UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
DEPARTAMENTO DE COMPUTAÇÃO E MATEMÁTICA

FREDERICO BARBOSA MUNIZ

**Caracterização de tecido do cólon por meio da
aplicação de aprendizagem profunda sobre imagens
hiperespectrais**

Ribeirão Preto–SP

2021

FREDERICO BARBOSA MUNIZ

**Caracterização de tecido do cólon por meio da aplicação de
aprendizagem profunda sobre imagens hiperespectrais**

Versão Corrigida

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP) da Universidade de São Paulo (USP), como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de Concentração: Computação Aplicada

Orientador: Joaquim Cezar Felipe

Ribeirão Preto–SP

2021

FREDERICO BARBOSA MUNIZ

**Characterization of colon tissue through the application of deep
learning on hyperspectral images**

Corrected Version

Dissertation presented to Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP) from the Universidade de São Paulo (USP), as part of the requirements to hold the Master of Science degree.

Field of Study: Applied Computing

Supervisor: Joaquim Cezar Felipe

Ribeirão Preto–SP

2021

Frederico Barbosa Muniz

Caracterização de tecido do cólon por meio da aplicação de aprendizagem profunda sobre imagens hiperespectrais. Ribeirão Preto–SP, 2021.

89p. : il.; 30 cm.

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Mestre em Ciências,

Área: Computação Aplicada.

Orientador: Joaquim Cezar Felipe

1. Aprendizagem profunda. 2. Imagens hiperespectrais. 3. Câncer do cólon. 4. Sistemas de apoio ao diagnóstico.

Frederico Barbosa Muniz

Caracterização de tecido do cólon por meio da aplicação de aprendizagem profunda
sobre imagens hiperespectrais

Modelo canônico de trabalho monográfico
acadêmico em conformidade com as normas
ABNT.

Trabalho aprovado. Ribeirão Preto–SP, 21 de janeiro de 2021:

Orientador:
Orientador

Professor
Convidado 1

Professor
Convidado 2

Ribeirão Preto–SP
2021

Este trabalho é inteiramente dedicado à memória do meu pai, Oloidio Muniz, que assim como outras milhões de pessoas no mundo, também perdeu a luta para o câncer. A aplicação aqui desenvolvida recebe o nome PAIH em sua homenagem.

Agradecimentos

Ao meu orientador, Prof. Dr. Joaquim Cezar Felipe pelo apoio, orientação, dedicação e ensinamentos durante todo o período de elaboração deste trabalho, inclusive durante suas férias e períodos de festividades. Agradeço também pela compreensão da distância entre minha residência e a universidade, bem como a atenção em verificar a cada semestre a viabilidade de alocação das disciplinas às sextas-feiras. Acredito que o grande docente é aquele que compreende que fatores externos à sala de aula contribuem para o bom aprendizado. Encerro esta pesquisa com um enorme conhecimento adquirido, graças a confiança depositada pelo senhor em meu trabalho. Serei eternamente grato.

Ao Prof. Dr. Luciano Bachmann e a equipe do Laboratório de Foto Biofísica do Departamento de Física da FFCLRP-USP pela disponibilização das imagens hiperespectrais utilizadas neste trabalho, bem como por toda a sua atenção nos momentos de dúvidas, tanto presencialmente quanto online.

Ao médico patologista da FMRP-USP, Prof. Dr. Sérgio Britto Garcia pela disponibilidade e contribuição imprescindível na análise e demarcação das imagens triadas. Sem a sua contribuição, não conseguiríamos avançar nesta pesquisa.

Ao Prof. Dr. Luiz Otavio Murta Junior e aos colegas do CSIM pelo suporte e disponibilização do servidor, recurso que agilizou muito a geração dos resultados deste trabalho. Também à todos os colegas e professores de disciplinas, tanto do DCM quanto do LApIS da EACH-USP pela troca de experiências.

À Profa. Dra. Fátima L. S. Nunes, por me incentivar grandemente no término dos meus estudos e me possibilitar o primeiro contato com a pesquisa científica, em especial na área de Processamento de Imagens Médicas. Com toda certeza, meu alicerce na ciência foi graças a você.

Ao Prof. Dr. Luciano Vieira de Araújo pela contribuição no meu primeiro artigo publicado no exterior, ainda durante a obtenção dos créditos na disciplina de LISI, momento que me possibilitou um grande amadurecimento científico para a escrita desta dissertação.

À minha família e a todas as pessoas próximas, que de certa forma contribuíram direta ou indiretamente para que este trabalho pudesse ser concluído. Em especial ao meu pai, que com toda a sua humildade e pouca instrução, nunca deixou de reconhecer e admirar em vida todas as minhas conquistas.

Por fim, à toda comunidade científica brasileira, que mesmo em um momento obscuro como o atual, ainda persiste com excelência visando o bem maior: a ciência.

“Do not go gentle into that good night”
(Dylan Thomas, 1951)

Resumo

MUNIZ, Frederico Barbosa. **Caracterização de tecido do cólon por meio da aplicação de aprendizagem profunda sobre imagens hiperespectrais**. 2021. 89p. Dissertação (Mestrado em Ciências) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2021.

O câncer é a segunda principal causa de morte no mundo e o câncer do cólon está entre os três tipos mais comuns. Normalmente o diagnóstico desta doença é realizado por meio de exames histológicos em lâminas contendo tecido de biópsia, coradas com Hematoxilina e Eusina. As pesquisas visando o diagnóstico baseado em imagens digitais de biópsia têm crescido com rapidez, alavancando o desenvolvimento e o aperfeiçoamento de métodos de processamento de imagens especialmente desenvolvidos ou adaptados para esta categoria. As imagens hiperespectrais, obtidas a partir da medida de absorbância em diferentes frequências de aplicação de raios infravermelhos sobre o tecido, por sua vez, apresentam relações com características histoquímicas valiosas. Neste projeto, foi desenvolvido um método para processamento de sinais hiperespectrais obtidos a partir de amostras de biópsia do tecido do cólon, com a modelagem e a implementação de um classificador para identificação de anomalias, utilizando técnicas de Aprendizagem Profunda, assim como o desenvolvimento de uma ferramenta computacional, visando auxiliar o patologista durante o processo de diagnóstico de câncer do cólon. A partir dos espectros de absorbância de infravermelho de cada pixel pertencente a regiões das amostras previamente identificadas como normais, cancerígenas ou inflamadas, o classificador desenvolvido atingiu valores de acurácia para estas três classes superiores a 94%, indicando que a ferramenta possa ser considerada promissora para o auxílio ao diagnóstico de lesões do cólon, fornecendo uma segunda opinião para os especialistas.

Palavras-chave: Aprendizagem profunda. Imagens hiperespectrais. Câncer do cólon. Sistemas de apoio ao diagnóstico.

Abstract

MUNIZ, Frederico Barbosa. **Colon tissue characterization through the application of deeplearning on hyperspectral images**. 2021. 89p. Dissertation (Master of Science) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2021.

Cancer is the second main cause of death in the world and colon cancer is among the most common ones. Usually the diagnosis of this disease is performed by histological examination on blades holding biopsy tissue, stained with Hematoxylin and Eosin. The researches aiming the diagnosis based on digital biopsy images has grown quickly, leveraging the development and improvement of processing images methods specially developed or adapted for this image category. The hyperspectral images, obtained from the absorbance measure in different frequencies of infrared ray applied to the tissue, on the other hand, present relations with valuable histochemical characteristics. In this project, we developed a method for processing hyperspectral signals obtained from biopsy samples of colon tissue, with the modeling and implementation of a classifier to identify anomalies, using Deep Learning techniques, as well as the development of a computational tool, aiming to assist pathologist during the colon cancer diagnosis process. From the infrared absorbance spectrum of each pixel of the sample areas previously identified as normal, cancerous or inflamed, the developed classifier achieved accuracy values for these three groups higher than 94%, indicating that the tool can be considered promising to support the diagnosis of colon injuries, providing a second opinion to the experts.

Keywords: Deep Learning. Hyperspectral Images. Colon Cancer. Computer-aided diagnosis.

Lista de Figuras

Figura 1 – Imagem microscópica com 40x de aumento de lâmina de biópsia de pólipo corada com H&E, utilizada em exames histológicos.	20
Figura 2 – Imagem hiperespectral de lâmina de biópsia de pólipo, gerada utilizando os valores de absorbância dos pixels nas frequências correspondentes ao número de onda igual a 1000 cm^{-1} , 1600 cm^{-1} , 2000 cm^{-1} e 2800 cm^{-1} respectivamente.	22
Figura 3 – Ampliação de uma imagem digital H&E.	27
Figura 4 – Matriz de dados de imagens: a) em escalas de cinza; b) em RGB; e c) hiperespectrais.	28
Figura 5 – a) Espectro Eletromagnético e; b) Comprimento de onda.	29
Figura 6 – a) Desdobramento das matrizes 3D de imagens em RGB e; b) em imagens hiperespectrais.	30
Figura 7 – Comparação entre dados obtidos por sensores multiespectrais e hiperespectrais.	31
Figura 8 – a) Imagem original; b) Imagem (a) binarizada e com posterior aplicação da operação de Erosão; c) Imagem (a) binarizada com posterior aplicação da operação de Dilatação e; d) Imagem (a) ajustada com aplicação do Método de Otsu.	34
Figura 9 – Neurônio Biológico.	39
Figura 10 – Neurônio Artificial.	40
Figura 11 – Funções de Ativação	41
Figura 12 – Rede com uma camada de entrada, uma camada escondida e uma camada de saída.	43
Figura 13 – Fluxo de Processamento do Algoritmo de Retropropagação do Erro. . .	43
Figura 14 – Arquitetura típica de uma rede convolucional, onde a camada é segmentada em diversos mapas de características com o objetivo de realizar a mesma operação sobre a imagem de entrada em regiões específicas. . .	47
Figura 15 – Resultado de uma convolução (direita) aplicada a uma imagem (esquerda).	47
Figura 16 – a) Rede sem aplicação do desligamento e; b) com efeito de desligamento.	51
Figura 17 – Esquema de funcionamento e dependência dos processos desenvolvidos na pesquisa.	58
Figura 18 – Conjunto de 27 imagens H&E utilizadas na identificação da imagens hiperespectrais na base de imagens e para demarcação das ROIs pelo patologista.	60

Figura 19 – Composição das principais entidades que fazem parte da estrutura de um arquivo de imagem hiperespectral no formato FSM.	60
Figura 20 – Matriz de dados da imagem hiperespectral.	61
Figura 21 – Simulação de como é feito o rearranjo de uma frequência hiperespectral para geração da uma imagem 2D: a) imagem hiperespectral FSM; b) matriz rearranjada de acordo com as resoluções dX e dY e; c) imagem 2D convertida.	62
Figura 22 – a) Simulação da demarcação realizada manualmente pelo patologista na imagem HE; b) ROI (quadrado vermelho) definida digitalmente a partir da demarcação manual, com dimensão de 50x50 pixels na imagem hiperespectral	62
Figura 23 – Banco de dados do módulo Gerador de banco de ROIs contendo informações das imagens hiperespectrais (da esquerda pra direita): nome da imagem, resolução horizontal, resolução vertical, coordenada da ROI no eixo X, coordenada da ROI no eixo Y e classe da imagem (CIS). . .	63
Figura 24 – Processo de remoção da parafina e classificação dos pixels.	66
Figura 25 – a) Interface de login. b) Interface inicial do sistema, após a autenticação. .	67
Figura 26 – Interfaces comparando a visão entre o usuário pesquisador (a) com a do usuário patologista (b)	68
Figura 27 – Resultado do pré-diagnóstico realizado pela RNA com arquitetura de Aprendizagem Profunda.	69
Figura 28 – a) imagem hiperespectral na frequência correspondente ao número de onda igual a 1600 cm^{-1} ; b) imagem H&E. Ambas fazem parte do mesmo exame, como mostra o círculo em destaque, mas possuem características espaciais distintas.	69
Figura 29 – Imagem hiperespectral e sua ROI, ambas na frequência correspondente ao número de onda igual a 1600 cm^{-1}	71
Figura 30 – Últimas colunas da base de ROIs. Na coluna 1627 constam a classificação CIS dos pixels.	71
Figura 31 – Comparação de acurácia entre os cenários propostos para a RNA com arquitetura de Aprendizagem Profunda.	73
Figura 32 – . Relação dos resultados mais altos e baixos por classe.	75
Figura 33 – Resultado do pré-diagnóstico realizado pelo Ferramental de Produção: a) imagem C18 com predominância de câncer (vermelho); b) imagem I1 com predominância de inflamação (verde) e; c) imagem S10 com predominância de tecidos saudáveis. Todas coincidiram com a classificação prévia indicada pelo patologista	76

Lista de Tabelas

Tabela 1	– Aplicações para Reconhecimento de Padrões	37
Tabela 2	– Especificações de <i>hardware</i> do ambiente de testes	59
Tabela 3	– Seleção de hiperparametros utilizados no processo de <i>Tuning</i>	72
Tabela 4	– Resultado do teste de pre-diagnóstico nas imagens utilizadas na extração das ROIs para a construção do classificador.	74

Lista de abreviaturas e siglas

2D	Two Dimensions
3D	Three Dimensions
AG	Algoritmo Genético
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
BPNN	Back-Propagation Neural Networks
CBIR	Content-Based Image Retrieval
CIS	Câncer, Inflamação e Saudável
CSS	Cascading Style Sheets
CSV	Comma-separated Values
DNA	Deoxyribonucleic Acid
FSM	Splitty Master Split File
FTIR	Fourier Transform Infrared
GB	Gigabyte
H&E	Hematoxilina e Eosina
HTML	HyperText Markup Language
ID	Identity
INCA	Instituto Nacional do Câncer
IRT	Infravermelho Térmico
JSON	JavaScript Object Notation
MLP	Multilayer Perceptron
OMS	Organização Mundial da Saúde
PAIH	Prediagnosis using Artificial Intelligence and Hyperspectral Image
PCA	Principal Components Analysis

PNG	Portable Network Graphics
PSMA	Prostate-specific Membrane Antigen
ReLU	Rectified Linear Units
RGB	Red, Green and Blue
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
RNN	Recurrent Neural Networks
ROI	Region of Interest
TIFF	Tagged Image File Format
UI	User Interface
WHO	World Health Organization

Sumário

	Introdução	18
1	FUNDAMENTAÇÃO TEÓRICA	27
1.1	Imagem Digital	27
1.1.1	Imagem Espectral	28
1.1.2	Imagem Hiperespectral	29
1.2	Limiarização de Imagens	32
1.2.1	Método de Otsu	33
1.3	Morfologia Matemática	33
1.3.1	Erosão	34
1.3.2	Dilatação	35
1.4	Reconhecimento de Padrões	36
1.5	Aprendizado de Máquina	38
1.6	Redes Neurais Artificiais	39
1.6.1	Perceptron multicamadas	42
1.6.2	Back-propagation	42
1.6.3	Aprendizagem Profunda	45
1.6.4	Redes Convolucionais	46
1.6.5	Treinamento em redes profundas	49
1.6.6	Pré-treinamento não supervisionado	50
1.6.7	Dropout	50
2	ESTADO DA ARTE	52
3	METODOLOGIA	56
3.1	Ambiente de testes	58
3.1.1	Base de imagens	59
3.2	Gerador do banco de ROIs	60
3.3	Treinamento e otimização da rede	63
3.4	Ferramental de Produção	65
3.5	Interface Web PAIH	67
4	RESULTADOS E DISCUSSÕES	70
4.1	Base de ROIs	70
4.2	Tuning	72
4.3	Avaliação	73

4.4	Ferramenta de Pré-diagnóstico	74
5	CONCLUSÕES	78
	Referências Bibliográficas	81

Introdução

Segundo os dados mais recentes da Organização Mundial de Saúde (OMS), o câncer é a segunda principal causa de morte no mundo e foi responsável por 9,6 milhões de mortes em 2018. A nível global, uma em cada seis mortes é relacionada à doença e aproximadamente 70% das mortes por câncer ocorrem em países de baixa e média renda. A apresentação tardia e o diagnóstico e tratamento inacessíveis são comuns. Em 2017, apenas 26% dos países de baixa renda relataram ter serviços de patologia disponíveis no setor público; por outro lado, mais de 90% dos países de alta renda relataram que os serviços de tratamento estão disponíveis, em comparação com menos de 30% dos países de baixa renda. O impacto econômico do câncer é significativo e está aumentando. No último levantamento global, o custo anual total da doença em 2010 foi estimado em aproximadamente US\$ 1,16 trilhão (PAHO, 2018).

No Brasil, segundo a publicação “*Estimativa 2020, Incidência de Câncer no Brasil*”, publicada em 2019 pelo Instituto Nacional do Câncer (INCA), para cada ano do triênio 2020 a 2022 aponta que ocorrerão por volta de 625 mil casos novos de câncer (450 mil, excluindo os casos de câncer de pele não melanoma). O câncer de pele não melanoma será o mais incidente (com 177 mil casos), seguido pelos cânceres de mama e próstata (66 mil cada), cólon e reto (41 mil casos), pulmão (30 mil casos) e estômago (21 mil casos). O cálculo global, segundo a metodologia de Mathers et al. (2003), estima a ocorrência de aproximadamente 685 mil casos novos. O câncer de pele não melanoma representará 27,1% de todos os casos de câncer em homens e 29,5% em mulheres. Já os tipos de câncer mais frequentes em homens - excluindo o câncer de pele não melanoma - serão próstata (29,2%), cólon e reto (9,1%), pulmão (7,9%), estômago (5,9%) e cavidade oral (5,0%). Nas mulheres - exceto também o câncer de pele não melanoma - os cânceres de mama (29,7%), cólon e reto (9,2%), colo do útero (7,4%), pulmão (5,6%) e tireoide (5,4%) corresponderão entre os principais. A estimativa da OMS indica que, em 2025, os casos de câncer cresçam em até 50% no Brasil em decorrência do aumento e do envelhecimento da população. Atualmente, a doença é a segunda causa de morte no país. Para o Brasil, estimam-se, para cada ano do triênio de 2020 a 2022, um total de 20.520 casos de câncer de cólon e reto em homens e 20.470 em mulheres. Esses valores correspondem a um risco estimado de 19,63 casos novos a cada 100 mil homens e 19,03 para cada 100 mil mulheres.

As mortes causadas pelo câncer do cólon estão entre os três tipos mais comuns em todo o mundo. No ano de 2018 foram registrados no mundo cerca de 1,8 milhão de casos, ficando atrás somente dos casos de mama e pulmão que já alcançaram também em 2018, mais de 2,09 milhões de casos cada. Dados do INCA de 2019 revelam que, sem considerar os tumores de pele não melanoma, o câncer de cólon e reto em homens é o segundo mais incidente nas Regiões Sudeste (28,62/100 mil) e Centro-Oeste (15,40/100 mil). Na Região Sul (25,11/100 mil), é o terceiro tumor mais frequente. Enquanto nas Regiões Nordeste (8,91/100 mil) e Norte (5,27/100 mil), ocupa a quarta posição. Para as mulheres, é o segundo mais frequente nas Regiões Sudeste (26,18/100 mil) e Sul (23,65/100 mil). Nas Regiões Centro-Oeste (15,24/100 mil), Nordeste (10,79/100 mil) e Norte (6,48/100 mil) é o terceiro mais incidente.

