foi mostrar que *deep features* da camada *activation_48_relu* forneceram um modelo com a melhor taxa de classificação na base UCSB. Além disso, nossa proposta indicou que associações com *deep features* de camadas intermediárias da arquitetura AlexNet também proporcionaram taxas expressivas nas bases UCSB e CR. Assim, é possível destacar que nossa proposta foi capaz de otimizar o processo de transferência de aprendizado e definir os modelos híbridos mais relevantes para classificar e reconhecer padrões em imagens

H&E. Por exemplo, o número total de atributos foi reduzido em 99,86% com o modelo híbrido definido a partir da camada de *activation_48_relu* (base UCSB). Além disso, as principais associações utilizaram no máximo 35 *deep features*. Acreditamos que essas condições e informações sejam contribuições importantes, pois possibilitam a utilização de modelos baseados em CNNs em conjuntos de dados com um número reduzido de imagens, sem *overfitting* ou mesmo o emprego de técnicas de aumento de dados para garantir a viabilidade do processo de classificação. A análise sistemática de camadas e *deep features* com a proposta de um robusto comitê de classificadores permitiu contornar estas restrições. Acreditamos que estes resultados e associações são relevantes para especialistas interessados na composição de técnicas de reconhecimento de padrões em imagens.

É importante enfatizar que os principais modelos híbridos também foram capazes de apresentar resultados dentre os relatados na literatura, bem como aos fornecidos por estratégias consolidadas. Mesmo que o método não tenha superado alguns resultados fornecidos por outras propostas, nossa abordagem foi capaz de indicar detalhes das camadas mais comumente selecionadas e suas características correspondentes, utilizando um número reduzido de atributos.

Um resumo das principais contribuições obtidas foi:

1. A definição de modelos que apresentam as melhores associações de camadas da CNN, *deep features* correspondentes via transferência de aprendizado e um robusto comitê de classificadores;
2. Indicação de modelos híbridos capazes de indicar soluções aceitáveis com um número reduzido de atributos e sem exigir o uso de aumento de dados para o processo de classificação;
3. A realização de uma análise sistemática de *deep features* capazes de prover soluções sem *overfitting*;
4. Indicações e informações de que *deep features* de camadas intermediárias também são capazes de fornecer taxas de acurácia significativas em diferentes conjuntos de imagens histológicas;
5. A apresentação de uma proposta capaz de fornecer modelos híbridos para classificar imagens histológicas coradas com H&E de tecido mamário, colorretal e hepático.

Dessa forma, nossos objetivos enumerados no Capítulo 1, foram atingidos. Em trabalhos futuros, pretendemos:

1. Comparar esta proposta com diferentes redes neurais residuais utilizando modelos GAN para aumento de dados;
2. Utilizar as melhores associações para o reconhecimento de padrão em imagens H&E de linfomas não-Hodgkin e detecção de Covid-19 em imagens radiográficas torácicas;
3. Definir modelos híbridos com base em representações obtidas via técnicas *gradient-weighted class activation mapping* e *locally-interpretable model-agnostic explanation*;
4. Propor modelos híbridos com o uso de atributos fractais multidimensionais.

Referências

ABRAHAM, B.; NAIR, M. S. Computer-aided detection of covid-19 from x-ray images using multi-cnn and bayesnet classifier. *Biocybernetics and biomedical engineering*, Elsevier, v. 40, n. 4, p. 1436–1445, 2020.

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.

AHMAD, H. M.; GHUFFAR, S.; KHURSHID, K. Classification of breast cancer histology images using transfer learning. In: IEEE. *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. [S.l.], 2019. p. 328–332.

ALOM, M. Z.; TAHA, T. M.; YAKOPCIC, C.; WESTBERG, S.; SIDIKE, P.; NASRIN, M. S.; ESESN, B. C. V.; AWWAL, A. A. S.; ASARI, V. K. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018.

ALPAYDIN, E. *Introduction to machine learning*. [S.l.]: MIT press, 2009.

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.

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, v. 12, n. 6, p. e0177544, 2017.

AREL, I.; ROSE, D. C.; KARNOWSKI, T. P. et al. Deep machine learning—a new frontier in artificial intelligence research. *IEEE computational intelligence magazine*, The University of Tennessee, v. 5, n. 4, p. 13–18, 2010.