O câncer do cólon é uma doença caracterizada pelo desenvolvimento insólito e descontrolado de células que são encontradas no intestino grosso. Se o tumor se estende para a parte inferior do cólon (reto), o câncer pode ser colorretal (INCA, 2020). A maioria dos tumores do cólon surgem de pequenas massas benignas chamadas pólipos adenomatosos, que se formam nas paredes internas do intestino. Alguns destes pólipos podem tornar-se tumores malignos do cólon ao longo do tempo se eles não são removidos por meio do procedimento de colonoscopia. As células tumorais do cólon invadem e causam danos ao tecido adjacente ao tumor, ocasionando sérias complicações. As células de tumores malignos podem viajar através da corrente sanguínea e sistema linfático e se espalhar para outras partes do corpo, causando a metástase.

Não há uma causa única do câncer de cólon. O risco de câncer de cólon é alto nos seguintes casos: idade superior a 60 anos, alimentação rica em gordura ou carne vermelha ou processada, uso de cigarro e álcool, existência de pólipos colorretais ou doença inflamatória intestinal, também chamada de doença de Crohn¹ ou colite ulcerativa. Algumas doenças hereditárias também aumentam o risco de desenvolver câncer de cólon. Dois dos mais frequentes são: a polipose adenomatosa familiar e o câncer colorretal hereditário sem polipose (ou síndrome de Lynch²). Assim como a maioria dos diagnósticos de câncer, grande parte das investigações de câncer de cólon são realizadas por meio de exames histológicos em lâminas contendo tecido de biópsia, coradas com Hematoxilina & Eusina (H&E) (Figura 1). A histologia é o estudo da anatomia microscópica de células e tecidos de plantas e animais, e de como essas estruturas se organizam para constituir os diferentes órgãos. A histologia tem um grande papel nas ciências médicas e biológicas, pois é através do estudo comparativo de tecidos saudáveis e patológicos que podemos realizar uma série de diagnósticos (JUNQUEIRA et. al, 2013; KIERSZENBAUM, 2012).

¹ Doença de Crohn é caracterizada pela inflamação do revestimento do trato digestivo, que muitas vezes se espalha profundamente nos tecidos afetados. Disponível em: <<https://www.nhs.uk/conditions/crohns-disease/>>

² A síndrome de Lynch é uma doença autossômica dominante, que prejudica o sistema de reparo do DNA. Disponível em: <https://www.cdc.gov/genomics/disease/colorectal_cancer/lynch.htm>

Figura 1 – Imagem microscópica com 40x de aumento de lâmina de biópsia de pólipos corada com H&E, utilizada em exames histológicos.

Fonte: próprio autor.

As pesquisas em torno de imagens digitais na medicina em geral têm crescido com rapidez. O Processamento de Imagens Digitais tem se tornado um elemento relevante na Ciência da Informação e muitas atividades científicas simplesmente não existiriam sem ele. A aplicação do Processamento de Imagens Digitais pode ser encontrada nas mais diversas áreas médicas, tais como Radiologia, Microscopia e Genética, entre outras (GONZALES et. al, 2007).

O rápido e contínuo avanço dos sistemas de aquisição, armazenamento e análise das imagens médicas digitais, juntamente com os sistemas de diagnóstico auxiliado por computador, alavancou o desenvolvimento e o aperfeiçoamento de métodos de processamento de imagens especialmente desenvolvidos ou adaptados para esta categoria de imagens (DOUGHERTY, 2009). O processamento e a análise de imagens médicas muitas vezes representam um desafio complexo, pois envolvem etapas como a aquisição da imagem, armazenamento, manipulação e a sua exibição. A análise de imagens propriamente dita envolve ainda extrair suas medidas, preferencialmente de modo automático, e utilizar estas medidas para a Recuperação de Imagem Baseada em Conteúdo (CBIR – do inglês *Content-Based Image Retrieval*) ou para a sua classificação automática. Procedimentos como este podem ser usados, por exemplo, para classificar uma imagem microscópica - obtida com incidência de infravermelho - em lâminas com amostras de biópsia do tecido do cólon em parafina, como: saudável, inflamada ou com câncer. Há trabalhos que utilizam o espectro infravermelho para a detecção de doenças e de outras características, utilizando diferentes configurações de equipamentos de infravermelho para esse fim. Alguns destes

trabalhos, abordando técnicas semelhantes às utilizadas nesta pesquisa, são descritas no capítulo 2 (Estado da Arte).

O princípio geral da espectroscopia infravermelha consiste em focar um feixe multifrequencial de radiação infravermelha sobre uma determinada amostra. Cada molécula da amostra absorve a radiação infravermelha em uma frequência específica, de acordo com seu perfil de vibração molecular. As frequências infravermelhas que não foram absorvidas geralmente passam pela amostra sem interação. A diferença entre o feixe infravermelho incidente e o feixe medido após a interação com as moléculas da amostra, permite identificar quais frequências foram absorvidas por ela. Dessa forma, é possível identificar quais moléculas ou ligações químicas estão presentes na amostra em questão (STUART, 2004). As frequências absorvidas são geralmente apresentadas como um espectro de absorção, que relaciona as frequências em unidades de números de onda, com a intensidade da absorção. As técnicas de infravermelho são geralmente referidas como Espectroscopia de Infravermelho com Transformada de Fourier (FTIR), uma vez que a Transformação de Fourier³ é comumente aplicada durante o cálculo do espectro.

O desenvolvimento de detectores multicanais na década de 1990 ampliaram as possibilidades de aplicação da espectroscopia no infravermelho (LASCH et al., 2006). Talvez o avanço mais significativo tenha sido o acoplamento de técnicas de microscopia com o sistema FTIR tradicional, culminando na criação da técnica denominada Microespectroscopia Infravermelha ou Micro-FTIR (LIMA, 2016). Os modelos de microscópios infravermelho escaneiam a amostra, coletando valores de absorbância para espectros de frequência em posições subsequentes. Primeiramente, o sistema mede o espectro infravermelho em um determinado ponto de amostragem. Então, a amostra é movida por um deslocador micrométrico e um segundo espectro é coletado na nova posição. Deslocamentos laterais típicos variam de 6 a 20 μm . Os espectros coletados são organizados de acordo com a posição em que foram medidos, podendo produzir uma imagem para cada diferente frequência. Cada pixel (ou ponto) escaneado possui um espectro infravermelho coletado naquele ponto da amostra. Como cada pixel contém um espectro infravermelho completo, o conjunto de dados forma um cubo multidimensional, comumente referido como uma imagem hiperespectral (LIMA, 2016). Ao final, o arquivo da imagem é construído não somente com informações espaciais, mas também com cada uma das frequências hiperespectrais. Neste estudo, todas as imagens hiperespectrais utilizadas possuem exatas 1626 frequências em um intervalo de número de onda, variando de 750 cm^{-1} a 4000 cm^{-1} (de 2 em 2). A Figura 2 apresenta imagens geradas em quatro diferentes frequências, aplicadas em uma das amostras que foram utilizadas neste trabalho.

O potencial dessa tecnologia tem sido abordado por estudos de prova de conceito

³ A transformada de Fourier é uma representação do sinal original no domínio da frequência. Disponível em: <<https://www.ime.usp.br/oliveira/ELE-Fourier1.pdf>>

Figura 2 – Imagem hiperespectral de lâmina de biópsia de pólipo, gerada utilizando os valores de absorbância dos pixels nas frequências correspondentes ao número de onda igual a 1000 cm^{-1} , 1600 cm^{-1} , 2000 cm^{-1} e 2800 cm^{-1} respectivamente.

Fonte: próprio autor.

nas últimas décadas. Além disso, usos biomédicos dessas informações espectroscópicas em geral têm sido relatados, mostrando que essas técnicas podem desempenhar um papel significativo na histopatologia, citologia, alvos de biópsia, alvos cirúrgicos, monitoramento de tratamento e estudos de drogas (BYRNE, 2015). Possíveis cenários de aplicação incluem a medição de núcleos-alvo para triagem citológica ou o mapeamento de áreas maiores para o diagnóstico de doenças. Devido ao grande volume de informações geradas por este tipo de imagem, essas tecnologias têm chamado a atenção nas áreas de pesquisa relacionadas à Ciência da Computação, havendo estudos sobre a análise dessas imagens com o auxílio de técnicas de Inteligência Artificial.

O Aprendizado de Máquina, por exemplo, é um subcampo da Ciência da Computação que evoluiu do estudo de Reconhecimento de Padrões e da Teoria do Aprendizado Computacional em Inteligência Artificial e que tem sido utilizado largamente em pesquisas relacionadas ao pré-diagnóstico de doenças. Ele explora o estudo e construção de algoritmos que podem aprender de seus erros e fazer previsões sobre dados. Tais algoritmos operam construindo um modelo a partir de amostras de entrada a fim de fazer previsões ou decisões guiadas pelos dados, ao invés de simplesmente seguindo instruções programadas. Há ainda uma série de outras subáreas dentro do Aprendizado de Máquina, como as Redes Neurais Artificiais (RNA), e áreas derivadas como as Redes Neurais Convolucionais (RNC) e a RNA com arquitetura de Aprendizagem Profunda ou Redes Neurais Profundas. A primeira consiste de modelos computacionais inspirados pelo sistema nervoso central e que são capazes de realizar o aprendizado bem como o Reconhecimento de Padrões. Já a segunda, além de apresentar praticamente as mesmas características das RNCs, possui ainda um conjunto de algoritmos que tentam modelar abstrações de alto nível de dados usando um grafo profundo com várias camadas de processamento, compostas de inúmeras transformações lineares e não lineares.

Assim, nesta pesquisa é apresentada uma proposta computacional para o auxílio ao diagnóstico de câncer do cólon que consiste em utilizar imagens hiperespectrais obtidas de lâminas com amostras de biópsia do tecido do cólon em parafina, caracterizando pixels

para que, a seguir, sejam aplicadas técnicas de RNA com arquitetura de Aprendizagem Profunda para identificar características das áreas da amostra, auxiliando os patologistas a atingirem diagnósticos mais confiáveis e precisos, por meio de uma ferramenta computacional gráfica, complementando os atuais exames histológicos. O restante deste capítulo está estruturado da seguinte forma: a próxima seção apresenta a motivação para o desenvolvimento da aplicação utilizando imagens hiperespectrais em conjunto com técnicas de RNA com arquitetura de Aprendizagem Profunda; as seções seguintes apresentam os objetivos do trabalho, a metodologia utilizada para desenvolvimento desta pesquisa e, por fim, a estrutura dos demais capítulos desta dissertação.

Motivação

Um dos objetivos de se realizar o processamento e a análise computacional de imagens médicas é o Reconhecimento de Padrões para auxílio ao diagnóstico. Ferramentas projetadas para esta tarefa, embora nunca prescindindo da avaliação do especialista, podem contribuir efetivamente para o aumento da precisão e a redução da subjetividade no processo de diagnóstico baseado em imagens. Recentemente, as análises baseadas em imagens oriundas de microscópios óticos, obtidas sobre lâminas de material histológico, têm sido sistematicamente exploradas, apesar das limitações observadas, intrínsecas ao próprio contexto deste tipo de modalidade de imagens. As imagens hiperespectrais obtidas a partir da incidência de ondas na gama de frequências do infravermelho sobre lâminas contendo amostras histológicas, oferecem uma gama de informações relevantes que podem complementar as informações obtidas com luz visível.

Esta hipótese de complementaridade representa a principal motivação da presente proposta. Este projeto investiga o alcance e as limitações referentes ao uso de imagens hiperespectrais na região do infravermelho para o Reconhecimento de Padrões e a identificação de regiões de amostras de tecidos que apresentam anomalias. Neste contexto, há que se considerar a complexidade inerente à grande quantidade de pixels a serem analisados, potencializada pelo fato de haver um espectro composto por centenas ou até milhares de diferentes frequências para cada pixel. Para processar o alto volume de informações, com o objetivo de construir um classificador, as técnicas de RNA com arquitetura de Aprendizagem Profunda surgem como alternativas eficientes e efetivas. Estas técnicas oferecem para uma determinada aplicação a possibilidade de aprender sobre as características e dissimilaridades de uma determinada imagem hiperespectral, a partir dos registros de classificações anteriores, que foram utilizadas para treinar um classificador sob um processo supervisionado.

A razão da escolha do uso de algoritmos de RNA com arquitetura de Aprendizagem Profunda vem do fato de que eles têm sido usados em diversos campos de aplicação, e têm

produzido alguns dos melhores resultados dentre as possíveis soluções com Aprendizado de Máquina (DENG, 2014). Com isso, dentre as técnicas mais implementadas em Inteligência Artificial, destaca-se a RNA com arquitetura de Aprendizagem Profunda, na qual, devido a sua robustez e à qualidade dos resultados obtidos, têm sido largamente utilizadas em aplicações que manipulam grandes volumes de dados (*Big Data*, por exemplo), e em especial para o Processamento de Imagens Digitais. Desta forma, a RNA com arquitetura de Aprendizagem Profunda apresentou-se como uma opção promissora para a aplicação neste trabalho.

Objetivos

Este trabalho tem por objetivo principal investigar, implementar e avaliar um método e uma ferramenta para caracterização de regiões anômalas de tecidos do cólon em amostras de biópsia, a partir de espectros obtidos pela aplicação de infravermelho sobre as mesmas, visando o auxílio ao diagnóstico de lesões inflamatórias e cancerígenas. A hipótese principal foi a de investigar se as imagens hiperespectrais permitem identificar regiões anômalas desses tecidos a partir do espectro dos seus pixels. Neste sentido, este trabalho possui os seguintes objetivos específicos:

- I. projetar um classificador baseado em RNA com arquitetura de Aprendizagem Profunda que, para um conjunto de regiões de interesse (ROI – do inglês *Region of Interest*) de imagens hiperespectrais, contendo pixels (com seu respectivo espectro) rotulados como câncer, inflamação ou saudável (aqui denominadas como classificação CIS), previamente marcados nas ROIs por um especialista (médico patologista) a partir da avaliação visual de imagens obtidas de lâminas coradas com H&E, identifique pixels candidatos a conter anomalias e, conseqüentemente, permita a marcação automática de regiões suspeitas de conter lesões;
- II. avaliar o método desenvolvido, visando obter indicadores de acurácia, para quantificar o grau de confiabilidade da proposta; e
- III. implementar uma ferramenta computacional que permita a utilização do método para classificar, identificar e apresentar ao usuário regiões contendo predominância de tecido cancerígeno, inflamado ou saudável (classificação CIS).

A utilização da RNA com arquitetura de Aprendizagem Profunda na implementação de uma nova ferramenta para o pré-diagnóstico de câncer de cólon, por meio de imagens hiperespectrais, otimizará o processamento dos dados, devido ao grande volume de informações obtidas em cada arquivo de imagem. Finalmente, o desenvolvimento desta

aplicação permitirá auxiliar patologistas a obterem apoio ao diagnóstico por meio de um recurso que complementa as informações obtidas a partir da inspeção visual das lâminas coradas com H&E. Com o desenvolvimento de uma ferramenta *Web*, de fácil distribuição e com suporte ao trabalho colaborativo, acreditamos que esta tecnologia possa ser disponibilizada de forma mais rápida e simples em ambientes de apoio ao diagnóstico.

Metodologia

Inicialmente, realizamos uma revisão bibliográfica sobre RNA com arquitetura de Aprendizagem Profunda, imagens hiperespectrais, câncer colorretal e sistemas de apoio ao diagnóstico. Adicionalmente, realizamos uma pesquisa sobre as principais ferramentas de Inteligência Artificial propostas na literatura. Todo este projeto foi desenvolvido com a colaboração de pesquisadores do Laboratório de Foto Biofísica do Departamento de Física da FFCLRP-USP, responsável pela aquisição das imagens hiperespectrais e pelo planejamento e definição dos experimentos.

Após completar os estudos de fundamentação teórica do trabalho, realizamos uma pesquisa de diferentes *frameworks* e bibliotecas de Inteligência Artificial para o suporte ao desenvolvimento da nossa ferramenta, que possibilitasse a classificação de pixels em imagens hiperespectrais e retornasse o resultado do pré-diagnóstico de forma gráfica, tornando acessível sua utilização. Estas atividades tiveram por objetivo construir uma base sólida de conhecimento para compreender os desafios das diferentes abordagens para a classificação de pixels em imagens hiperespectrais em conjunto com técnicas de Inteligência Artificial, bem como para planejar e projetar a criação da nova abordagem proposta nesta dissertação e, conseqüentemente, da ferramenta de pré-diagnóstico.

Após os estudos que embasaram o desenvolvimento deste trabalho, organizamos um conjunto de imagens hiperespectrais contendo diferentes amostras de biópsia do tecido do cólon em parafina que serão explicadas com mais detalhes no capítulo 3 (Metodologia). Além da versão hiperespectral, as imagens de biópsia também possuem uma versão digital no formato TIFF (*Tagged Image File Format*), correspondente a coloração H&E, utilizada para identificação visual das imagens hiperespectrais, uma vez que esses arquivos são compostos de matrizes de dados e não possuem uma versão digital para a sua identificação. As imagens hiperespectrais - ou arquivos das imagens - possuem um formato específico denominado FSM (*Splitty Master Split File*) e precisaram ser convertidas no aplicativo *MatLab* para o formato CSV (*Comma-separated Values*) para que pudessem ser manipuladas no ambiente de desenvolvimento proposto. Após efetuada a conversão, foi possível gerar a imagem hiperespectral digital (também no formato TIFF) em qualquer uma das suas 1626 frequências disponíveis.

Em seguida, levantamos os requisitos para o uso desses arquivos com uma tecnologia de Inteligência Artificial que melhor trabalhasse com o modelo de RNA com arquitetura de Aprendizagem Profunda. Assim foi definida a escolha do *Keras*, uma biblioteca de RNA de código aberto escrita na linguagem de programação *Python*. O *Keras* foi projetado para permitir o desenvolvimento rápido de RNA com arquitetura de Aprendizagem Profunda, por ser fácil de usar, graças a uma estrutura modular, extensível e orientada pelo *Google TensorFlow*, uma biblioteca de código aberto para Aprendizado de Máquina, aplicável a uma ampla variedade de tarefas. Após a fase de análise dos requisitos, iniciamos o desenvolvimento da abordagem e consequentemente da ferramenta de pré-diagnóstico. A ferramenta desenvolvida nesta pesquisa, denominada PAIH (acrônimo do inglês *Prediagnosis using Artificial Intelligence and Hyperspectral image*), foi programada utilizando a linguagem de programação *Python* e suas bibliotecas aplicáveis à Web. O PAIH corresponde a integração dos diversos módulos que compõem este trabalho e maiores detalhes sobre sua estrutura e processo funcional estão descritos no capítulo 3 (Metodologia).

Finalmente, de modo a validar o funcionamento da nova ferramenta, desenvolvemos uma prova de conceito dividida em duas etapas. A primeira etapa consistiu na execução de um dos métodos mais utilizados no processo de comprovação em sistemas baseados no Aprendizado de Máquina: a Validação Cruzada. Este processo foi utilizado para demonstrar o uso da ferramenta a fim de obter resultados similares em frações diferentes da base de ROIs do classificador. Já a segunda etapa consistiu no uso de imagens hiperespectrais que não fizeram parte da base de treinamento utilizada no classificador. Este processo foi então utilizado para demonstrar o uso da ferramenta PAIH em uma situação real de pré-diagnóstico, quando uma imagem hiperespectral - que não faz parte do conjunto de imagens hiperespectrais utilizadas no classificador - é utilizada diretamente na aplicação gráfica por um usuário especialista.

Estrutura do documento

O restante do documento está estruturado da seguinte forma: o capítulo 1 apresenta a Fundamentação Teórica necessária à compreensão dos principais conceitos utilizados neste trabalho; o capítulo 2 apresenta o Estado da Arte a esta dissertação e que serviram como referência neste estudo; o capítulo 3 apresenta a Metodologia com a nossa proposta de classificação de pixels em imagens hiperespectrais, utilizando RNA com arquitetura de Aprendizagem Profunda no pré-diagnóstico em exames de amostras de biópsia do tecido do cólon em parafina; no capítulo 4 são apresentados os Resultados e Discussões da pesquisa e, por fim, o capítulo 5 apresenta a Conclusão do trabalho, assim como propostas sobre trabalhos futuros.

Fundamentação Teórica

Este capítulo apresenta uma visão geral acerca de seis conjuntos fundamentais de conceitos necessários à compreensão do contexto e das soluções tecnológicas para o desenvolvimento deste trabalho. O capítulo está estruturado da seguinte forma: a seção 1.1 apresenta uma visão geral sobre Imagem Digital, Espectral e Hiperespectral; a seção 1.2 apresenta uma visão geral sobre Limiarização; a seção 1.3 apresenta uma visão geral sobre Morfologia Matemática; a seção 1.4 apresenta uma visão geral sobre Reconhecimento de Padrões; a seção 1.5 apresenta uma visão geral sobre Aprendizado de Máquina; e por fim, a seção 1.6 apresenta uma visão geral sobre Redes Neurais Artificiais e Aprendizagem Profunda.

1.1 Imagem Digital

Uma imagem digital é uma imagem $f(x,y)$ discretizadas tanto em coordenadas espaciais quanto em brilho. Uma imagem digital pode ser considerada como sendo uma matriz cujos índices de linhas e de colunas identificam um ponto na imagem, e o correspondente valor do elemento da matriz identifica o nível de cinza naquele ponto (GONZALES, 2002). O menor elemento formador de uma imagem em duas dimensões (de coordenadas x e y) é o pixel, enquanto em imagens tridimensionais (de coordenadas x , y e z) é o denominado voxel. Na Figura 3 é possível observar que a ampliação sucessiva da imagem fornece a visualização dos pixels que compõem a mesma.

Figura 3 – Ampliação de uma imagem digital H&E.

Fonte: próprio autor.

As imagens digitais são formadas pelas coordenadas espaciais e valores numéricos para cada pixel que representam a escala de cores utilizada. Quando a imagem está representada na escala de cinza, tem-se uma matriz de dados bidimensional e cada pixel está associado a um valor de intensidade na escala de cinza (Figura 4a). Já para imagens em RGB (do inglês *Red*, *Green*, *Blue*), formadas a partir dos três canais de cores de mesmo nome, obtém-se uma matriz de dados tridimensional, ou matrizes concatenadas, cada uma com valores diferentes de intensidade - em que cada pixel terá três valores de intensidade (Figura 4b), (SILVA, 2013; PRATS-MONTALBÁN et al., 2011).

Figura 4 – Matriz de dados de imagens: a) em escalas de cinza; b) em RGB; e c) hiperespectrais.

Fonte: adaptado de SILVA (2013).

1.1.1 Imagem Espectral

Uma imagem espectral é aquela que reproduz um objeto em função do comprimento de onda que está a refletir ou emitir do objeto em questão, constituindo um conjunto de imagens representadas com diferentes comprimentos de onda. Já a resolução espectral é um conceito próprio para os sistemas sensores denominados de multiespectrais. Segundo Novo (1989), a resolução espectral é uma medida da largura das faixas espectrais e da sensibilidade do sistema sensor em distinguir entre dois níveis de intensidade do sinal de retorno. Os autores Jensen et al. (2018) destacam dois pontos importantes nesta tecnologia: o comprimento de onda detectado pelo sensor e a quantidade de faixas espectrais. A Figura 5a apresenta as diferentes regiões do espectro eletromagnético, destacando a faixa do visível, enquanto a Figura 5b mostra o comprimento de onda detectado pelas bandas de dois sistemas sensores (vermelho e azul) utilizados em sensoriamento remoto.

O primeiro sistema sensor (vermelho) tem muitas bandas espectrais e uma grande sensibilidade espectral. O outro sistema sensor (azul), possui poucas bandas e uma menor sensibilidade espectral. Comparando os dois sistemas sensores, observa-se que o primeiro pode caracterizar e distinguir melhor um objeto na imagem do que o outro sistema.

Figura 5 – a) Espectro Eletromagnético e; b) e Comprimento de onda.

Fonte: JENSEN et al. (2018).

Portanto, quanto maior o número de bandas e menor a largura do intervalo, maior a discriminação do alvo na cena e melhor a resolução espectral (MELO, 2002).

1.1.2 Imagem Hiperespectral

As imagens multiespectrais ou hiperespectrais fornecem não somente as informações espaciais da amostra, mas também dados a respeito da sua composição química e possuem um espectro de alta resolução ou alguns comprimentos de onda associados a cada pixel da imagem (SILVA, 2013). Esses espectros podem ser obtidos por diversas técnicas, como Infravermelho, Fluorescência, Raman, UV-vis etc. As técnicas de Raman e Infravermelho, geralmente, são as mais utilizadas (DE JUAN, 2009). Em especial, a imagem hiperespectral é uma tecnologia que combina imagens digitais convencionais e espectroscopia em um único sistema. No entanto, esta técnica vai além da aquisição convencional de informações espectrais e espaciais de um objeto simultaneamente. A técnica de imagem consiste na aquisição de informações de dados espaciais a partir de objetos usando uma câmera digital, enquanto a espectroscopia consiste em adquirir as características espectrais de um objeto para descrever a intensidade da luz que emergem de suas moléculas em diferentes comprimentos de onda e, assim, proporcionar uma impressão digital precisa do objeto (NUBIATO, 2016; SIVA, 2013).

Todavia, antes de realizar a manipulação de imagens hiperespectrais, é preciso convertê-la em uma matriz de dados que permita a análise de seus dados por meio de outras tecnologias. Desta forma, é preciso rearranjar a matriz 3D (para imagens em RGB, multiespectrais e hiperespectrais) em uma matriz bidimensional, onde cada pixel será considerado como uma amostra e os valores de intensidade (ou frequências) serão dispostos em linhas (Figura 6a e 6b), respeitando a ordem sugerida (SILVA, 2013). Aplicações como o *MatLab* permitem esse tipo de interação e rearranjo.

Figura 6 – a) Desdobramento das matrizes 3D de imagens em RGB e; b) em imagens hiperespectrais.

Fonte: Adaptado de SILVA (2013).

Segundo Silva (2013), de modo geral, as imagens hiperespectrais possuem uma grande vantagem com relação a aquisição dos espectros de uma amostra, pois variações espectrais (pixel a pixel) fornecem informações a respeito da distribuição de componentes em sua superfície. Considerando o uso de imagens em duas dimensões, adicionando uma nova dimensão de informação "espectro", os dados de uma imagem hiperespectral podem ser percebidos como um cubo de dados tridimensional (CHAO et al., 2001). Sendo assim, como já mencionado, o sistema fornece informação espacial, juntamente com a informação espectral para cada pixel ou linha de pixels na imagem. No caso das imagens hiperespectrais em escala de cinza, o hipercubo é criado quando centenas de imagens individuais são empilhadas umas sobre as outras, sendo que cada imagem no hipercubo contém uma quantidade de informações sobre o objeto analisado. Cada pequeno retângulo do cubo representa a região de integração espacial e espectral de um pixel da imagem.

A diferença entre imagem multiespectral e imagem hiperespectral é representada na Figura 7, onde os dados de imagens multiespectrais e hiperespectrais são vistos como cubos, com duas dimensões (x , y), representando a posição espacial, e uma terceira dimensão (λ), representando o comprimento de onda (SCHOWENGERDT, 2006). Na imagem multiespectral, a dimensão espectral é menor e de bandas espectrais mais largas, enquanto que na imagem hiperespectral há uma relativa continuidade espectral, com bandas espectrais estreitas. Pode-se observar ainda, que o eixo λ (vetor de comprimento de onda) da

imagem hiperespectral é formado por dezenas de bandas espectrais que coletam medidas de radiância espectral. Desse modo, essas imagens são capazes de obter medidas com alto nível de detalhamento espectral, tornando possível a identificação de componentes, através dos padrões de absorção e espalhamento da radiação eletromagnética. Essas informações podem ser utilizadas para caracterizar o objeto de forma mais confiável do que a imagem tradicional ou técnicas espectroscópicas (KUMAR et al., 2009; QUEVEDO et al., 2010). Assim, a combinação de imagens e espectroscopia gera uma imagem hiperespectral, possibilitando determinar as características físicas, bem como as suas distribuições espaciais (ELMASRY, 2008).

Figura 7 – Comparação entre dados obtidos por sensores multiespectrais e hiperespectrais.

Fonte: SCHOWENGERDT (2006).

Na análise de imagens hiperespectrais, é importante realizar determinadas correções nos dados para aumentar a qualidade dos resultados. Técnicas de pré-processamento são bastante utilizadas para corrigir alguns problemas típicos, não só relacionados com as variações de medidas de imagens, mas também com as variações comuns observadas em um conjunto clássico de espectros. Ferramentas usuais como normalização, suavização e derivadas podem ser empregadas no experimento (SILVA, 2013). Há também outras ferramentas mais específicas para corrigir pixels anômalos, outliers e realizar compressões de imagens e que podem contribuir para o aumento da qualidade dos resultados (DE JUAN, 2009).

Para a redução do conjunto de dados sem perda significativa da informação, existem, por exemplo, técnicas de compressão como a PCA (do inglês - *Principal Components Analysis*), seleção de características baseada em Wavelets, etc, (PRATS-MONTALBÁN et al., 2011, SILVA 2013). Todavia, existem também duas formas simples de se realizar a análise de uma imagem hiperespectral a partir de uma redução significativa de dados, sem submetê-la a técnicas de compressão. A primeira é construir uma imagem contendo apenas as informações de absorbância ou transmitância de um único comprimento de onda

que deve estar diretamente associado ao composto que se deseja analisar. Já a segunda maneira, consiste na construção de um mapa de distribuição global da imagem, onde para cada pixel teremos um valor numérico que corresponde à soma de todos os valores de absorbância ou transmitância em todos os comprimentos de onda. Porém, é importante ressaltar que a utilização de qualquer uma das duas formas descritas implica na perda da riqueza de informação fornecida pela imagem hiperespectral (DE JUAN, 2009; SILVA, 2013).

Em geral, após submeter os dados às técnicas adequadas de pré-processamento, é possível melhorar a análise utilizando ferramentas para aperfeiçoar a resolução da imagem. No caso de imagens hiperespectrais, a resolução está associada à capacidade de identificar a informação química para cada pixel. Ou seja, os métodos de resolução procuram as contribuições individuais de cada componente de um sistema de forma que a combinação delas representem a resposta instrumental do sistema (DE JUAN, 2003; SILVA 2013).

1.2 Limiarização de Imagens

A limiarização de imagens é uma ferramenta de segmentação muito simples e computacionalmente rápida. É indicada para aplicações que exijam a identificação e extração do objeto da imagem, cujo tempo computacional consumido seja um fator relevante. As técnicas de segmentação estão divididas em dois conceitos distintos: similaridade (agrupamento de regiões homogêneas) ou descontinuidades (detecção de bordas). A limiarização é uma forma de agrupamento de regiões homogêneas (GONZALEZ et al., 2000).

Outros pesquisadores procuram classificar as técnicas de limiarização observando quais informações são usadas na escolha dos extremos e o número de limiares utilizados. Segundo a classificação de Weska (1978), os métodos de limiarização possuem três grupos: global, quando é determinado usando apenas as intensidades; local, se for usada as intensidades e alguma informação espacial da imagem; e dinâmico, se usar ambas as informações mais as coordenadas espaciais. Zhang (1996) classifica os algoritmos de segmentação de imagens também em duas classes distintas: métodos analíticos (baseados na análise de princípios, requisitos, complexidade e propriedades do algoritmo) e métodos empíricos (caracterizados apenas pela qualidade de seus resultados). Sahoo et. al. (1988), por sua vez, divide a técnica em dois grandes grupos: globais e locais. Para os autores, os limiares globais são aqueles que dividem toda a imagem usando apenas um limiar. Já os locais dividem a imagem em subimagens e para cada subimagem é atribuído um limiar. Esta técnica também é conhecida como limiar multiníveis. Esta classificação usa como critério a quantidade de limiares selecionados para a imagem. Segundo Gonzalez (2000), a limiarização multiníveis é geralmente mais trabalhosa que o limiar global, pela dificuldade em encontrar múltiplos limiares que determinem efetivamente as regiões de interesse, especial-

mente quando houver vários grupos de objetos na imagem. Por outro lado, a limiarização global alcança bons resultados apenas quando a iluminação da imagem é relativamente uniforme e as regiões de interesse, representadas pelos objetos, possuam uma diferença significativa de intensidade do seu fundo (contraste). Muitos métodos de limiarização têm sido propostos, além de variações destes métodos.

1.2.1 Método de Otsu

A principal característica do método de Otsu é a maximização da variância entre as classes da imagem, isto é, a maximização da separação entre o objeto e o fundo da imagem. A operação de limiar é vista como o particionamento dos pixels de uma imagem em duas classes. Uma das vantagens deste método é que ele não se restringe ao tipo de histograma da imagem. Isto é, pode ser aplicado em histogramas unimodais, bimodais ou multimodais, mas possui melhor desempenho em imagens com maior variância de intensidade (OTSU, 1978). Considerando as medidas de desempenho uniformidade e forma, este método apresenta bons resultados para imagens com histogramas bimodais e unimodais.

O Otsu é classificado como um método orientado à uniformidade do objeto. Sua principal desvantagem, está no desempenho sensível à presença de ruídos na imagem. Por outro lado, o Otsu é uma das melhores escolhas para aplicações gerais em tempo real em Visão Computacional. Sua implementação é sequencial e determina um limiar global (SAHOO et. al., 1988; LEE et. al., 1990). O conceito proposto é de iterar por todos os valores possíveis para a extremidade em uma imagem, buscando aquele que minimiza a soma da variância intraclasse da imagem. Esse valor irá corresponder ao melhor limiar para o caso, separando frente e fundo e atribuindo uma cor para cada classe. A variância intraclasse para um limiar possível t pode ser calculada com a Equação 1.1:

$$\sigma_w^2 = w_b \sigma_b^2 + w_f \sigma_f^2 \quad (1.1)$$

Onde:

w é o peso para cada classe. Essa medida corresponde à probabilidade que um pixel possui de pertencer a classe b (*background*) ou f (*foreground*).

1.3 Morfologia Matemática

De modo geral, podemos classificar as operações morfológicas em três grandes grupos: filtragem, segmentação ou medidas de imagens. Todas essas operações são construídas através de dois operadores fundamentais da morfologia: a erosão e a dilatação. Esses dois

operadores podem ser considerados como essenciais na morfologia, já que todos os outros operadores podem ser expressos em termos de erosão e dilatação (NEVES et al., 2013). A Figura 8 mostra os efeitos de dois algoritmos (além do já explicado Otsu) que foram utilizados neste projeto e que serão explicados nas subseções seguintes, sendo: Erosão e Dilatação, ambos aplicados na Figura 8a, após sua conversão para imagem binária (Figura 8b).

Figura 8 – a) Imagem original; b) Imagem (a) binarizada e com posterior aplicação da operação de Erosão; c) Imagem (a) binarizada com posterior aplicação da operação de Dilatação e; d) Imagem (a) ajustada com aplicação do Método de Otsu.

Fonte: próprio autor.

1.3.1 Erosão

Uma das primeiras perguntas que podem ser feitas quando se compara um elemento estruturante e um conjunto ou objeto é se o elemento estruturante está totalmente dentro do objeto, uma vez que o objeto erodido é o lugar geométrico dos centros dos elementos estruturantes onde esta resposta é afirmativa (NEVES, 2003). Isso qualifica a erosão binária com os seguintes efeitos: a) diminuir os objetos; b) eliminar objetos menores que o elemento estruturante; c) aumentar os buracos; e d) permitir a separação de objetos próximos (FACON, 1996).

A erosão pode ser representada segundo a Equação 1.2 abaixo, onde considerando x como o conjunto a ser erodido e $\beta(x)$ como o objeto estruturante escolhido, a erosão (ε) pode ser definida como o lugar geométrico dos pontos x nos quais β é incluído em x , quando sua origem está posicionada em x (SOILLE, 1999).

$$\varepsilon_{\beta}(x) = \{x | \beta_x \subseteq x\} \quad (1.2)$$

Segundo Facon (1996), os efeitos da erosão em tons de cinza são: a) escurecer a imagem; b) alargar e engordar os vales (padrões escuros); c) conectar vales próximos; d) reduzir e às vezes eliminar picos (padrões claros); e e) separar picos próximos. Um

elemento estruturante volumétrico (β_v), este tem um valor de cinza associado a cada ponto de seu domínio de definição, e pode-se definir a erosão para esse caso como (SOILLER, 1999):

$$\frac{[\varepsilon_{\beta_v}(f)](x)}{b \in \beta_v} = \min \{f(x + b) - \beta_v(b)\} \quad (1.3)$$

Os valores de cinza de um elemento estruturante volumétrico devem ser escolhidos de acordo com os valores de intensidade da imagem. Devido à dificuldade de se estabelecer essa condição de antemão, e ao aumento das operações computacionais por conta da subtração, o elemento estruturante volumétrico é raramente utilizado na prática (NEVES et al., 2013).

1.3.2 Dilatação

A dilatação é o operador dual da erosão (NEVES et al., 2013). Segundo Facon, (1996), os efeitos da dilatação binária são: a) engordar o objeto; b) preencher pequenos buracos; c) conectar objetos próximos. Considerando x como o conjunto a ser dilatado, β o elemento estruturante escolhido, a dilatação (∂) pode ser definida como o local dos pontos x onde β toca x quando sua origem coincide com x (SOILLE, 1999):

$$\partial_{\beta}(x) = \{x\beta_x \cap x \neq \emptyset\} \quad (1.4)$$

Já os efeitos da dilatação em níveis de cinza são: a) clarear a imagem; b) alargar e engordar os picos (padrões claros); c) conectar picos próximos; d) reduzir e as vezes eliminar vales (padrões escuros); e e) separar vales próximos (FACON, 1996). A dilatação por um elemento estruturante não plano é definida por (SOILLE, 1999):

$$\frac{[\partial_{\beta_v}(f)](x)}{b \in \beta_v} = \max \{f(x + b) + \beta_v(b)\} \quad (1.5)$$

Segundo Neves et al. (2013), diferente de imagens erodidas ou dilatadas por elementos estruturais planos, as imagens dilatadas ou erodidas por elementos estruturantes volumétricos não estão limitadas pelos valores mínimos ou máximos dos tons de cinza da imagem de entrada, pois a erosão com elemento estruturante em tons de cinza pode levar a imagens com resultados negativos, devendo ser usados moderadamente.