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.

BA, J.; CARUANA, R. Do deep nets really need to be deep? In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2654–2662.

- BARDOU, D.; ZHANG, K.; AHMAD, S. M. Classification of breast cancer based on histology images using convolutional neural networks. *Ieee Access*, IEEE, v. 6, p. 24680–24693, 2018.
- BARTL, E.; REZANKOVA, H.; SOBISEK, L. Comparison of classical dimensionality reduction methods with novel approach based on formal concept analysis. In: SPRINGER. *International Conference on Rough Sets and Knowledge Technology*. [S.l.], 2011. p. 26–35.
- BENTAIEB, A.; HAMARNEH, G. Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 3, p. 792–802, 2017.
- BIANCONI, F.; KATHER, J. N.; REYES-ALDASORO, C. C. Experimental assessment of color deconvolution and color normalization for automated classification of histology images stained with hematoxylin and eosin. *Cancers*, Multidisciplinary Digital Publishing Institute, v. 12, n. 11, p. 3337, 2020.
- BIOSYSTEMS, L. *What is digital pathology*. 2021. <<https://www.leicabiosystems.com/knowledge-pathway/what-is-digital-pathology/>>. Accessed: 2021-06-19.
- BOLÓN-CANEDO, V.; SÁNCHEZ-MARONO, N.; ALONSO-BETANZOS, A.; BENÍTEZ, J. M.; HERRERA, F. A review of microarray datasets and applied feature selection methods. *Information Sciences*, Elsevier, v. 282, p. 111–135, 2014.
- BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 111–118.
- BOUZIANE, A.; BOUMALI, S.; BERKANE, N.; GUENDOZ, F. S. A hybrid approach for automatic breast cancer detection. In: IEEE. *2020 International Conference on e-Health and Bioengineering (EHB)*. [S.l.], 2020. p. 1–4.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- CECOTTI, H.; BELAÏD, A. Rejection strategy for convolutional neural network by adaptive topology applied to handwritten digits recognition. In: IEEE. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. [S.l.], 2005. p. 765–769.
- CESSIE, S. L.; HOUWELINGEN, J. C. V. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 41, n. 1, p. 191–201, 1992.
- CHOUGRAD, H.; ZOUAKI, H.; ALHEYANE, O. Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing*, Elsevier, v. 392, p. 168–180, 2020.
- CLEARY, J. G.; TRIGG, L. E. K*: An instance-based learner using an entropic distance measure. In: *Machine Learning Proceedings 1995*. [S.l.]: Elsevier, 1995. p. 108–114.

COCCIA, M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. *Technology in Society*, Elsevier, v. 60, p. 101198, 2020.

COHEN, J. P.; MORRISON, P.; DAO, L.; ROTH, K.; DUONG, T. Q.; GHASSEMI, M. Covid-19 image data collection: Prospective predictions are the future. *arXiv pre-print arXiv:2006.11988*, 2020.

COLDPLAY. *Fix You*. [S.l.]: Parlophone Records, 2005. Letra de Música.

CONCI, A.; AZEVEDO, E.; LETA, F. R. Computação gráfica—teoria e prática,[v. 2]. *Rio de Janeiro*, 2008.

CRUZ-ROA, A.; BASAVANHALLY, A.; GONZÁLEZ, F.; GILMORE, H.; FELDMAN, M.; GANESAN, S.; SHIH, N.; TOMASZEWSKI, J.; MADABHUSHI, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2014: Digital Pathology*. [S.l.], 2014. v. 9041, p. 904103.

DABASS, M.; VIG, R.; VASHISTH, S. Five-grade cancer classification of colon histology images via deep learning. In: *CRC Press*. [S.l.: s.n.], 2018. p. 18.

DABEER, S.; KHAN, M. M.; ISLAM, S. Cancer diagnosis in histopathological image: Cnn based approach. *Informatics in Medicine Unlocked*, Elsevier, p. 100231, 2019.

DAPSON, R.; HOROBIN, R. Dyes from a twenty-first century perspective. *Biotechnic & Histochemistry*, Taylor & Francis, v. 84, n. 4, p. 135–137, 2009.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, v. 7, n. Jan, p. 1–30, 2006.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255.