1.4 Reconhecimento de Padrões

O Reconhecimento de Padrões é a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em um número de categorias ou classes (THEODORIDIS, 1999). Essa área trabalha com os seguintes conceitos: a) Padrão: entidade, objeto, processo ou evento que pode assumir um nome; b) Classificação: atribuir classes para as amostras, baseado em suas características; c) Classe: conjunto de padrões que possuem características semelhantes; d) Classificadores: ferramentas utilizadas para classificar ou descrever padrões ou objetos a partir de um conjunto de propriedades ou características; e) Característica ou Atributo: dado extraído de uma amostra por meio de medida e/ou processamento; e) Ruído: distorção, falha ou diferenciação que ocorre na obtenção dos dados. Uma das formas de se realizar o Reconhecimento de Padrões é por meio dos algoritmos de Aprendizado de Máquina.

Há duas maneiras de se projetar um classificador (CONNEL et al., 2001), sendo: i) Classificação Supervisionada: o padrão de entrada é identificado como um membro de uma classe pré-definida, ou ii) Classificação Não Supervisionada: são selecionadas amostras representativas para cada uma das classes que se deseja classificar. Nela entende-se que já conhecemos o padrão e classes que estamos procurando. Já na Classificação Não Supervisionada, o conjunto de dados é particionado em grupos, baseados em características específicas, tais que os pontos dentro de um grupo (*cluster*) sejam mais similares entre si do que com os pontos de outros grupos.

O padrão a ser reconhecido é comparado com os modelos armazenados. A medida de similaridade é frequentemente uma correlação ou uma função de distância. Muitas vezes o modelo por si mesmo aprende a partir do conjunto de treinamento. Este modelo é computacionalmente caro, mas a atual disponibilidade de recursos computacionais permite que essas abordagens sejam mais facilmente viabilizadas (JAIN et al., 2000). O Reconhecimento de Padrões tem inúmeras aplicações (JIANG et al., 2004; LOPES, 2012). A Tabela 1 mostra exemplos de domínios do problema com suas respectivas classes de padrões. Com o avanço e a disponibilidade de vários recursos computacionais, tornou-se viável o projeto e a utilização de elaborados métodos de análise e classificação de padrões. Em muitas aplicações, não existe somente uma única abordagem para que a classificação seja ótima e, por isso, a combinação de várias abordagens de classificadores é uma prática bastante usada.

O projeto de um sistema de Reconhecimento de Padrões envolve, essencialmente, três etapas: 1) aquisição de dados (extração de características) e pré-processamento (seleção das características mais discriminativas); 2) representação de dados e; 3) tomada de decisão (construção de um classificador). A escolha de técnicas de pré-processamento, esquema de representação e método para a tomada de decisão, depende do domínio do

Tabela 1 – Aplicações para Reconhecimento de Padrões

Domínio do Problema	Aplicação	Padrão de Entrada	Classes e Padrões
Bioinformática	Análise de sequência	DNA/Sequência de Proteína	Tipos de conhecidos de genes/padrões
Mineração de dados	Busca por padrões significantes	Pontos em um espaço multi-dimensional	Compactar e separar grupos
Classificação de documentos	Busca na Internet	Documento texto	Categorias semânticas (negócios, entre outros)
Análise de documento de imagem	Máquina de leitura para cegos	Imagem de documento	Caracteres alfanuméricos, palavras
Automação industrial	Inspeção de placas de circuito impresso	Intensidade ou alcance de imagem	Natureza do produto (defeituosa ou não)
Recuperação de base de dados multimídia	Busca Internet	Vídeo clip	Gêneros de vídeo (p.e. ação, diálogo, entre outros)
Reconhecimento biométrico	Identificação pessoal	Face, íris, impressão digital	Usuários autorizados para controle de acesso
Sensoriamento remoto	Prognóstico da produção de colheita	Imagem multi-espectral	Categorias de aproveitamento de terra, desenvolvimento de padrões de colheita
Reconhecimento de voz	Inquérito por telefone sem assistência de operador	Voz em forma de onda	Palavras faladas

Fonte: LOPES, 2012

problema. Um problema bem definido e suficientemente detalhado, onde se tem pequenas variações intra-classes e grandes variações inter-classes, produzirá representações compactas de padrões e, conseqüentemente a estratégia de tomada de decisão será simplificada. A escolha de uma abordagem para o Reconhecimento de Padrões não é uma tarefa simples e muitas vezes ela conta com a experiência do projetista. Vale observar que elas não são necessariamente independentes, levando a diversas possibilidades para o projeto de sistemas híbridos (FU, 1983). Além disso, na literatura de Reconhecimento de Padrões, às vezes a mesma abordagem possui diferentes interpretações.

1.5 Aprendizado de Máquina

O Aprendizado de Máquina é definido de diferentes formas na literatura. Para Dehaven (2017), ele pode ser definido como um campo de estudo da Inteligência Artificial, relacionado com o desenvolvimento de sistemas computacionais capazes de fazer previsões e tomar decisões a partir de um conjunto de dados. Outras definições, como o caso de Ávila (2013), apresentam o conceito de Aprendizado de Máquina como sendo o desenvolvimento de algoritmos computacionais capazes de aprender e fazer previsões com base num conjunto de dados. Dentre as abordagens de Aprendizado de Máquina com melhores resultados no processamento de informação imagética, o autor Ávila (2013) aponta: *k-Nearest Neighbor Classification*, *Support Vector Machines*, o modelo *Bag-of-Visual-Words* e as RNA com arquitetura de Aprendizagem Profunda. Já para Calsaverini (2016), o Aprendizado de Máquina surge da intersecção da Ciência da Computação com a Estatística, na medida em que pretende construir máquinas que resolvam problemas e extrair pressuposições confiáveis de dados. Porém, embora todas as definições apresentadas auxiliem ao entendimento do conceito, elas levantam uma questão fundamental: o que se pode entender por uma máquina capaz de aprender? O autor Calsaverini (2016), expõe a seguinte resposta:

(...) dizemos que uma máquina aprende em relação a uma determinada tarefa T , métrica de desempenho P e tipo de experiência E , se o sistema confiável melhora seu desempenho P na tarefa T , seguinte E . Dependendo de como especificamos T , P e E de experiência, a tarefa de aprendizagem também pode ser chamada por nomes como mineração de dados, descoberta autônoma, banco de dados de atualização, programação, por exemplo, etc. (p. 1. Tradução livre do autor).

A principal distinção entre os processos de aprendizagem usados em Aprendizado de Máquina diz respeito à Aprendizagem Supervisionada ou a Aprendizagem Não Supervisionada (ÁVILA, 2013). Exemplos típicos da Aprendizagem Supervisionada são a classificação e a regressão, quando o alvo a prever é, respectivamente, uma categoria ou um valor. Para Calsaverini (2016), uma tarefa típica da Aprendizagem Não Supervisionada é o *clustering*, consistindo no agrupamento de pontos de um conjunto de dados em “grupos naturais” ou proceder ao “agrupamento hierárquico” desses mesmos pontos, como por exemplo o *Bag-of-Visual-Words* que usa um algoritmo de Aprendizagem Não Supervisionada na obtenção do “dicionário visual” para o seu funcionamento. Já nas áreas de Processamento de Imagens Digitais, atualmente as RNA com arquitetura de Aprendizagem Profunda vem sendo o sistema com o melhor desempenho em reconhecimento de imagens (CALSAVERINI, 2016; BEZERRA, 2016; DENG et al., 2014; JOST, 2016; YANG et al., 2017).

1.6 Redes Neurais Artificiais

As RNAs são técnicas de Inteligência Artificial inspiradas na estrutura do cérebro humano, e que adquirem conhecimento através do treinamento, chamado de aprendizado (KRUG, 2008). Normalmente, uma RNA é composta por várias unidades de processamento (neurônios), que são conectadas por canais de comunicação e associadas a um peso. Essas unidades de processamento trabalham apenas com as informações recebidas por suas conexões (BENZECRY, 2005). As RNAs são aplicadas no desenvolvimento de soluções de sistemas complexos, tais como datilógrafo fonético, reconhecimento de imagens, de voz, de caracteres, de faces, além de outras aplicações (BRAGA et al., 2007; SOUSA, 2002; CASTELANO, 2006).

As RNAs têm inspiração mais especificamente nos neurônios biológicos (Figura 9), que são formados por um corpo celular chamado Soma; possuem ramificações conhecidas como Dendritos que transmitem os sinais das extremidades até o corpo celular (terminais de entrada), e os Axônios que transmitem sinais do corpo celular até as extremidades (terminais de saída). A comunicação entre os neurônios é dada através das Sinapses, que é a região de contato entre os neurônios, assim transmitindo impulsos nervosos entre eles quando atingido um limiar de ação (BARRETO, 2002).

Figura 9 – Neurônio Biológico.

Fonte: Disponível em <<https://sites.icmc.usp.br/andre/research/neural/>>.

O Teorema da Aproximação Universal proposto por Hornik (1991), garante que RNAs com uma única camada oculta são aproximadores universais e são estruturas capazes de aproximar qualquer função contínua, contanto que se possa definir a quantidade suficiente de neurônios nessa única camada intermediária. Uma das motivações para a investigação de RNAs compostas de muitas camadas veio da neurociência, mais especificamente do estudo da parte do cérebro denominada córtex visual (DICARLO et al., 2012). Sabe-se que quando algum estímulo visual chega à retina, o sinal correspondente percorre

uma sequência de regiões do cérebro, e que cada uma dessas regiões é responsável por identificar características específicas na imagem correspondente ao estímulo visual. Em particular, neurônios das regiões iniciais são responsáveis por detectar formas geométricas simples na imagem, tais como cantos e bordas, enquanto neurônios nas regiões finais têm a atribuição de detectar formas gráficas mais complexas, compostas das formas gráficas simples detectadas por regiões anteriores. Dessa forma, cada região de neurônios (que é análoga ao conceito de camada em RNAs) combina padrões detectados pela região imediatamente anterior para formar características mais complexas. Esse processo se repete através das regiões até que os neurônios da região final têm a atribuição de detectar características de mais alto nível de abstração, tais como faces ou objetos específicos (BENGIO et al., 2013).

Para configuração e elaboração de uma RNA, devem ser especificadas as características dos neurônios, a topologia e as regras de treinamento. A adaptação dos pesos iniciais e a aprendizagem do seu comportamento são especificadas pelas regras de treinamento. Os algoritmos de treinamento (ou aprendizado) de uma RNA têm a característica de ajustar de forma iterativa os pesos das conexões entre os neurônios até que os pares de entradas e saídas esperados, no caso de Aprendizado Supervisionado, por exemplo, sejam obtidos e as relações matemáticas de causa e efeito possam ser estabelecidas. Quando a configuração de um determinado problema apresentado a RNA muda, e o modelo do desempenho não é mais adequado à situação, então é possível submeter a RNA a mais treinamentos com novas condições de entrada e saída para a melhora do seu desempenho (KRÖSE et al., 1996; HAYKIN, 1999). A Figura 10 apresenta o modelo geral de um neurônio artificial e nela é possível identificar três elementos básicos.

Figura 10 – Neurônio Artificial.

Fonte: próprio autor.

O primeiro modelo artificial de um neurônio biológico foi proposto no trabalho pioneiro de Mcculloch et al. (1943). O modelo de um neurônio artificial proposto por ele é uma simplificação do que se sabia na época, a respeito do neurônio biológico. Como já citado resumidamente, no neurônio artificial, que é baseado no neurônio biológico, trocam-se os Dendritos por Entradas, cuja ligação com o corpo celular artificial é dada por elementos chamados de Peso que simulam as Sinapses. Os sinais captados pelas entradas

são processados pela função Soma, e o limiar de ação do neurônio foi substituído pela transferência. O primeiro elemento é um conjunto de sinapses ou elos de conexão, cada qual caracterizada por um peso w_{kj} , com $k = 1$ e $j = 1, 2, \dots, m$. O segundo elemento diz respeito a um somatório para os sinais de entrada, ponderado pelas sinapses do neurônio (combinação linear), tal que:

$$v_k = \sum_{j=0}^m w_{kj} \cdot x_j \quad (1.6)$$

O terceiro elemento é uma função de ativação, que restringe a amplitude de saída de um neurônio, limitando o intervalo de saída para um valor finito.

$$y_k = \varphi(v_k) \quad (1.7)$$

Um neurônio k pode ser descrito matematicamente através das Equações 1.6 e 1.7, onde φ representa a função de ativação do neurônio artificial. Esta função existe para buscar um ajuste ótimo ou modelo para um problema. Após o trabalho precursor de Mcculloch, foram propostos outros modelos, permitindo saídas não necessariamente iguais a 0 ou 1 e com diferentes funções de ativação. A Figura 11 apresenta três tipos de funções de ativação. Em (a) a função logística sigmoide (*logsig*); (b) a função tangente hiperbólica (*tansig*); e em (c) a função linear de ativação (*purelin*).

Figura 11 – Funções de Ativação

A função logística (Equação 1.8) sigmoide pode assumir valores entre 0 e 1, onde a é o parâmetro de inclinação da função sigmoide e v é o valor de ativação do neurônio:

$$f(v) = \frac{1}{1 + e^{(-av)}} \quad (1.8)$$

A Equação 1.9 representa a função tangente hiperbólica assumindo valores entre 1 e -1, onde a é o parâmetro de inclinação da curva, b são os limites inferiores e superiores

e v é o valor de ativação, ou seja,

$$f(v) = a \frac{e^{(bv)} - e^{(-bv)}}{e^{(bv)} + e^{(-bv)}} \quad (1.9)$$

A função de ativação linear é definida pela Equação 1.10 onde a é um número real que define a saída linear para os valores de entrada, y é a saída e x é a entrada. Assim,

$$y = ax \quad (1.10)$$

Por muito tempo foi um consenso na comunidade de RNAs usar funções sigmóides para implementar a ativação de unidades ocultas, de modo a induzir não linearidades (BEZERRA, 2016). Entretanto, esse consenso foi questionado por Nair et al. (2010), que propuseram o uso de uma alternativa mais simples, a função de ativação retificada linear (Equação 1.11). Neurônios que usam essa função de ativação são chamados de Unidades Retificadoras Lineares ou *Rectified Linear Units* (ou simplesmente ReLU).

$$g(z) = \max(0, z) \quad (1.11)$$

1.6.1 Perceptron multicamadas

As redes Perceptron Multicamadas (MLP – do inglês *Multilayer Perceptron*) apresentam um poder computacional muito maior do que as redes sem camadas intermediárias, pois possuem a capacidade de tratar dados que não são linearmente separáveis (HAYKIN, 1999). Um dos principais aspectos deste tipo de rede está relacionado com a função de ativação utilizada (BRAGA et al., 2000). A MLP consiste em: um conjunto de unidades sensoriais que compõe a camada de entrada; uma ou mais camadas ocultas de neurônios computacionais; e uma camada de saída de neurônios computacionais. Na Figura 12 é apresentada uma RNA com uma camada escondida. Por simplicidade, esta arquitetura é referida como uma rede 3 – 4 – 2, isto é, 3 neurônios de entrada, 4 neurônios escondidos e 2 neurônios de saída (YIODA, 2000).

1.6.2 Back-propagation

O algoritmo mais popular das RNAs é o de Retropropagação do Erro (*Error Back-propagation*). A maioria dos métodos de aprendizado para RNAs do tipo MLP utilizam variações deste algoritmo (BRAGA et al., 2000). Sua aprendizagem consiste em dois

Figura 12 – Rede com uma camada de entrada, uma camada escondida e uma camada de saída.

Fonte: Adaptado de YIODA (2000).

passos através das diferentes camadas da rede, conforme a Figura 13: um passo à frente (*forward*), a propagação, e um passo para trás, a retropropagação (*backward*).

Figura 13 – Fluxo de Processamento do Algoritmo de Retropropagação do Erro.

Fonte: Adaptado de DENG et al. (2014).

No passo para frente (*forward*) encontram-se os seguintes sub-passos: (1) a entrada é apresentada à primeira camada da rede, a camada C_0 ; (2) para a camada C_i a partir da camada de entrada, após os neurônios da camada C_i ($i > 0$) calcularem seus sinais de saída, estes servem como entrada para a definição das saídas produzidas pelos neurônios da camada C_{i+1} ; (3) as saídas produzidas pelos nodos da última camada são comparadas às saídas desejadas. De acordo com Braga et al. (2000), o passo para trás (*backward*) envolve as seguintes etapas a partir da última camada, até chegar na camada de entrada: (1) Os neurônios da camada atual ajustam seus pesos de forma a reduzir seus erros nas respostas desejadas dos padrões e; (2) o erro de um neurônio das camadas intermediárias é calculado utilizando os erros dos neurônios da camada seguinte conectados a ele, ponderados pelos pesos das conexões entre eles. Ainda segundo o autor, além das redes já citadas, existe também as Redes Neurais Recorrentes, explicadas mais a frente.

O algoritmo de Retropropagação do Erro é baseado na regra delta proposto por Windrow et al. (1960), sendo por isto também chamada de regra delta generalizado. Este algoritmo propõe uma forma de definir o erro dos nodos das camadas intermediárias, possibilitando o ajuste de seus pesos. Os ajustes dos pesos são realizados utilizando-se o

Método do Gradiente⁴. A derivação da regra delta generalizada é semelhante à derivação da regra delta. A função custo a ser minimizada é uma função de erro, definida pela soma dos erros quadráticos e representada por

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (1.12)$$

onde: E é a medida de erro total, p é o número de padrões, k é o número de unidades de saída, d_i é a i -ésima saída desejada e y_i é a i -ésima saída gerada pela rede. Esta equação define o erro total cometido pela rede, ou a quantidade em que, para todos os padrões p de um dado conjunto, as saídas geradas pela rede diferem das saídas desejadas. A regra delta generalizada requer que as funções de ativação utilizadas pelos neurônios sejam contínuas, diferenciáveis e, geralmente, não-decrescentes da entrada total recebida pelo nodo. A Equação 1.13 ilustra o cálculo do valor de ativação, tal que:

$$y_j^p = f_j(v_j^p) \quad (1.13)$$

onde:

$$v_j^p = \sum_{i=1}^n x_i^p w_{ji} \quad (1.14)$$

Na Equação 1.14, a constante n representa o número de conexões de entrada do neurônio j para camada seguinte e w_{ji} , os pesos entre estas conexões.

Todavia, o algoritmo de Retropropagação do Erro apresenta uma série de dificuldades ou deficiências que desestimula uma maior disseminação do seu uso. O principal problema diz respeito à lentidão do algoritmo para superfícies mais complexas. Não é raro o algoritmo convergir para mínimos locais. Os mínimos locais são pontos na superfície de erro que apresentam uma solução estável, embora não sejam a saída correta. Desde que o algoritmo de Retropropagação do Erro foi popularizado, algumas pesquisas vêm buscando técnicas tanto para acelerar o algoritmo quanto para reduzir a incidência de mínimos locais. Dentro destas técnicas destaca-se o algoritmo de Levenberg-Marquardt. O algoritmo é eficiente quando se trata de redes que não possuem mais do que algumas centenas de conexões a serem ajustadas, uma vez que ele trabalha de forma a evitar o mal-condicionamento de inversão de matrizes. (HAGAN, 1994).