ELLEUCH, M.; MAALEJ, R.; KHERALLAH, M. A new design based-svm of the cnn classifier architecture with dropout for offline arabic handwritten recognition. *Procedia Computer Science*, Elsevier, v. 80, p. 1712–1723, 2016.

ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, Nature Publishing Group, v. 542, n. 7639, p. 115–118, 2017.

FAN, R.-E.; CHANG, K.-W.; HSIEH, C.-J.; WANG, X.-R.; LIN, C.-J. Liblinear: A library for large linear classification. *Journal of machine learning research*, v. 9, n. Aug, p. 1871–1874, 2008.

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.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, NIH Public Access, v. 33, n. 1, p. 1, 2010.

FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, Taylor & Francis, v. 32, n. 200, p. 675–701, 1937.

FU, R.; LI, B.; GAO, Y.; WANG, P. Content-based image retrieval based on cnn and svm. In: IEEE. *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*. [S.l.], 2016. p. 638–642.

GAD, A. F. Convolutional neural networks. In: *Practical Computer Vision Applications Using Deep Learning with CNNs*. [S.l.]: Springer, 2018. p. 183–227.

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.

GETREUER, P. A survey of gaussian convolution algorithms. *Image Processing On Line*, v. 2013, p. 286–310, 2013.

GOLATKAR, A.; ANAND, D.; SETHI, A. Classification of breast cancer histology using deep learning. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 837–844.

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

GUIMARÃES, D. P.; MANTUAN, L. A.; OLIVEIRA, M. A. D.; JUNIOR, R. L.; COSTA, A. M. D.; ROSSI, S.; FAVA, G.; TAVEIRA, L. N.; GIARDINA, K. M.; TALARICO, T. et al. The performance of colorectal cancer screening in brazil: the first two years of the implementation program in barretos cancer hospital. *Cancer Prevention Research*, AACR, v. 14, n. 2, p. 241–252, 2021.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.

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.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Identity mappings in deep residual networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 630–645.

HINTON, G.; DENG, L.; YU, D.; DAHL, G.; MOHAMED, A.-r.; JAITLY, N.; SENIOR, A.; VANHOUCHE, V.; NGUYEN, P.; KINGSBURY, B. et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, v. 29, 2012.

HINTON, G. E. Learning multiple layers of representation. *Trends in cognitive sciences*, Elsevier, v. 11, n. 10, p. 428–434, 2007.

HINTON, G. E. Deep belief networks. *Scholarpedia*, v. 4, n. 5, p. 5947, 2009.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

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

JAYACHANDRAN, S.; GHOSH, A. Deep transfer learning for texture classification in colorectal cancer histology. p. 173–186, 2020.

JIANG, Y.; CHEN, L.; ZHANG, H.; XIAO, X. Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PloS one*, Public Library of Science San Francisco, CA USA, v. 14, n. 3, p. e0214587, 2019.

JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. *arXiv preprint arXiv:1302.4964*, 2013.

KASSANI, S. H.; KASSANI, P. H. A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell*, Elsevier, v. 58, p. 76–83, 2019.

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

KHAN, M. A.; JAVED, M. Y.; SHARIF, M.; SABA, T.; REHMAN, A. Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In: IEEE. *2019 international conference on computer and information sciences (ICCIS)*. [S.l.], 2019. p. 1–7.

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

KITTLER, J.; ALKOOT, F. M. Sum versus vote fusion in multiple classifier systems. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 25, n. 1, p. 110–115, 2003.

- KITTLER, J.; HATEF, M.; DUIN, R. P.; MATAS, J. On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 20, n. 3, p. 226–239, 1998.
- KOLTER, J. Z.; NG, A. Y. Regularization and feature selection in least-squares temporal difference learning. In: *Proceedings of the 26th annual international conference on machine learning*. [S.l.: s.n.], 2009. p. 521–528.
- KONONENKO, I.; ROBNIK-SIKONJA, M.; POMPE, U. Relieff for estimation and discretization of attributes in classification, regression, and ilp problems. *Artificial intelligence: methodology, systems, applications*, Citeseer, p. 31–40, 1996.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- KUMAR, S.; SHARMA, S. Sub-classification of invasive and non-invasive cancer from magnification independent histopathological images using hybrid neural networks. *Evolutionary Intelligence*, Springer, p. 1–13, 2021.
- LAI, M. Deep learning for medical image segmentation. *arXiv preprint arXiv:1505.02000*, 2015.
- LAL, T. N.; CHAPELLE, O.; WESTON, J.; ELISSEEFF, A. Embedded methods. In: *Feature extraction*. [S.l.]: Springer, 2006. p. 137–165.
- LE, Q. V. Building high-level features using large scale unsupervised learning. In: IEEE. *2013 IEEE international conference on acoustics, speech and signal processing*. [S.l.], 2013. p. 8595–8598.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation*, MIT Press, v. 1, n. 4, p. 541–551, 1989.
- LEE, J.-G.; JUN, S.; CHO, Y.-W.; LEE, H.; KIM, G. B.; SEO, J. B.; KIM, N. Deep learning in medical imaging: general overview. *Korean journal of radiology*, v. 18, n. 4, p. 570–584, 2017.
- LI, G.; LIU, J.; JIANG, C.; ZHANG, L.; LIN, M.; TANG, K. Relief r-cnn: Utilizing convolutional features for fast object detection. In: SPRINGER. *International Symposium on Neural Networks*. [S.l.], 2017. p. 386–394.
- LI, K.; WANG, L. Ensemble methods of face recognition based on bit-plane decomposition. In: IEEE. *2009 International Conference on Computational Intelligence and Natural Computing*. [S.l.], 2009. v. 1, p. 194–197.

Li, Y.; Lan, L.; Hu, S.; Gao, P.; Dai, X.; He, P.; Li, X.; Peng, J. Fully printed top-gate metal–oxide thin-film transistors based on scandium-zirconium-oxide dielectric. *IEEE Transactions on Electron Devices*, v. 66, n. 1, p. 445–450, 2019.

LITJENS, G.; SÁNCHEZ, C. I.; TIMOFEEVA, N.; HERMSEN, M.; NAGTEGAAL, I.; KOVACS, I.; KAA, C. H.-V. D.; BULT, P.; GINNEKEN, B. V.; LAAK, J. V. D. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, Nature Publishing Group, v. 6, p. 26286, 2016.

MACENKO, M.; NIETHAMMER, M.; MARRON, J. S.; BORLAND, D.; WOOSLEY, J. T.; GUAN, X.; SCHMITT, C.; THOMAS, N. E. A method for normalizing histology slides for quantitative analysis. In: IEEE. *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. [S.l.], 2009. p. 1107–1110.

MAHBOD, A.; ELLINGER, I.; ECKER, R.; SMEDBY, Ö.; WANG, C. Breast cancer histological image classification using fine-tuned deep network fusion. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 754–762.

MAJTNER, T.; YILDIRIM-YAYILGAN, S.; HARDEBERG, J. Y. Combining deep learning and hand-crafted features for skin lesion classification. In: IEEE. *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. [S.l.], 2016. p. 1–6.

MATHWORKS. *Transfer Learning with Deep Network Designer*. 2021. <<https://www.mathworks.com/help/deeplearning/ug/transfer-learning-with-deep-network-designer.html>>. Accessed: 2021-04-26.

MATLAB. *9.6.0.1072779 (R2019a)*. Natick, Massachusetts: The MathWorks Inc., 2019.

MCALLISTER, P.; ZHENG, H.; BOND, R.; MOORHEAD, A. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Computers in biology and medicine*, Elsevier, v. 95, p. 217–233, 2018.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 807–814.

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

NG, A. Y. Feature selection, l_1 vs. l_2 regularization, and rotational invariance. In: *Proceedings of the twenty-first international conference on Machine learning*. [S.l.: s.n.], 2004. p. 78.