⁴ O Método do Gradiente é um método numérico usado em otimização para encontrar um mínimo local de uma função, onde em cada passo se toma a direção negativa ao gradiente. Disponível em: <<https://sites.icmc.usp.br/andretta/ensino/aulas/sme0100-2-12/aula5-gradientes.pdf>>

1.6.3 Aprendizagem Profunda

O conceito de aprendizagem profunda pode ser empregado em modelos diferentes de RNAs. É o termo usado para denotar o problema de treinar RNAs que realizam o aprendizado de características de forma hierárquica, de tal forma que características nos níveis mais altos da hierarquia sejam formadas pela combinação de características de mais baixo nível (JÜRGEN, 2015; DENG et al., 2014; AREL et al., 2010; GOODFELLOW et al., 2016; LECUN et al., 2015; BENGIO, 2009). Muitos problemas práticos têm sido resolvidos por RNAs compostas de uma única camada intermediária (HAYKIN, 2007). As RNAs não são uma novidade no campo do Aprendizado de Máquina, sendo o Perceptron, criado em 1957 por Rosenblatt, considerado como o primeiro passo rumo às atuais RNAs (DEHAVEN, 2017).

Contudo, o desenvolvimento deste tipo de rede ficou comprometido pelas capacidades computacionais da época, nomeadamente no que concerne à utilização de uma arquitetura de Aprendizagem Profunda, uma vez que a mesma implica múltiplas camadas de unidades de processamento, no caso, os já citados neurônios artificiais (BEZERRA, 2016; CALSAVERINI, 2016; JOST, 2016). O renovado interesse ocorrido a partir de meados da década de 2000 nas RNAs com arquitetura de Aprendizagem Profunda, foi possível devido a uma combinação de vários fatores (DENG et al., 2014). Os três mais determinantes foram: a) os avanços significativos no desenvolvimento de *hardware*, especialmente nas unidades de processamento gráfico; b) o desenvolvimento de algoritmos e técnicas mais eficientes para treinar RNAs; e c) a disponibilização crescente de grandes conjuntos de dados, nomeadamente bancos de imagens rotulados como *ImageNet* ou *Places* (BEZERRA, 2016; CALSAVERINI, 2016; DENG, 2014). A razão para que a Aprendizagem Profunda apresente melhor desempenho tanto na classificação de imagens como no reconhecimento de objetos, reside no fato de tornarem possível extrair automaticamente e com elevado grau de confiabilidade, informação visual de médio e alto nível, tal como a detecção de limites de objetos que se intersectam e partes de um objeto (DENG, 2014).

Há ainda outra motivação, dessa vez teórica, para o estudo da Aprendizagem Profunda. Em Larochelle et al. (2009), usando funções lógicas (*boolean functions*) com base em estudos anteriores, os autores demonstram que é possível representar qualquer função lógica por meio de um circuito lógico de uma única camada, embora essa representação possa requerer uma quantidade de unidades na camada intermediária que cresce exponencialmente com o tamanho da entrada. Eles também demonstraram que, ao permitir que o circuito cresça em quantidade de camadas intermediárias, essas mesmas funções lógicas problemáticas poderiam ser representadas por uma quantidade de unidades ocultas que cresce polinomialmente com o tamanho da entrada. Há três classes de RNAs muito populares em Aprendizagem Profunda: as Redes Autocodificadoras, um tipo de RNA usada

para aprender codificações de dados eficientes de maneira Não Supervisionada; as Redes Neurais Recorrentes, que - apesar de não precisarem ser classificadas necessariamente como redes profundas - são um outro tipo de RNA projetada para reconhecer padrões em sequências de dados e; as Redes Neurais Convolucionais, explicadas com mais detalhes nas subseções a frente.

As Redes Autocodificadoras (*Autoencoders*) pertencem a classe de RNAs que são treinadas de forma Não Supervisionada para aprender características a partir de um conjunto de dados. Elas são treinadas não para prever alguma classe, mas sim para identificar na sua saída a própria entrada (BEZERRA, 2016). Estas características são úteis para uso posterior em tarefas de Aprendizado Supervisionado e outras relacionadas à Visão Computacional. Esse tipo de rede profunda permite o aprendizado de representações mais claro de um conjunto de dados, sendo muito úteis no contexto do aprendizado de codificações mais compactas de um conjunto de dados (SALAKHUTDINOV et al., 2006). Além dessas características, as *Autoencoders* podem ainda ser usadas para iniciar os pesos de uma rede orientada por Aprendizagem Profunda no uso posterior em uma tarefa de Aprendizado Supervisionado. Já as Redes Neurais Recorrentes (RNN - do inglês *Recurrent Neural Networks*) fazem parte de uma extensa classe de RNAs, em que o estado depende tanto da entrada quanto do seu estado atual. Esses algoritmos consideram tempo e sequência, constituindo assim uma dimensão temporal. Nas RNNs, a propriedade de possuir um estado dinâmico possibilita realizar análises computacionais, dependente do contexto, aprendendo a partir de dependências de longo prazo (BEZERRA, 2016). As RNNs têm sido aplicadas recentemente em uma variedade de problemas e são muito utilizadas em aplicações de: geração de textos, reconhecimento de fala, tradução automática, entre outros.

1.6.4 Redes Convolucionais

Como citado previamente, as RNCs se inspiram no funcionamento do córtex visual (LE-CUN et al., 1998; ZEILER et al., 2014). Elas se baseiam em alguns princípios básicos, como: os campos receptivos locais, compartilhamento de pesos, convolução e nas subamostras. Em uma RNC, sua camada é segmentada em diversos mapas de características (Figura 14), em que cada unidade de um mapa tem o objetivo realizar a mesma operação sobre a imagem de entrada, com cada unidade aplicando essa operação em uma região específica dessa imagem (BEZERRA, 2016).

A convolução acontece quando cada unidade em um mapa de característica realiza uma operação. No exemplo do trabalho de Bezerra (2016), considere que a imagem de entrada contém 28 x 28 pixels em escala de cinza. A Figura 15a apresenta o resultado da aplicação de um tipo particular de convolução sobre a imagem exemplo.

Figura 14 – Arquitetura típica de uma rede convolucional, onde a camada é segmentada em diversos mapas de características com o objetivo de realizar a mesma operação sobre a imagem de entrada em regiões específicas.

Fonte: Adaptado de LECUN (1998).

Figura 15 – Resultado de uma convolução (direita) aplicada a uma imagem (esquerda).

Fonte: Adaptado de BEZERRA (2016).

Outra operação importante utilizada em uma RNC é a subamostragem, que se refere a redução da resolução de uma imagem, sem alterar significativamente o seu aspecto. Na RNC, a subamostragem reduz a dimensionalidade de um mapa de característica fornecido como entrada e produz outro mapa de característica, constituindo uma espécie de resumo. Há várias formas de subamostragem aplicáveis a um mapa de característica como, por exemplo: o *Max Pooling* (seleção do valor máximo), *Average Pooling* (cálculo da média) ou a *L2-Pooling* (norma do conjunto) etc.

Na forma mais comum de arquitetar uma RNC, a rede é organizada em estágios. Cada estágio é composto por uma ou mais camadas de convolução em sequência, seguidas por uma camada de subamostragem, que por sua vez é seguida - de forma opcional - por uma camada de normalização. Uma RNC pode conter vários estágios empilhados após a camada de entrada, correspondente à imagem. Após o estágio final da rede, são adicionadas uma ou mais camadas completamente conectadas. A Figura 14 apresenta um exemplo esquemático de uma RNC na qual, após a camada de entrada (que corresponde aos pixels da imagem), temos uma camada de convolução composta de 6 mapas de características (representados como planos na figura), seguida de uma camada de subamostragem, completando o primeiro estágio. Essa figura ainda ilustra um segundo estágio antes das duas camadas completamente conectadas (BEZERRA, 2016).

Na subamostragem (*Subsampling Layer* ou *Pooling Layer*), o objetivo é realizar a

agregação das saídas (ou ativações) de um conjunto de unidades da camada anterior. É possível mostrar que camadas de subamostragem resultam em uma rede mais robusta a transformações espaciais. Segundo Jarrett et al. (2009), as RNCs também podem conter ainda as denominadas camadas de normalização de contraste local (LCN - do inglês *Local Contrast Normalization*), em que, quando utilizada, essa camada é posicionada na saída da camada de subamostragem. Uma camada LCN normaliza o contraste de uma imagem de forma não linear. Ao invés de realizar uma normalização global, a camada LCN aplica a normalização sobre regiões locais da imagem, considerando cada pixel por vez. Essa normalização pode corresponder a subtrair a média da vizinhança de um pixel particular e dividir pela variância dos valores de pixel dessa vizinhança. Esta transformação equipara a RNC com a invariância de brilho, propriedade útil no contexto de reconhecimento de imagens (BEZERRA, 2016).

Outro tipo de camada em uma RNC é a completamente conectada. Em arquiteturas de RNCs modernas, é comum encontrar uma ou duas camadas desse tipo antes da camada de saída. Juntamente com as camadas de convolução e subamostragem, as camadas totalmente conectadas geram descritores de características da imagem que podem ser mais facilmente classificados pela camada de saída, que normalmente usam a função *Softmax*. Outras formas de função de ativação, com a função *Softmax*, foram propostas para amenizar algumas dessas desvantagens sofridas pela função de ativação logística sigmoide, citada anteriormente (NWANKPA et al., 2018).

A função *Softmax* é outro tipo de função de ativação usada na computação neural. Essa função é utilizada para calcular a probabilidade de distribuição de um vetor de números reais. Ela é usada geralmente em modelos multi-classes onde retorna probabilidades de cada classe, com a classe alvo tendo a maior probabilidade. A função *Softmax* aparece principalmente nas camadas de saída das arquiteturas de Aprendizagem Profunda. A principal diferença entre função logística sigmoide e a *Softmax* é que a função sigmoide é usada na classificação binária enquanto o *Softmax* é usado para tarefas de classificação multivariada (NWANKPA et al., 2018).

A função *Softmax* produz uma saída que é um intervalo de valores entre 0 e 1, com a soma das probabilidades igual a 1. A função *Softmax* é calculada (Equação 1.15) usando a equação (NWANKPA et al., 2018):

$$f(U_j) = \frac{\exp(U_j)}{\sum_j \exp(U_j)} \quad (1.15)$$

Segundo Bezerra (2016), é importante notar que os pesos dos mapas de características em cada camada de convolução são aprendidos durante o treinamento. Por conta das operações de convolução e subamostragem serem diferenciáveis, uma RNC pode ser tam-

bém treinada por meio do algoritmo de Retropropagação do Erro, com a diferença de que os pesos são atualizados considerando a média dos gradientes dos pesos compartilhados. Durante esse treinamento, é comum aplicar a técnica de desligamento às camadas completamente conectadas para evitar o sobreajuste. Por meio das RNC, é possível realizar a detecção de padrões em imagens de forma mais adequada, no sentido de que características dos dados dessa natureza podem ser exploradas para obter melhores resultados. Desde a origem das RNCs, esse tipo de RNA tem se mostrado promissor em tarefas que envolvem Reconhecimento Visual (LECUN et al., 2015).

1.6.5 Treinamento em redes profundas

Mesmo sendo considerado o modelo de RNA mais vantajoso, no início dos anos 2000 o treinamento de RNAs com arquitetura de Aprendizagem Profunda era uma tarefa de difícil execução, devido a dois fatores.

O primeiro deles é a Dissipação dos Gradientes (HOCHREITER et al., 2001), causado pela saturação de unidades, uma vez que durante a aplicação do Algoritmo de Retropropagação, o gradiente da ativação de cada unidade oculta é calculado e utilizado como um termo multiplicativo para atualização dos pesos correspondentes. Entretanto se a unidade em questão está próxima da saturação, então a derivada parcial resultante será um valor próximo de zero. Isso faz com que o gradiente com relação às pré-ativações seja também próximo de zero, porque seu cálculo envolve multiplicar pela derivada parcial. Isso significa que o gradiente retropropagado para as camadas anteriores será cada vez mais próximo de zero, de tal forma que quanto mais próximo à camada de entrada, maior a quantidade de gradientes próximos de zero. Como resultado, à medida que o gradiente do erro é retropropagado seu valor decresce exponencialmente, de tal forma que o aprendizado nas camadas mais próximas à entrada se torna muito lento. Esse problema afeta não apenas o aprendizado em redes profundas, mas também em redes recorrentes.

O segundo complicador é a consequência do número maior de parâmetros para adequar, uma vez que os parâmetros de uma RNA são seus pesos, e quanto mais camadas, mais pesos. De fato, quanto maior a quantidade de parâmetros a ajustar em um modelo, maior o risco de sobreajuste. O aprendizado de modelos por meio do treinamento de redes profundas tem sido tradicionalmente tratado como um processo empírico. Soluções para os problemas mencionados acima envolvem o uso de diversos procedimentos que foram surgindo ao longo dos anos e que em conjuntos passaram a contar como fatores importantes no sucesso da RNA com arquitetura de Aprendizagem Profunda, além claro, do aumento do poder computacional e das tecnologias computacionais como um todo (BEZERRA, 2016).

1.6.6 Pré-treinamento não supervisionado

A proposta do Pré-treinamento Não Supervisionado é o de iniciar os pesos de uma rede por meio de uma abordagem Não Supervisionada (HINTON, 2007; BENGIO et al., 2007). Segundo Erhan (2010), esse procedimento permite posicionar os pesos da rede em uma região do espaço de parâmetros tal que seja mais fácil para métodos de otimização baseados em gradientes encontrarem um valor ótimo para o objetivo durante o treinamento supervisionado. Durante sua aplicação, camadas intermediárias são adicionadas à rede iterativamente, onde em cada iteração, uma nova camada intermediária é adicionada à rede. A nova camada intermediária adicionada deve capturar regularidades nos padrões de ativação das unidades contidas na camada anterior. Esse procedimento pode ser aplicado para a construção de redes com um número arbitrário de camadas intermediárias.

Após a fase de Pré-treinamento Não Supervisionado, a camada de saída é então adicionada à rede. Por fim, a rede completa é treinada da forma supervisionada usual. Essa segunda fase é denominada sintonia fina (*fine-tuning*), (BEZERRA, 2016). Uma vantagem do pré-treinamento é que é possível tirar proveito da existência de dados não rotulados, que em geral são mais comuns do que dados rotulados. Outra vantagem é que normalmente o aprendizado com pré-treinamento converge mais rapidamente. No trabalho que se propôs originalmente o uso de pré-treinamento (HINTON et al., 2006), os autores realizaram o treinamento não supervisionado por meio de máquinas restrita de Boltzmann⁵. Posteriormente, outros autores propuseram usar pré-treinamento por meio de Redes *Autoencoders* (RANZATO et al., 2006; BENGIO et al., 2007).

1.6.7 Dropout

O *Dropout* é outra técnica também aplicável ao treinamento de uma RNA para reduzir o risco de sobreajuste (HINTON et al., 2012; SRIVASTAVA et al., 2014). Na *forward propagation*, antes de um padrão ser apresentado à camada intermediária de índice k , cada entrada de um vetor binário $m(k)$ é gerado por meio de uma distribuição uniforme com parâmetro p ($p = 0,5$ é um valor usual). Esse vetor é então usado como uma máscara para desativar algumas das unidades da camada k . Mais especificamente, a pré-ativação da camada k ocorre de forma usual, mas a sua ativação é modificada ligeiramente: $h(x) = g(a(k)(x)) \cdot m(k)$. Efetivamente, essa modificação faz com que a ativação das unidades correspondentes a valores iguais a zero na máscara $m(k)$ seja também igual a zero (BALDI et al., 2013). A Figura 16 representa o efeito da aplicação do desligamento: a imagem da esquerda é a configuração original da rede e a da direita corresponde à configuração após

⁵ A máquina de Boltzmann restrita é uma rede estocástica amplamente utilizada para compor redes profundas. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2016/pdf/155985.pdf>>

o desligamento de algumas unidades.

Figura 16 – a) Rede sem aplicação do desligamento e; b) com efeito de desligamento.

Fonte: Adaptado de SRIVASTAVA et al. (2014)

A aplicação de desligamento requer a definição de um hiperparâmetro na probabilidade p que controla a intensidade de desligamento. Se $p = 1$ implica nenhum desligamento, e quanto menor o valor de p , maior a taxa de desligamento. Valores típicos de p para as unidades ocultas estão na faixa de 0,5 a 0,8 (SRIVASTAVA et al., 2014). De acordo com Konda et al., 2015, o procedimento de desligamento pode ser interpretado como uma forma adaptativa de acréscimo de dados (Dyk et al., 2001).

Estado da Arte

Aqui apresentamos alguns trabalhos que de algum modo se relacionam com os temas abordados nesta dissertação. Existe uma série de pesquisas relacionadas ao uso de RNAs para o diagnóstico de casos de câncer e outras doenças, inclusive em conjunto com imagens espectrais. O levantamento bibliográfico teve por objetivo avaliar o estado da arte em pesquisas relacionadas ao uso - preferencialmente em conjunto - das tecnologias de RNA com arquitetura de Aprendizagem Profunda, imagens hiperespectrais, Reconhecimento de Padrões e no pré-diagnóstico informatizado em casos de câncer do cólon. As pesquisas foram obtidas a partir da seguinte string de busca:

("Document Title": Hyperspectral AND (Imag OR Absorbance OR Spectrum OR Analys*) AND (Diagn* OR Colorect* OR Cancer OR Colon) AND (Neural Network OR Pattern Recognition OR Deep*))*

Após revisados e devidamente triados os artigos mais similares a esta pesquisa, foi possível uma análise prévia do estado arte. Há diferentes estudos sobre o uso de imagens espectrais para análise de tecidos cancerígenos e, boa parte, em conjunto com algum tipo de RNA. Todavia, a análise do estado da arte também possibilitou que outras técnicas computacionais - diretamente relacionadas com a nossa pesquisa - pudessem ser analisadas e aplicadas ao projeto. Um exemplo, é o estudo de Xiang et al. (2009). Nele foi analisado o espectro de imagens de câncer do endométrio. Nesta pesquisa, além de imagens espectrais – similares às utilizadas em nosso estudo - os autores utilizaram uma RNA de Propagação de Retorno (BPNN, do inglês *Back-Propagation Neural Networks*) e seus efeitos na classificação na imagem dos tecidos. O ponto forte de contribuição deste artigo foi o fato do modelo de RNA utilizado no projeto ter sido obtido pelo processo de otimização do número de neurônios na camada oculta, da taxa de aprendizagem e das épocas de aprendizagem por meio da técnica de *Tuning*. A técnica contribuiu diretamente para o nosso projeto, por proporcionar uma forma de selecionar os melhores parâmetros da RNA com arquitetura de Aprendizagem Profunda utilizada na aplicação. Pesquisando complementarmente o trabalho de Xiang, foi possível constatar a grande utilização de

Algoritmos Genéticos (AGs) no processo de *Tuning* para otimização de RNAs. Os AGs também são orientados seguindo os conceitos da biologia, tal qual os neurônios artificiais. De acordo com Murray (1994), o AG é a técnica de otimização mais adequada para solucionar problemas de otimização de RNAs. Apesar de não ocorrerem com frequência, os AGs são ainda capazes de encontrar uma solução ótima global de um problema de otimização em complexos espaços de busca multidimensionais (PHAM et al., 2000). Por fim, os resultados do estudo de Xiang, mostraram que a sua BPNN, utilizada em conjunto com imagens espectrais, pôde fornecer um método eficiente para o diagnóstico precoce do câncer endometrial.