- NIU, X.-X.; SUEN, C. Y. A novel hybrid cnn–svm classifier for recognizing handwritten digits. *Pattern Recognition*, Elsevier, v. 45, n. 4, p. 1318–1325, 2012.
- NOVITASARI, D. C. R.; HENDRADI, R.; CARAKA, R. E.; RACHMAWATI, Y.; FANANI, N. Z.; SYARIFUDIN, A.; TOHARUDIN, T.; CHEN, R. C. Detection of covid-19 chest x-ray using support vector machine and convolutional neural network. *Commun. Math. Biol. Neurosci.*, v. 2020, p. Article–ID, 2020.
- PAPASTERGIOU, T.; ZACHARAKI, E. I.; MEGALOOIKONOMOU, V. Tensor decomposition for multiple-instance classification of high-order medical data. *Complexity*, Hindawi, v. 2018, 2018.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011.
- PEEMEN, M.; MESMAN, B.; CORPORAAL, H. Speed sign detection and recognition by convolutional neural networks. In: *Proceedings of the 8th international automotive congress*. [S.l.: s.n.], 2011. p. 162–170.
- POLIKAR, R. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, IEEE, v. 6, n. 3, p. 21–45, 2006.
- RAHHAL, M. M. A.; BAZI, Y.; ABDULLAH, T.; MEKHALFI, M. L.; ALHICHRI, H.; ZUAIR, M. Learning a multi-branch neural network from multiple sources for knowledge adaptation in remote sensing imagery. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 10, n. 12, p. 1890, 2018.
- RAKHLIN, A.; SHVETS, A.; IGLOVIKOV, V.; KALININ, A. A. Deep convolutional neural networks for breast cancer histology image analysis. In: SPRINGER. *international conference image analysis and recognition*. [S.l.], 2018. p. 737–744.
- REHMAN, A.; KHALID, Z.; ASGHAR, M. A.; KHAN, M. J. et al. Kinship verification using deep neural network models. In: IEEE. *2019 International Symposium on Recent Advances in Electrical Engineering (RAEE)*. [S.l.], 2019. v. 4, p. 1–6.
- RIBEIRO, M. G.; NEVES, L. A.; NASCIMENTO, M. Z. do; ROBERTO, G. F.; MARTINS, A. S.; TOSTA, T. A. A. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. *Expert Systems with Applications*, Elsevier, v. 120, p. 262–278, 2019.
- 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*, Elsevier, v. 84, p. 134–143, 2019.

ROBNIK-ŠIKONJA, M.; KONONENKO, I. An adaptation of relief for attribute estimation in regression. In: *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*. [S.l.: s.n.], 1997. v. 5, p. 296–304.

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

SAGI, O.; ROKACH, L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 8, n. 4, p. e1249, 2018.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, Public Library of Science, v. 10, n. 3, p. e0118432, 2015.

SANTOS, F. P. dos; PONTI, M. A. Robust feature spaces from pre-trained deep network layers for skin lesion classification. In: *IEEE. 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2018. p. 189–196.

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, F. P. dos; RIBEIRO, L. S.; PONTI, M. A. Generalization of feature embeddings transferred from different video anomaly detection domains. *Journal of Visual Communication and Image Representation*, Elsevier, v. 60, p. 407–416, 2019.

SANTOS, L. F. S. D.; NEVES, L. A.; ROZENDO, G. B.; RIBEIRO, M. G.; NASCIMENTO, M. Z. do; TOSTA, T. A. A. Multidimensional and fuzzy sample entropy (sampenmf) for quantifying h&e histological images of colorectal cancer. *Computers in biology and medicine*, Elsevier, v. 103, p. 148–160, 2018.

SAXENA, S.; SHUKLA, S.; GYANCHANDANI, M. Breast cancer histopathology image classification using kernelized weighted extreme learning machine. *International Journal of Imaging Systems and Technology*, Wiley Online Library, 2020.

SCHÖLKOPF, B.; SMOLA, A. J.; BACH, F. et al. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. [S.l.]: MIT press, 2002.

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.

SHI, Z.; HAO, H.; ZHAO, M.; FENG, Y.; HE, L.; WANG, Y.; SUZUKI, K. A deep cnn based transfer learning method for false positive reduction. *Multimedia Tools and Applications*, Springer, v. 78, n. 1, p. 1017–1033, 2019.

SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1285–1298, 2016.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, Springer, v. 6, n. 1, p. 1–48, 2019.

SIEGEL, R. L.; MILLER, K. D.; JEMAL, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, Wiley Online Library, v. 70, n. 1, p. 7–30, 2020.

SILVA, I. N. D. C. J. A. G. D. *Estimativa 2020: incidência do Câncer no Brasil*. 2020. <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document/estimativa-2020-incidencia-de-cancer-no-brasil.pdf>>. Acesso em: 12/05/2021.

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.

SMITH, C. Our debt to the logwood tree: the history of hematoxylin. *MLO: medical laboratory observer*, v. 38, n. 5, p. 18–20, 2006.

SPANHOL, F. A.; OLIVEIRA, L. S.; PETITJEAN, C.; HEUTTE, L. Breast cancer histopathological image classification using convolutional neural networks. In: *IEEE. 2016 international joint conference on neural networks (IJCNN)*. [S.l.], 2016. p. 2560–2567.

SRINIDHI, C. L.; CIGA, O.; MARTEL, A. L. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, Elsevier, p. 101813, 2020.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, 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.

TAN, C.; SUN, F.; KONG, T.; ZHANG, W.; YANG, C.; LIU, C. A survey on deep transfer learning. In: *SPRINGER. International conference on artificial neural networks*. [S.l.], 2018. p. 270–279.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introdução ao datamining: mineração de dados*. [S.l.]: Ciência Moderna, 2009.

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.

TITFORD, M. The long history of hematoxylin. *Biotechnic & histochemistry*, Taylor & Francis, v. 80, n. 2, p. 73–78, 2005.

TITORIYA, A.; SACHDEVA, S. Breast cancer histopathology image classification using alexnet. In: IEEE. *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*. [S.l.], 2019. p. 708–712.

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.

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

VOGADO, L. H. S.; VERAS, R. D. M. S.; ANDRADE, A. R.; ARAUJO, F. H. D. D.; SILVA, R. R. V. e; AIRES, K. R. T. Diagnosing leukemia in blood smear images using an ensemble of classifiers and pre-trained convolutional neural networks. In: IEEE. *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2017. p. 367–373.

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. Theorem of the ugly duckling. *Pattern Recognition: Human and Mechanical*, 1985.

WITTEN, I. H.; FRANK, E. Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, ACM New York, NY, USA, v. 31, n. 1, p. 76–77, 2002.

XIE, M.; JEAN, N.; BURKE, M.; LOBELL, D.; ERMON, S. Transfer learning from deep features for remote sensing and poverty mapping. v. 30, n. 1, 2016.

XUE, D.-X.; ZHANG, R.; FENG, H.; WANG, Y.-L. Cnn-svm for microvascular morphological type recognition with data augmentation. *Journal of medical and biological engineering*, Springer, v. 36, n. 6, p. 755–764, 2016.

YANG, M.; KPALMA, K.; RONSIN, J. *A survey of shape feature extraction techniques*. [S.l.]: In-Tech, 2008.

YU, C. H.; LEE, H. S.; LARA, E.; GAN, S. Adult learning across usa, canada, and new zealand: A cross-cultural study of piaac. *International Education Studies*, ERIC, v. 12, n. 5, p. 1–16, 2019.

YU, X.; DONG, H. Ptl-cfs based deep convolutional neural network model for remote sensing classification. *Computing*, Springer, v. 100, n. 8, p. 773–785, 2018.

ZAHN, J. M.; POOSALA, S.; OWEN, A. B.; INGRAM, D. K.; LUSTIG, A.; CARTER, A.; WEERARATNA, A. T.; TAUB, D. D.; GOROSPE, M.; MAZAN-MAMCZARZ, K. et al. Agemap: a gene expression database for aging in mice. *PLoS genetics*, Public Library of Science San Francisco, USA, v. 3, n. 11, p. e201, 2007.

ZENG, Z.; ZHANG, H.; ZHANG, R.; YIN, C. A novel feature selection method considering feature interaction. *Pattern Recognition*, Elsevier, v. 48, n. 8, p. 2656–2666, 2015.

ZHANG, L.; ZHOU, W.-D. Sparse ensembles using weighted combination methods based on linear programming. *Pattern Recognition*, Elsevier, v. 44, n. 1, p. 97–106, 2011.

ZHANG, Z.; JIANG, M.; YE, N. Effective multiplicative updates for non-negative discriminative learning in multimodal dimensionality reduction. *Artificial Intelligence Review*, Springer, v. 34, n. 3, p. 235–260, 2010.

ZHANG, Z.; XING, F.; SU, H.; SHI, X.; YANG, L. Recent advances in the applications of convolutional neural networks to medical image contour detection. *arXiv preprint arXiv:1708.07281*, 2017.

ZHENG, L.; YANG, Y.; TIAN, Q. Sift meets cnn: A decade survey of instance retrieval. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 5, p. 1224–1244, 2017.