Há também trabalhos semelhantes com o uso de RNAs em conjunto com tecnologias do infravermelho. Na pesquisa de Lijun et al. (2017), foi proposto um método de detecção em imagem infravermelha, combinando segmentação e RNAs, baseado na análise do processamento deste tipo de imagem. Já na pesquisa de Janssens et al. (2017), foi proposto um estudo sobre RNA com arquitetura de Aprendizagem Profunda em conjunto com Infravermelho Térmico (IRT). Nele, concluiu-se que usando este tipo de RNA, treinadas em regiões importantes das imagens de IRT, é possível identificar uma série de condições específicas e que podem potencialmente levar a novas compreensões deste tipo de artefato. Por fim, no estudo de Vanderbeek et al. (1994), é apresentado um esquema de detecção analógica para o processamento de sinal de alta velocidade de interferogramas infravermelhos remotos passivos. O sistema é composto por três etapas sucessivas, as quais são realizadas no domínio analógico. O primeiro estágio é uma RNA analógica que calcula a transformada de Fourier do interferograma. A segunda fase insere as partes reais e imaginárias da transformada de Fourier e calcula a sua magnitude. A fase final é uma RNA que usa o espectro de magnitude para determinar a presença ou ausência de hexafluoreto de enxofre nas imagens.

Outras pesquisas trazem exemplos de experimentos voltados à detecção e classificação de ROIs em imagens espectrais. No trabalho de Sammouda et al. (2015), foi desenvolvido uma aplicação protótipo para a detecção e classificação de tumores de próstata, utilizando espectros de infravermelho em imagens da próstata. Esta técnica de imagem óptica baseia-se em estimular os endógenos cromóforos ou aplicação de agentes de contraste, capazes de direcionar o câncer celular. No trabalho é proposto um método de segmentação de imagens utilizando o denominado Antígeno de Membrana Específico da Próstata (PSMA). Nele, uma RNA classifica os pixels em clusters distintos. Os resultados preliminares deste trabalho, demonstraram que a segmentação proposta pôde melhorar a prática clínica já existente durante o processo de identificação de áreas na imagem da próstata. No estudo de Nie et al. (1991), foi demonstrado a utilidade de algoritmos de Reconhecimento de Padrões em diagnósticos e classificação de tecidos doentes utilizando a espectroscopia Raman. A espectroscopia Raman é uma técnica fotônica de alta resolução que pode proporcionar, em poucos segundos, informações químicas e estruturais

de praticamente qualquer material, composto orgânico ou inorgânico permitindo assim sua identificação. No estudo, a espectroscopia Raman foi capaz de detectar com precisão tecidos doentes e saudáveis, a partir de suas impressões digitais.

Por fim, dois trabalhos trazem projetos mais próximos da nossa pesquisa, utilizando RNAs e imagens hiperespectrais para a detecção de anomalias em tecidos cancerígenos. Em Ortega et al. (2011), foi desenvolvido um método para diagnóstico de tumores do sistema nervoso a partir de imagens hiperespectrais de tecidos cerebrais humanos. As amostras biológicas utilizadas neste trabalho consistem em biópsias de tecido cerebral humano, ressecados durante a cirurgia. Uma vez que a biópsia é diagnosticada, o tecido é classificado de acordo com a classificação da *World Health Organization* (WHO) para tumores do sistema nervoso. Uma vez diagnosticado um determinado tecido, os patologistas definem as ROIs utilizando canetas coloridas. Tecidos tumorais são marcados com a cor vermelha, e tecidos saudáveis com a cor azul. Ainda sobre o projeto de Ortega, para demarcar cada espectro como saudáveis ou tumorais - de acordo com a identificação prévia do patologista - uma interface gráfica foi desenvolvida, permitindo a exibição de uma imagem RGB que possibilita ao usuário recortar uma ROI que contenha apenas tecidos diagnosticados. Depois que uma região da imagem for selecionada, uma janela de diálogo é mostrada e o usuário informa se a ROI é saudável (marcada como azul) ou com tumor (marcado como vermelho). Na sequência, os espectros dos pixels selecionados são extraídos do hipercubo e os rótulos patológico são adicionados ao conjunto de dados hiperespectrais. O banco de dados criado foi classificado utilizando um algoritmo de RNA com arquitetura de Aprendizagem Profunda, a fim de determinar se é possível distinguir entre tecido saudável e tecido tumoral usando apenas sua informação espectral. Na função de ativação foi selecionada a função sigmoide hiperbólica tangente na camada de saída. Este método serviu de inspiração na primeira versão da aplicação PAIH, mas que posteriormente foi substituída pela função de ativação multi-classes *Softmax*, explicada nos capítulos a frente. Por fim, os resultados do trabalho de Ortega et al., demonstraram que as classes (tumor e saudável) foram classificadas corretamente. Estes resultados foram obtidos sem o uso de uma cadeia de pré-processamento antes da classificação, que provavelmente aumentaria a qualidade e o desempenho dos classificadores. Foi demonstrado que o sistema de aquisição desenvolvido é capaz de registrar dados hiperespectrais contidos em lâminas biológicas preservadas. Resultados preliminares mostraram ainda que foi possível discriminar os tecidos entre saudável e com tumor, utilizando a informação espectral do tecido.

Por fim, no trabalho de Signoroni et al. (2017), foi projetada e treinada uma RNA com arquitetura de Aprendizagem Profunda para identificação de patógenos a partir de infecções urinárias, por meio de um banco de imagens hiperespectrais. A RNA com arquitetura de Aprendizagem Profunda foi avaliada e comparada com outras soluções convencionais de classificação. A proposta da aplicação é a de possibilitar a identifica-

ção de bactérias patogênicas crescendo sobre placas de cultura por meio de uma imagem hiperespectral. Em simultâneo foram utilizadas técnicas de compressão nas imagens hiperespectrais com o objetivo de minimizar a ocupação de dados, evitando a redução da precisão durante o processo de classificação. O primeiro foco deste trabalho está na arquitetura da RNA com arquitetura de Aprendizagem Profunda. A arquitetura proposta alcançou uma alta precisão de classificação quando aplicado em uma base de dados. O sistema obteve uma precisão de 99,7% após 50.000 iterações de treinamento. Assim, neste trabalho, foi provado que é possível empregar uma RNA com arquitetura de Aprendizagem Profunda para resolver o problema da classificação de bactérias patogênicas em imagens hiperespectrais.

Esta pesquisa bibliográfica mostrou que há um conjunto de pesquisas correlacionadas ao uso de RNAs, Reconhecimento de Padrões e imagens hiperespectrais no auxílio ao diagnóstico em casos de câncer. Todavia nenhuma dessas pesquisas tratou especificamente o projeto proposto nesta dissertação, que tem como proposta o uso de uma RNA com arquitetura de Aprendizagem Profunda em conjunto com imagens hiperespectrais para caracterização de tecido do cólon em três classes diferentes (câncer, inflamada ou saudável), visando o auxílio ao diagnóstico.

Metodologia

Os atuais exames histológicos de biópsia do cólon passam por vários processos, desde a coleta do tecido - por meio do processo de colonoscopia - até a análise da imagem H&E pelo patologista. O tempo do diagnóstico depende tanto da qualidade da coleta do material quanto da habilidade do profissional. Como demonstrado no capítulo 2 (Estado da Arte), há diversos trabalhos que propõem a otimização deste processo em vários tipos de exames histológicos, utilizando tecnologias como imagens hiperespectrais e RNAs. Porém nenhuma destas pesquisas tratou especificamente do uso de imagens hiperespectrais em conjunto com a tecnologia de RNAs com arquitetura de Aprendizagem Profunda para caracterização de tecidos do cólon em três classes diferentes (câncer, inflamada e saudável).

Este trabalho propõe o desenvolvimento de uma ferramenta que possa auxiliar no diagnóstico de câncer do cólon utilizando RNAs com arquitetura de Aprendizagem Profunda para analisar pixels em imagens hiperespectrais obtidas de lâminas com amostras de biópsia do tecido do cólon em parafina, categorizando-os automaticamente nas classes: câncer, inflamada ou saudável (classificação CIS). O objetivo da aplicação é o de auxiliar a formulação de diagnósticos mais confiáveis e precisos, por meio de uma ferramenta computacional gráfica de pré-diagnóstico, complementando os atuais exames histológicos. A proposta também tem como hipótese comprovar que as imagens hiperespectrais permitem identificar regiões de tecidos contendo indícios de câncer, inflamação ou tecido saudável a partir do espectro dos seus pixels. A metodologia utilizada para propor a comprovação do método depende da execução de vários processos nos módulos desenvolvidos. Assim, o sistema utilizado no experimento proposto é composto por cinco módulos, sendo:

- I. **Gerador do banco de ROIs** – sistema que efetua a montagem automática da base de dados que foi utilizada no módulo *Treinamento do Classificador* contendo somente as ROIs digitais (e suas respectivas 1626 frequências hiperespectrais) demarcadas anteriormente de forma manual pelo patologista. O sistema gera automaticamente - em um único arquivo - todas as ROIs oriundas de uma base de imagens hiperespectrais categorizadas na classificação CIS.

- II. **Otimização e Validação Cruzada** – a definição manual dos melhores hiperparâmetros para uma RNA com arquitetura de Aprendizagem Profunda pode ser um trabalho oneroso. Para que o módulo de *Treinamento do Classificador* fosse executado de forma mais precisa, nesta segunda etapa foi proposto a implementação de um sistema que seleciona automaticamente os melhores hiperparâmetros a partir de um recurso nativo do *Keras*: o atributo *best_params* da função *GridSearchCV*. Uma vez definido os melhores hiperparâmetros, é de suma importância validar a RNA com arquitetura de Aprendizagem Profunda utilizando diferentes frações do banco de ROIs por meio de um sistema autenticador que utilize como metodologia a Validação Cruzada.
- III. **Treinamento do Classificador** – após todas as etapas para geração do banco de ROIs, otimização e validação dos melhores hiperparâmetros da RNA com arquitetura de Aprendizagem Profunda, esta implementação gera o arquivo contendo os pesos e o treinamento final que será utilizado no *Ferramental de Produção*.
- IV. **Ferramental de Produção** – a implementação permite carregar na entrada uma imagem hiperespectral de biópsia de câncer do cólon, possibilitando a classificação de cada um dos seus pixels de acordo com a classificação CIS treinada no módulo de *Treinamento do Classificar*, gerando na saída uma nova imagem digital colorizada contendo o resultado do pré-diagnóstico.
- V. **Interface Web PAIH** – interface *Web* amigável, de fácil distribuição e que possibilita ao profissional patologista e pesquisadores selecionar e pré-diagnosticar imagens hiperespectrais de forma visual, rápida e simples.

A Figura 17 apresenta o processo de funcionamento da aplicação proposta e a dependência entre os módulos desenvolvidos, onde:

- a) É o arquivo da imagem hiperespectral em seu formato padrão (FSM);
- b) o arquivo é então aberto na aplicação *MatLab* para análise dos metadados;
- c) o arquivo FSM é convertido e exportado para o formato CSV;
- d) base (diretório) de dados hiperespectrais no formato CSV;
- e) módulo *Gerador de banco de ROIs*;
- f) banco de ROIs, também no formato CSV;
- g) módulo de *Treinamento do Classificador*;
- h) arquivo JSON contendo a estrutura da RNA;
- i) módulo *Ferramental de Produção*;
- j) interfaces *Web* da aplicação;

- k) resultado visual do pré-diagnóstico;
- l) partes que constituem a *Interface Web PAIH*.

Figura 17 – Esquema de funcionamento e dependência dos processos desenvolvidos na pesquisa.

Fonte: próprio autor.

3.1 Ambiente de testes

Para minimizar o tempo da fase de experimentação desta pesquisa, foram utilizados dois ambientes de testes distintos. O primeiro é um computador convencional, destinado as tarefas de processamento mais leves como: o *Gerador do banco de ROIs*, o *Ferramental de Produção* e a *Interface Web PAIH*. O segundo é um servidor dedicado, utilizado nas tarefas que exigiram maior tempo e consumo de processamento como: a *Otimização e Validação Cruzada* e o *Treinamento do Classificador*. O servidor dedicado pertence ao Departamento de Computação e Matemática (DCM) da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP) da Universidade de São Paulo (USP). As especificações técnicas destes ambientes constam na Tabela 2.

Em ambos os ambientes de testes foram instalados os mesmo programas e bibliotecas, permitindo que todos os módulos desenvolvidos pudessem ser executados corretamente em qualquer ambiente. Os algoritmos de todos os módulos foram escritos utilizando a linguagem de programação *Python* em sua versão 3.7.0. A biblioteca de Aprendizado de Máquina *Google Tensorflow 2.0* e a API de RNA *Keras 2.4*, completam os requisitos básicos do ambiente de teste para esta pesquisa. Outras tecnologias suplementares

complementaram a estrutura dos módulos desenvolvidos e serão especificadas em suas respectivas subseções.

Tabela 2 – Especificações de *hardware* do ambiente de testes

Item	Computador convencional	Servidor Dedicado
Sistema Operacional	<i>Windows 10 Education</i>	<i>Linux Manjaro 20.1</i>
Processador	<i>Intel i7 7700k 4.20Ghz</i>	<i>Intel Xeon 6130 OURO 2.1Ghz</i>
Disco	1x SSD 240GB	2x SSD 320GB RAID 0
Placa de vídeo	<i>NVidia GTX 1080TI 11GB</i>	<i>NVidia QUADRO PAS-CAL P5000 16GB</i>
Memória	16GB de RAM	500GB de RAM

Fonte: próprio autor

3.1.1 Base de imagens

Neste trabalho foram utilizadas imagens hiperespectrais de tecido do cólon coletadas por meio de exames histológicos em lâminas contendo tecido de biópsia, assim como imagens digitais coradas com H&E. Todas as imagens foram fornecidas pelos pesquisadores do Laboratório de Foto Biofísica do Departamento de Física da FFCLRP-USP.

As imagens foram geradas inicialmente no formato FSM e separadas em três grupos: 9 imagens classificadas como cancerígenas; 9 imagens classificadas como inflamadas e; 9 imagens classificadas como saudáveis. Os arquivos possuem resoluções diferentes uns dos outros, mas cada pixel em todas as imagens possui exatamente um espectro composto com 1626 frequências, sendo esta a entrada do classificador da RNA de aprendizagem profunda em formato de matriz: pixels nas linhas e frequências nas colunas da entidade, originando uma imagem hiperespectral.

Cada imagem hiperespectral é acompanhada de uma imagem digital H&E (Figura 18) para facilitar a sua identificação. As imagens H&E foram utilizadas para a marcação das ROIs pelo patologista, no processo de rotulação dos pixels. Ao final, essas imagens foram reaproveitadas para identificação dos exames na *Interface Web PAIH*.

O artefato principal deste trabalho são os arquivos das imagens hiperespectrais. Assim, a primeira etapa da pesquisa foi encontrar uma forma de efetuar a leitura e a manipulação desses arquivos, afim de compreender sua estrutura. A extração dos dados por pixel dessas imagens, que serão utilizados para o aprendizado de características pela RNA com arquitetura de Aprendizagem Profunda, foi realizada através de um extrator desenvolvido com o *MatLab* pelos pesquisadores do Laboratório de Foto Biofísica do Departamento de Física da FFCLRP-USP. Por se tratar de um formato de arquivo atípico,

Figura 18 – Conjunto de 27 imagens H&E utilizadas na identificação da imagens hiperespectrais na base de imagens e para demarcação das ROIs pelo patologista.

Fonte: próprio autor.

a primeira etapa consistiu na conversão do formato FSM para CSV. As razões desta conversão serão explicadas no capítulo 4 (Resultados e Discussões).

3.2 Gerador do banco de ROIs

Ao abrir uma imagem hiperespectral no formato FSM na aplicação *MatLab* o sistema apresenta algumas entidades (Figura 19) que possuem, além de informações sobre os pixels da imagem hiperespectral, outras informações importantes sobre o arquivo como, por exemplo, a quantidade de frequências existentes. Esta etapa do trabalho teve como objetivo compreender a composição dessas entidades:

Figura 19 – Composição das principais entidades que fazem parte da estrutura de um arquivo de imagem hiperespectral no formato FSM.

r	46513x1626 double
dX	193
dY	241
wn	1626x1 double

Fonte: próprio autor.

As principais entidades alocadas no metadados de uma imagem hiperespectral são:

- wn – São as frequências (total de 1626) referentes ao intervalo de número de onda, variando de 750 cm^{-1} a 4000 cm^{-1} (de 2 em 2);
- dY – é a resolução vertical em pixel da imagem hiperespectral. No exemplo da Figura 19 são demonstrados 241 pixels;
- dX – é a resolução horizontal em pixel da imagem hiperespectral. No exemplo da Figura 19 são demonstrados 193 pixels;

- iv. r - é a matriz contendo a quantidade de pixels da imagem (dY multiplicado por dX) pela quantidade de frequências.

A matriz r , principal entidade entre as demais é a que efetivamente armazena os pixels da imagem hiperespectral e todas as 1626 frequências de cada pixel. Na Figura 20 é possível visualizar um trecho da estrutura desta entidade na aplicação *Matlab*.

Figura 20 – Matriz de dados da imagem hiperespectral.

	1	2	3	4	5	6	7	8	9	10	11
1	-0.2574	-0.1552	0.1403	-0.2029	-0.1615	-0.1141	-0.4660	-0.5075	-0.3563	-0.1133	0.2229
2	-0.4533	-0.4007	-0.3708	-0.3209	-0.2590	-0.0521	-0.0330	-0.0578	-0.2524	-0.0893	0.3451
3	0.2187	-0.2260	-0.2278	-0.1137	-0.0764	-0.1018	0.2527	0.3530	0.1861	-0.0416	-0.1691
4	-0.1574	-0.0875	-0.1274	-0.1116	-0.0631	-0.1621	-0.0762	-0.1004	-0.4304	-0.6219	-0.5078
5	-0.2889	-0.3753	-0.2845	-0.0278	-0.1808	-0.2443	-0.4025	-0.3711	-0.2570	-0.1921	-0.2699
6	-0.2773	-0.2798	-0.2669	-0.1703	0.0563	-0.2107	-0.2224	-0.1905	-0.1522	-0.1629	-0.2754
7	-0.2934	-0.3043	-0.3191	-0.2232	-0.0807	0.0249	0.2148	-0.1364	-0.2842	-0.2040	-0.0956
8	-0.2371	-0.0304	-0.0795	-0.1516	-0.0815	0.0048	-0.2291	-0.3558	0.0526	0.2734	0.3358
9	-0.3145	-0.3779	-0.4003	-0.3164	-0.3039	-0.4625	-0.3609	-0.0212	-0.0503	-0.2581	-0.1659
10	-0.1287	0.2902	-0.1807	-0.3353	-0.3154	-0.2641	-0.1568	-0.1102	-0.1399	-0.0170	0.1788
11	-0.4411	-0.3727	-0.1191	0.0239	-0.1916	-0.4059	-0.1419	0.3321	0.0122	-0.3343	-0.2897
12	-0.0178	0.4196	0.3436	-0.1449	-0.3205	-0.3489	-0.1451	-0.0450	-0.0188	-0.2671	-0.2524
13	-0.1410	-0.2687	-0.2640	-0.2396	-0.3376	-0.3857	-0.3835	-0.1495	0.4237	0.2917	-0.0042
14	-0.2530	-0.3050	-0.3336	-0.4318	-0.3629	-0.2793	-0.1538	-0.3683	-0.4680	-0.3762	-0.1234
15	-0.2184	-0.2622	0.0727	0.0427	-0.1433	-0.2132	-0.0630	0.0758	0.2239	-0.0385	0.1476
16	-0.1247	-0.1051	-0.1597	-0.2524	-0.1344	0.0869	-0.0563	-0.1765	-0.2313	-0.2104	-0.0789
17	-0.1491	-0.1626	-0.2459	-0.4576	-0.3531	-0.2035	0.1706	-0.1333	-0.3225	-0.4274	-0.3554
18	-0.0879	-0.1586	-0.3922	-0.4612	-0.3905	-0.1310	0.3163	0.1416	-0.1039	-0.1558	-0.2375

Fonte: próprio autor.

Cada coluna da Figura 20 representa uma frequência hiperespectral do pixel e cada linha representa um pixel da imagem. Para localizar corretamente o valor de absorvância dos pixels na frequência correspondente ao número de onda, basta verificar na entidade *wn* o valor correspondente a coluna da matriz na entidade r (1 a 1626). Outra alternativa é convertendo os valores por meio da equação: $750 + (2 * \text{coluna da matriz } r) - 2$, em que o resultado será um valor de onda em cm^{-1} , por exemplo: $750 + (2 * 1026) - 2 = 2800 \text{ cm}^{-1}$. Assim como exposto no trabalho de Silva (2013), para visualizar a imagem em uma das 1626 frequências disponíveis, é necessário converter uma determinada coluna (frequência) para uma nova matriz de dados com largura e altura iguais as informações contidas nas entidades dY e dX , especificado nos metadados do arquivo. A aplicação *Gerador do banco de ROIs* possui um algoritmo que efetua automaticamente este rearranjo, mas ela também pode ser executada facilmente com a função *imshow(m(x : y : λ))* do *MatLab*, onde: *imshow()* é a função para exibição de imagens 2D; m é a matriz da imagem hiperespectral; x é a quantidade de pixels horizontais a serem exibidos; y a quantidade de pixels verticais a serem exibidos e λ a frequência hiperespectral do eixo z da matriz tridimensional. A Figura 21 simula como é feita esta conversão.

Uma vez analisada a estrutura do arquivo FSM da imagem hiperespectral, o próximo passo foi entregar ao patologista todas as 27 imagens H&E (sendo 9 de cada classificação CIS) para que ele realiza-se a demarcação manual das regiões identificadas visualmente com indícios de câncer, inflamação ou de tecido saudável. Por se tratarem de demarcações manuais livres, as regiões marcadas pelo patologista (Figura 22a) não

Figura 21 – Simulação de como é feito o rearranjo de uma frequência hiperespectral para geração de uma imagem 2D: a) imagem hiperespectral FSM; b) matriz rearranjada de acordo com as resoluções dX e dY e; c) imagem 2D convertida.

Fonte: próprio autor.

possuem a mesma quantidade de pixels. Logo, foi definido um tamanho digital padrão para todas as ROIs, a partir da região central de cada demarcação manual, no tamanho de 50x50 pixels (Figura 22b), garantindo visualmente que cada ROI, posteriormente demarcada na imagem hiperespectral estivesse localizada dentro da região evidenciada pelo patologista na imagem H&E. Essa definição facilitou a construção da base de ROIs, uma vez que todas as 27 ROIs agora possuem a mesma resolução em pixels.

Figura 22 – a) Simulação da demarcação realizada manualmente pelo patologista na imagem HE; b) ROI (quadrado vermelho) definida digitalmente a partir da demarcação manual, com dimensão de 50x50 pixels na imagem hiperespectral

Fonte: próprio autor.

Na geração da base de ROIs, utilizada para treinar o classificador da RNA com arquitetura de Aprendizagem Profunda, foi preciso agrupar as informações de resolução e das coordenadas de todas as 27 ROIs - especificadas com 50x50 pixels - em cada imagem e armazená-las em um banco de dados (Figura 23) para que o módulo pudesse gerar a

base de ROIs automaticamente a partir de todas as imagens hiperespectrais da base de imagens. O programa *ImageJ*⁶ foi utilizado para localizar as coordenadas das ROI.

Figura 23 – Banco de dados do módulo Gerador de banco de ROIs contendo informações das imagens hiperespectrais (da esquerda pra direita): nome da imagem, resolução horizontal, resolução vertical, coordenada da ROI no eixo X, coordenada da ROI no eixo Y e classe da imagem (CIS).

	1	2	3	4	5	6
1	image	x	y	coord_x	coord_y	classe_id↓
2	c4	193	241	102	99	c↓
3	c6	104	253	118	12	c↓
4	i1	204	253	144	104	i↓
5	i2	551	283	63	277	i↓
6	s12	272	128	37	199	s↓
7	s15	182	232	155	124	s↓

Fonte: próprio autor.

3.3 Treinamento e otimização da rede

Após a construção da base de ROIs, o próximo passo foi a definição da arquitetura da RNA de Aprendizagem Profunda. Apesar da popularidade deste tipo de RNA, há ainda um problema comum em como definir uma arquitetura ideal de hiperparâmetros⁷ para uma determinada aplicação. Os hiperparâmetros são as informações que definem a estrutura geral de uma RNA com arquitetura de Aprendizagem Profunda como: quantidade de épocas, função de ativação, otimizadores, etc. Testar manualmente diferentes arquiteturas pode tornar o trabalho oneroso, além de demandar tempo de desenvolvimento (NEARY, 2018).

Utilizamos na otimização dos hiperparâmetros a biblioteca de Aprendizado de Máquina *scikit-learn*. No *scikit-learn*, a técnica de definição dos hiperparâmetros - chamada de *Tuning* - é fornecida por meio da classe *GridSearchCV*, onde, ao utilizá-la, deve-se fornecer um dicionário de hiperparâmetros a serem avaliados no argumento *param_grid* da função. O resultado dos melhores hiperparâmetros é definido por meio de um método automatizado de Validação Cruzada logo após a avaliação de cada modelo individual (conjunto de parâmetros) do dicionário. Na documentação do *Keras* há dezenas de possíveis parâmetros que podem ser testados. Aqui destacaremos somente os melhores indicados após o *Tuning*.

⁶ *ImageJ* é um programa de processamento de imagem *Java* de domínio público. Disponível em: <<https://imagej.nih.gov/ij/docs/intro.html>>

⁷ Existem dezenas de hiperparâmetros disponíveis para o *Keras*. Nesta seção destacaremos somente aqueles utilizados no projeto. Outras informações pode ser adquiridas na documentação do *Keras*. Disponível em: <<https://keras-team.github.io/keras-tuner/documentation/hyperparameters/>>

A Validação Cruzada é uma técnica muito utilizada, principalmente para estimar o erro de testes associados ao ajuste de um determinado método de aprendizagem estatística e Aprendizado de Máquina, evitando o chamado superajuste (*overfitting*). Neste trabalho, esta técnica foi utilizada tanto para a definição dos hiperparâmetros quanto para a validação da base de ROIs. Existem vários tipos de Validação Cruzada para RNAs e o mais popular entre eles é o método *holdout*. Este método consiste em dividir o total dos dados destinados ao classificador em duas partes, sendo uma designada para o treinamento da RNA e a outra para o seu teste. A parcela da base destinada ao treinamento é utilizada para estimação dos parâmetros e, normalmente corresponde a no máximo 70% do volume total dos dados. Os outros 30% são então utilizados para estimar a acurácia do modelo (AZEVEDO, 2018).

O modelo *holdout* foi utilizado na fase inicial deste projeto e serviu como referência de acurácia até então. Porém, nesta última fase optamos por uma Validação Cruzada mais analítica, principalmente por se tratar de um volume de dados muito maior daquele utilizado na primeira versão do experimento. Assim escolhemos a Validação Cruzada do tipo K-Fold. Este método implica em dividir aleatoriamente os dados em uma quantidade de K grupos (denominados Folds) e de tamanhos semelhantes. O processo é então repetido K vezes, de modo que cada um dos K grupos sejam utilizados uma vez como dado de teste para validação do modelo, resultando em K estimativas do erro de teste (JAMES et al., 2013).

De acordo com os estudos de James et al. (2013), o resultado final do K-Fold é o desempenho médio do classificador nos K testes. A justificativa de repetir os testes múltiplas vezes é o de aumentar a confiabilidade da precisão do classificador da RNA com arquitetura de Aprendizagem Profunda. Para produzir estimativas das taxas tanto de erro de testes quanto de variância, não tendenciando a valores muito altos, é comum utilizar a Validação Cruzada K-Fold definindo $K = 5$ ou $K = 10$ (utilizado neste experimento).

Uma vez estabelecida a base de ROIs, a otimização da RNA com arquitetura de Aprendizagem Profunda e a Validação Cruzada K-Fold a próxima etapa foi o *Treinamento do Classificador*. Tal como exposto nos capítulos anteriores, o objetivo do classificador da RNA com arquitetura de Aprendizagem Profunda desenvolvido nesta pesquisa é o de analisar os pixels de uma imagem hiperespectral em todas as suas 1626 frequências, bem como a sua classificação CIS, possibilitando que o módulo *Ferramental de Produção* possa pré-diagnosticar qualquer imagem hiperespectral de câncer do cólon analisada. Foi importante estabelecer uma forma de armazenar em disco o treinamento da RNA com arquitetura de Aprendizagem Profunda, uma vez que este processo pode levar horas ou até dias de processamento. O *Keras* permite que o modelo final da RNA com arquitetura de Aprendizagem Profunda possa ser salvo em disco, utilizando o formato de arquivo JSON. O treinamento desta pesquisa foi salvo com o nome de *nn_structure.json*, como

exposto na Figura 17h. O *Keras* permite que os pesos estabelecidos também possam ser gravados em um arquivo separado no formato de arquivo HDF5⁸. Assim como o modelo, os pesos também foram salvos em um arquivo distinto de nome *nn_weights.h5*. Ao final, esses dois arquivos serão carregados no módulo *Ferramental de Produção*, possibilitando a execução do algoritmo de classificação.

3.4 Ferramental de Produção

O Ferramental de Produção é, como o próprio nome sugere, o *core* do sistema proposto neste projeto. Este é o módulo responsável por carregar o classificador da RNA com arquitetura de Aprendizagem Profunda (arquivo JSON, gerado no módulo Treinamento do Classificador), que permite classificar os pixels de uma imagem hiperespectral e fornecer o pré-diagnóstico do exame de biópsia de câncer do cólon. Na primeira etapa do processo, o Ferramental de Produção faz o carregamento dos arquivos *nn_struture.json* e *nn_weights.h5*.

Um dos desafios deste projeto foi o de separar as regiões de parafina dos tecidos. É importante que o classificador analise somente pixels relativos as áreas da biópsia, uma vez que foram treinadas somente regiões correspondentes a tecidos no módulo *Treinamento do Classificador*. Qualquer material diferente de tecidos de biópsia poderia ocasionar um resultado indesejado no pré-diagnóstico, resultando em falsos positivos. Sendo assim, a etapa seguinte teve como objetivo separar da imagem hiperespectral as áreas de tecido da biópsia, das áreas correspondentes a parafina. Para isso foram utilizados três métodos de Processamento de Imagens Digitais: o limiar (*threshold*) por OTSU, a Dilatação e a Erosão. O propósito foi o de criar uma máscara binária a partir de uma matriz de dados, utilizando uma frequência hiperespectral específica como base. Nesta fase utilizamos a frequência correspondente ao número de onda igual a 1600 cm^{-1} na definição da máscara binária, por possuir um nível de nitidez e detalhamento equilibrado, segundo a recomendação de pesquisadores do Laboratório de Foto Biofísica. Por fim, a proposta foi a de utilizar esta imagem digital 2D e separar o objeto (tecido) do fundo da imagem (parafina), utilizando a máscara binária e posteriormente efetuar a calibração necessária de brilho, picos e vales utilizando os métodos de Dilatação e Erosão, afim de reduzir a perda de material de tecido (OTSU, 1978; FACON, 1996; SOILLE, 1999).

Todo o processo (Figura 24) utilizou a biblioteca OpenCV⁹ disponível para a linguagem de programação *Python*. A biblioteca possui dezenas de funções de *Thresholding*

⁸ O HDF5 armazena eficientemente grandes matrizes de números e é o formato padrão utilizado pelo *Keras*. Disponível em: <<https://docs.h5py.org/en/stable/>>

⁹ OpenCV é uma biblioteca multiplataforma livre para o desenvolvimento de aplicativos na área de Visão computacional. Disponível em: <https://docs.opencv.org/master/d6/d00/tutorial_py_root.html>

que facilitam o processo de desenvolvimento. Uma vez finalizada a criação da máscara binária, a partir da frequência correspondente ao número de onda igual a 1600 cm^{-1} e utilizando as técnicas propostas (OTSU, Dilatação e Erosão), o sistema compara as coordenadas de pixels da imagem hiperespectral com as da máscara binária onde, caso o valor da máscara seja igual a 0, o classificador ignora este pixel e o *Ferramental de Produção* define sua cor como preto. Nesta etapa o sistema já inicia também a classificação dos pixels para o pré-diagnóstico. Para isso o sistema gera outras 3 matrizes: uma para os pixels classificados como cancerígenos (mC), outra para os inflamados (mI) e por fim uma última para os saudáveis (mS). Para realizar este processo, o *Ferramental de Produção* envia para o classificador o pixel e todas as frequências hiperespectrais para análise. Se o pixel for classificado como cancerígeno, o *Ferramental de Produção* colore o pixel na cor vermelha somente na matriz mC. Se o pixel candidato for classificado como inflamado, sua coordenada na matriz mI é colorizada como verde. Por fim, pixels classificados como saudáveis, levam a colorização como azul na matriz mS.

Figura 24 – Processo de remoção da parafina e classificação dos pixels.

Fonte: próprio autor.

Ao final do processo, o sistema gera uma imagem digital no formato PNG (*Portable Network Graphics*), em que cada canal ou matriz do canal RGB corresponde a colorização realizada pela RNA com arquitetura de Aprendizagem Profunda (mC, mI e mS), entregando no final a colorização dos pixels por categoria. O *Ferramental de Produção* também apresenta ao usuário a porcentagem de pixels classificados em cada categoria. Essas etapas consitituem o pré-diagnóstico do sistema e serão detalhados no capítulo 4 (Resultados e Discussões).

3.5 Interface Web PAIH

A última etapa do projeto teve por objetivo o desenvolvimento da camada de UI (Interface do usuário - do inglês *User Interface*) que possibilite uma interação simples com o sistema. A UI opera como frontend do *Ferramental de Produção*, que em uma visão mais técnica, trata-se da camada de *backend* do sistema. A junção de ambas as camadas complementam a aplicação PAIH.

O PAIH foi desenvolvido com o objetivo de tornar simples a interação do usuário com a *engine* da RNA com arquitetura de Aprendizagem Profunda, proposta nesta pesquisa. O *frontend* foi programado utilizando as bibliotecas aplicáveis a *Web* (*Flask*) da linguagem de programação *Python*, bem como um conjunto de padrões e práticas de desenvolvimento que facilitam a estruturação e o reuso do código, tais como padrões de projeto (*design-patterns*) (EVANS et al., 2004) e outras tecnologias para Internet, como HTML5, CSS, JSON, Javascript e AJAX. Na interface inicial (Figura 25), o PAIH permite que dois tipos de atores possam utilizar a ferramenta sendo: um pesquisador ou patologista.

Figura 25 – a) Interface de login. b) Interface inicial do sistema, após a autenticação.

Fonte: próprio autor.

Dependendo do perfil selecionado, as informações apresentadas no sistema sofrem alterações de acesso. Em geral, usuários com nível de pesquisador possuem acesso irrestrito aos arquivos do sistema, já usuários com nível de patologista podem manipular somente as imagens hiperespectrais em CSV, visualizar imagens digitais H&E e executar pré-diagnósticos (Figura 26). A limitação de acesso ocorre pois não há necessidade de

que usuários patologistas tenham acesso, por exemplo, aos arquivos FSM, uma vez que eles são úteis somente em atividades de pesquisa, principalmente visando contribuições em trabalhos futuros. Com isso, seu acesso por terceiros viola as regras da aplicação.

Figura 26 – Interfaces comparando a visão entre o usuário pesquisador (a) com a do usuário patologista (b)

Fonte: próprio autor.

A aplicação reconhece e organiza automaticamente diretórios e subdiretórios contendo um conjunto de arquivos compatíveis com o sistema, tais como: imagens hiperespectrais (FSM), imagens digitais (PNG ou TIFF) e arquivos CSV. Esses diretórios ficam alocados em um servidor, fazendo do PAIH um repositório de exames de biópsia de tecido do cólon em nuvem. Ao selecionar uma imagem hiperespectral na caixa de seleção, o usuário pode pressionar o botão *Prediagnose* para que a engine do *Ferramental de Produção* seja acionada e o pré-diagnóstico processado. Ao final do processo, o sistema exibe a imagem H&E, a hiperespectral na frequência correspondente ao número de onda igual a 1600 cm^{-1} , o resultado do pré-diagnóstico colorizado em uma imagem PNG na classificação CIS e o percentual de acerto da RNA com arquitetura de Aprendizagem Profunda (Figura 27).

Ainda na Figura 27 é possível notar que a imagem hiperespectral e a imagem H&E não possuem a mesma resolução, proporção e rotação e isso fica ainda mais claro no círculo de destaque na Figura 28. Esse é um ponto importante a ser mencionado, para que não haja dúvidas em relação a autenticidade do pré-diagnóstico realizado pelo PAIH. Esse detalhe, apesar de comum e de não apresentar quaisquer problemas no processo de classificação da RNA com arquitetura de Aprendizagem Profunda, é presente em toda a base de imagens utilizada nesta pesquisa.

Figura 27 – Resultado do pré-diagnóstico realizado pela RNA com arquitetura de Aprendizagem Profunda.

Fonte: próprio autor.

Figura 28 – a) imagem hiperespectral na frequência correspondente ao número de onda igual a 1600 cm^{-1} ; b) imagem H&E. Ambas fazem parte do mesmo exame, como mostra o círculo em destaque, mas possuem características espaciais distintas.

Fonte: próprio autor.

Resultados e Discussões

Este capítulo apresenta os resultados obtidos com o desenvolvimento do classificador proposto. Na seção 4.1 são apresentados os resultados do processo de geração da base de ROIs; na seção 4.2 as definições de hiperparâmetros no processo de *Tuning*; na seção 4.3 os resultados da Validação Cruzada, e; na seção 4.4 o resultado do pré-diagnóstico gerado pela RNA com arquitetura de Aprendizagem Profunda proposta na pesquisa.

4.1 Base de ROIs

Ao final, foram extraídas 27 ROIs (uma de cada imagem hiperespectral disponível), sendo 9 cancerígenas, 9 inflamadas e 9 saudáveis. Cada ROI possui dimensão de 50×50 pixels (total de 2500 pixels) e cada pixel possui 1626 frequências a serem analisados pela RNA com arquitetura de Aprendizagem Profunda. Em suma, os pixels são as instâncias a serem analisadas e as frequências são os atributos de cada instância. O número total de pixels da base de ROIs foi de 67500 pixels, com um volume em disco totalizando aproximadamente 1.2GB.

Como forma de validar o algoritmo, a aplicação gera em um diretório a ROI em formato PNG (Figura 29). Apesar da comprovada eficiência do algoritmo na execução do recorte da ROI, achamos importante realizar esta prova de conceito para minimizar a probabilidade de erros no processo.

Todos os pixels foram organizados na base de ROIs seguindo o mesmo esquema explicado na Figura 21, onde as linhas representam os pixels e as colunas as frequências. A última coluna da Figura 30 é a classificação CIS de cada pixel da ROI.

É importante ressaltar o motivo da conversão do arquivo FSM para CSV neste projeto. Como comentado, por se tratar de um formato de arquivo atípico, a primeira etapa consistiu na conversão do formato FSM para outro mais popular, que pudesse usu-

Figura 29 – Imagem hiperespectral e sua ROI, ambas na frequência correspondente ao número de onda igual a 1600 cm^{-1} .

Fonte: próprio autor.

Figura 30 – Últimas colunas da base de ROIs. Na coluna 1627 constam a classificação CIS dos pixels.

1622	1623	1624	1625	1626	1627
1622	1623	1624	1625	1626	1627
0.0013993	0.0021655000000000003	0.0023673	0.0020245000000000003	0.0017721	c↓
0.0088324	0.0090552000000000001	0.0094907000000000001	0.0095867	0.0090589	c↓
0.0131680000000000001	0.013624	0.0140759999999999998	0.0144669999999999999	0.0150229999999999998	c↓
0.018637	0.019352	0.019415	0.019512	0.0205449999999999997	c↓
0.028894	0.028582	0.028539	0.029187	0.029699	c↓
0.01871	0.0185919999999999997	0.018613	0.019253	0.0200669999999999998	c↓
-0.00269530000000000003	-0.0025176	-0.00254619999999999998	-0.0019602	-0.00098545	c↓
-0.011356	-0.011502	-0.011512	-0.0108209999999999999	-0.010494	c↓
-0.014547	-0.0142849999999999999	-0.0143550000000000001	-0.013963	-0.013147	c↓
-0.0185270000000000002	-0.0186770000000000003	-0.019194	-0.0188449999999999997	-0.0176770000000000002	c↓
-0.022469	-0.0224389999999999997	-0.021996	-0.021764	-0.0216449999999999998	c↓
-0.0243800000000000002	-0.02377	-0.0231860000000000002	-0.0226789999999999998	-0.0227490000000000002	c↓
-0.014343	-0.0146270000000000001	-0.0144030000000000001	-0.013973	-0.0136009999999999999	c↓
0.040005	0.0391699999999999996	0.039137	0.039518	0.039929	c↓
0.088403	0.089145	0.089653	0.089896	0.0901360000000000001	c↓
0.079789	0.080065	0.080647	0.081201	0.081684	c↓
0.054347	0.0544830000000000004	0.055254	0.0562059999999999999	0.056892	c↓

Fonte: próprio autor.

fruir de todos os recursos dos pacotes *Numpy*¹⁰ e *Pandas*¹¹ da linguagem de programação *Python*. Assim, foi definida a extensão CSV, um formato de arquivo muito comum, de fácil manipulação e compatível com diversos recursos do *Python*. A conversão foi realizada utilizando a função de exportação da aplicação *MatLab*, a qual possibilitou identificar informações importantes (metadados) acerca dos arquivos das imagens.

¹⁰ *NumPy* é um pacote para a linguagem de programação *Python* com suporte a *arrays*, matrizes multi-dimensionais e funções matemáticas. Disponível em: <<https://numpy.org/about/>>

¹¹ *Pandas* é uma ferramenta de código aberto para análise e manipulação de objetos do tipo *DataFrame*. Disponível em: <<https://pandas.pydata.org/about/index.html>>

4.2 Tuning

Com base no manual da API do *Keras* bem como nos trabalhos correlatos descritos no capítulo 2 (Estado da Arte), foram testadas três configurações distintas utilizando a função *GridSearchCV*. Cada conjunto de hiperparâmetros foi denominado aqui como cenário. Ao final do processo, a função *GridSearchCV* retorna qual o melhor hiperparâmetro a ser utilizado na base de treinamento da RNA com arquitetura de Aprendizagem Profunda, a partir do atributo *best_params*. O processo de *Tuning* realizado foi executado no ambiente de testes do servidor dedicado e todo o processo teve duração de aproximadamente 20 dias e 6 horas. Foram testados os seguintes hiperparâmetros:

- *units*: refere-se a quantidade de neurônios utilizados nas camadas ocultas;
- *batch_size*: é o número de exemplos de treinamento usados em uma iteração;
- *activation*: define a melhor função de ativação;
- *optimizer*: algoritmo utilizado para maximizar o desempenho da rede na descida do gradiente; e
- *epochs*: é o ciclo de processamento do cenário proposto na arquitetura, em que são repetidas até ser alcançado algum critério de parada, normalmente o fato da mudança de pesos se tornar muito pequena.

Na Tabela 3 estão especificados os cenários da seleção de hiperparâmetros utilizados no *Tuning*:

Tabela 3 – Seleção de hiperparametros utilizados no processo de *Tuning*

Hiperparâmetro	Cenário 1	Cenário 2	Cenário 3
<i>units</i>	1626	2439	2439
<i>batch_size</i>	16	32	64
<i>activation</i>	<i>tanh</i>	<i>relu</i>	<i>relu</i>
<i>optimizer</i>	<i>adam</i>	<i>adadelta</i>	<i>adam</i>
<i>epochs</i>	500	1000	3000

Fonte: próprio autor

É importante ressaltar que há trabalhos como em Wani et al., (2017) e Ranjit et al. (2019) que propõem novos métodos e discussões sobre a predefinição de hiperparâmetros para *Tuning* de forma a otimizar o desempenho de uma RNA. Neste trabalho optamos pelo uso da função *GridSearchCV*, mas sem deixar de lado a importância da otimização dos hiperparâmetros, visto que o objetivo nesta fase da pesquisa ainda parte do princípio de validarmos a proposta de classificação dos pixels. Assim, outros modelos de otimização poderão ser testados em trabalhos futuros.

4.3 Avaliação

A Validação Cruzada K-Fold apresentou ótimos resultados de acurácia nos três cenários propostos no processo de *Tuning*. Como mostra o gráfico da Figura 31, apesar de haver boa diferença de acurácia entre o cenário 1 e os outros dois, a diferença de acurácia entre o cenário 2 e o cenário 3 foi praticamente nula, quando ultrapassada 1000 épocas de processamento, oscilando em valores decimais em torno de 94%. O Cenário 1 manteve-se com acurácia em torno de 80%, mesmo chegando a 3000 épocas.

Figura 31 – Comparação de acurácia entre os cenários propostos para a RNA com arquitetura de Aprendizagem Profunda.

Fonte: próprio autor.

Com isso, mantivemos como escolha o Cenário 3, que apesar de possuir o dobro dos valores do hiperâmetro *batch_size* e três vezes mais épocas que o Cenário 2, ele não apresentou uma diferença significativa em tempo de processamento, atingindo uma acuraria de 94,4%, no caso, 0,2% mais eficiente que o Cenário 2, que apresentou 94,2% de acurácia.

4.4 Ferramenta de Pré-diagnóstico

O primeiro teste executado na ferramenta PAIH foi o pré-diagnóstico das 27 imagens hiperespectrais que foram utilizadas na extração das ROIs para a construção do classificador. O objetivo desta etapa foi o de analisar se a classe predominante coincidiu (ou não) com a classificação prévia indicada pelo patologista. A Tabela 4 apresenta os resultados na seguinte ordem: i) ID da Imagem: identifica o arquivo selecionado, onde a primeira inicial do arquivo corresponde a sua classe (CIS); ii) Classe: corresponde a classificação prévia indicada pelo patologista (CIS) e; iii) Acuracia do pré-diagnóstico da RNA em cada classe (Câncer, Inflamada ou Saudável): é porcentagem de pixels classificados corretamente em relação à classe (destacada em verde).

Tabela 4 – Resultado do teste de pre-diagnóstico nas imagens utilizadas na extração das ROIs para a construção do classificador.

ID da imagem	Classe	Câncer	Inflamação	Saudável
C11	C	99.9%	0,0%	0.1%
C13	C	92,0%	0.2%	7.8%
C18	C	89.6%	6.3%	4.1%
C20	C	96.4%	0.5%	3.1%
C21	C	71.2%	16.9%	11.9%
C4	C	99.5%	0,0%	0.5%
C6	C	73.3%	3.8%	23,0%
C8	C	76.1%	20.3%	3.6%
C9	C	88.6%	7.2%	4.2%
I11	I	7.6%	85.8%	6.6%
I15	I	1,0%	68.1%	30.9%
I1	I	0.8%	98.2%	1,0%
I20	I	8.5%	80.4%	11.1%
I25	I	0.9%	75.4%	23.7%
I26	I	4.9%	94.2%	0.9%
I2	I	0.4%	83.3%	16.3%
I5	I	9.5%	66.4%	24.1%
I8	I	4.3%	89.3%	6.4%
S12	S	0.1%	0.4%	99.5%
S15	S	0.1%	4,0%	95.9%
S16	S	9.1%	10.5%	80.4%
S17	S	0.2%	4.8%	95,0%
S18	S	0.4%	2.5%	97.1%
S21	S	2,0%	13.7%	84.3%
S2	S	2.6%	43.2%	54.2%
S5	S	8.1%	8.2%	83.7%
S9	S	3.3%	16.2%	80.5%

Fonte: próprio autor

Todas as 27 imagens obtiveram predominância em relação ao diagnóstico previamente realizado pelo patologista, com a maioria da acurácia entre 80% e 99%, indicando a eficiência do classificador desenvolvido. Todavia, analisando cada caso, na classe de arquivos com iniciais em C (cancerígenos), a acurácia mais alta foi do arquivo C11 (Figura 32a), atingindo 99,9% de acerto e a acurácia mais baixa foi do arquivo C21 (Figura 32b), atingindo 71,2% de acerto. A maioria dos arquivos desta classe obtiveram valores acima de 88% de acerto, sendo a classe mais bem classificada. Em relação aos arquivos com iniciais em I (inflamados), a acurácia mais alta foi obtida no arquivo I1 (Figura 32c), atingindo 98,2% de acerto e a acurácia mais baixa foi do arquivo I5 (Figura 32d), atingindo 66,4% de acerto. Nesta classe a maioria dos arquivos também obtiveram valores acima de 80% de acerto. E por fim, na classe de arquivos com iniciais em S (saúdáveis), a acurácia mais alta foi obtida no arquivo S12 (Figura 32e), atingindo 99,5% de acerto e a mais baixa no arquivo S2 (Figura 32f), atingindo 54,2% de acerto, sendo esta a mais baixa classificação entre todas as classes. Nesta classe a maioria dos arquivos também obtiveram valores de acurácia acima de 80%.

Figura 32 – . Relação dos resultados mais altos e baixos por classe.

Fonte: próprio autor.

Por fim, como forma de comprovar a eficiência do classificador do *Ferramental de Produção* da aplicação PAIH no processo de validação de pré-diagnóstico, bem como de toda a hipótese proposta nesta pesquisa, não participaram do teste final de pré-diagnóstico nenhuma das 27 imagens utilizadas para geração da base de ROIs e treinadas no classificador. Para este teste, foram separadas outras 3 imagens inéditas, uma de cada classe, também fornecidas pelos pesquisadores do Laboratório de Foto Biofísica do Departamento de Física da FFCLRP-USP, sendo: C18 (cancerígena), I1 (inflamada) e S10 (saúdável).

As imagens originais passaram pelo processo de conversão do formato FSM para CSV para que pudessem ser lidas pela aplicação. Os resultados são apresentados na Figura 33.

Figura 33 – Resultado do pré-diagnóstico realizado pelo Ferramental de Produção: a) imagem C18 com predominância de câncer (vermelho); b) imagem I1 com predominância de inflamação (verde) e; c) imagem S10 com predominância de tecidos saudáveis. Todas coincidiram com a classificação prévia indicada pelo patologista

Fonte: próprio autor.

Os resultados coincidiram com o diagnóstico manual realizado pelo patologista. Como é possível visualizar, a Figura 33a possui forte predominância na cor vermelha. Isso resulta que a RNA com arquitetura de Aprendizagem Profunda classificou a imagem hiperespectral corretamente, com 92,6% da área do tecido classificada como câncer; 7,4% de área classificada como inflamação (verde) e 0% de área saudável (azul). Já a Figura 33b, possui predominância na cor verde, totalizando 71,9% da área do tecido classificada como inflamação, 3% de câncer e 25,1% de área saudável. E finalmente, a Figura 33c é predominantemente azul, com 81,3% de regiões classificadas como tecido saudável; 17,7% de câncer e 1% de inflamação.

É comum que em áreas cancerígenas também haja a ocorrência de tecido com inflamação. Isso explica o índice de 7,4% de tecidos inflamados na Figura 33a. Já os pontos em vermelho nas Figura 33b e 33c, apesar de mínimos, podem indicar um falso positivo do classificador. A propósito, essas considerações também valem para os testes descritos na Tabela 5. Há de se considerar que 27 ROIs é uma quantidade razoável, por tratar-se de classificar cada pixel, para testes de conceito em pesquisas relacionadas a RNA com arquitetura de Aprendizagem Profunda, mas que para um sistema de produção seriam necessários mais novos casos que pudessem compor a base de ROIs e consequentemente melhorar o treinamento do classificador. Assim, quanto maior o conjunto de treinamento melhor será o resultado do classificador e quanto maior o conjunto de testes mais exata será a estimativa do erro.

Uma solução para esses casos, seria a inclusão de Algoritmos de Homogeneização. Assim, o diagnóstico de uma imagem hiperespectral na classificação CIS poderia ser dividida em regiões homogêneas em relação a uma dada característica de cor para que, posteriormente, as diversas regiões possam ser analisadas de forma individual ou combinadas em macrorregiões.

Por fim, tendo como base os ambientes de testes especificados na Tabela 2, o custo computacional também foi algo importante a se observar. O processamento do pré-diagnóstico do *Ferramental de Produção* se demonstrou muito eficiente, com tempo médio de aproximadamente 45 segundos por exame, sendo executado no ambiente do computador convencional (*Intel i7*, 16GB de RAM e GPU *NVidia GTX 1080ti*). No ambiente do servidor dedicado (*Intel Xeon 6130*, 320GB de RAM e GPU *NVidia Quadro P5000*), esse tempo teve duração praticamente igual. Isso demonstrou que apesar da complexidade do processo de classificação, o sistema se mostra suficientemente rápido em um provável ambiente de produção. Todavia, em todos os testes realizados, o processo de classificação apresentou um consumo relativamente alto de memória RAM. No computador convencional (16GB de RAM), o processo de classificação apresentou consumo em torno de 3% a 5% de memória RAM por pré-diagnóstico (480MB a 800MB) e essa média manteve-se no servidor dedicado. Considerando um ambiente *Web*, onde dezenas ou centenas de exames possam ser realizados ao mesmo tempo e conseqüentemente possa elevar esses valores a números muito mais altos, acreditamos que seja importante uma análise mais criteriosa deste consumo em um ambiente de produção.

Conclusões

O objetivo deste trabalho foi demonstrar, investigar, implementar e avaliar um método e uma ferramenta para caracterização de regiões anômalas e saudáveis de tecidos do cólon em amostras de biópsia, a partir de espectros obtidos pela aplicação de infravermelho sobre as mesmas, visando o auxílio ao diagnóstico de lesões inflamatórias e cancerígenas. A hipótese principal foi a de investigar se as imagens hiperespectrais permitem identificar regiões anômalas nestes tecidos a partir do espectro dos seus pixels. Para tanto, foi implementado uma RNA com arquitetura de Aprendizagem Profunda capaz de classificar pixels em imagens hiperespectrais, pré-diagnosticando como: cancerígenas, inflamadas ou saudáveis. A partir dos testes do classificador e dos experimentos realizados, podemos concluir que essas tecnologias trabalhadas em conjunto permitem essa caracterização dos tecidos possibilitando o auxílio ao pré-diagnóstico de exames histológicos de câncer do cólon.

Além disso, foi proposta e implementada uma ferramenta *Web* (denominada PAIH) que torna o arquivamento de imagens hiperespectrais mais organizado e a manipulação da estrutura da RNA com arquitetura de Aprendizagem Profunda mais simples e acessível aos potenciais usuários do sistema, como patologistas e pesquisadores.

A estrutura e os processos de usabilidade da ferramenta permitem ainda que outras pesquisas relacionadas ao pré-diagnóstico de doenças, utilizando imagens hiperespectrais também possam ser futuramente incorporadas em sua engine. Algumas vantagens que podemos destacar da aplicação PAIH e dos módulos auxiliares desenvolvidos compreendem:

- Um mecanismo de geração de ROIs a partir de imagens hiperespectrais simples e eficiente, capaz de gerar dezenas de recortes em poucos minutos.
- A acuraria e os limites dos recursos do PAIH podem ser reavaliados a partir da geração de novas estruturas e hiperparâmetros de RNA com arquitetura de Apre-

dizagem Profunda, bastando somente a importação de um novo arquivo JSON para o ferramental.

- Pesquisadores podem utilizar diretamente o *backend* do *Ferramental de Produção* e gerar resultados de pré-diagnóstico em lote de forma rápida e eficiente.
- O módulo *Web* pode ser acessado por qualquer dispositivo que possua um *Web Browser* padrão, uma vez que o sistema não exige a instalação de nenhum tipo de *add-on* ou *cookie* para o seu funcionamento.
- Possui níveis de acesso distintos entre diferentes papéis de usuários, facilitando a manipulação dos dados.

Entre as desvantagens da aplicação podemos citar a dependência da aplicação *MatLab* para a conversão dos arquivos FSM para CSV. Este pode ser um empecilho, devido à dependência da utilização da aplicação para este fim. Outro ponto em desvantagem é o tempo elevado de treinamento que a RNA com arquitetura de Aprendizagem Profunda necessita para o aprendizado da classificação dos pixels de uma base de ROIs. O cenário proposto para o treinamento nesta pesquisa teve duração de aproximadamente 48 horas no ambiente do servidor dedicado. Apesar de ser um tempo relativamente comum no âmbito das RNAs, isso significa que utilizar novas imagens de entrada como artefatos passíveis de treinamento futuro podem demandar de uma logística criteriosa entre o tempo de uso da aplicação e o período de treinamento, evitando possíveis *overhead* de processamento no ambiente. Todavia, à medida que se conheçam mais informações sobre os dados e cenários, ambientes e métodos mais sofisticados poderão ser testados.

É importante ressaltar que o processo automático aqui proposto não provê o resultado final de exames histológicos de câncer do cólon, pois apenas a interpretação do patologista pode dar real significado aos dados, gerando assim o diagnóstico do exame. Logo, o ferramental desenvolvido nesta pesquisa servirá como ferramenta de apoio ao trabalho dos patologistas, gerando uma segunda opinião, com possível aumento da precisão e redução da subjetividade do processo.

Trabalhos Futuros

Como propostas para trabalhos futuros, podemos citar:

- Estudar uma forma de ler e gerar ROIs diretamente dos arquivos FSM das imagens hiperespectrais, com o objetivo de agilizar o processo de geração da base de ROIs.

- Implementar técnicas de co-registro para que a imagem H&E contendo o resultado do pré-diagnóstico seja aglutinada a sua respectiva imagem hiperespectral.

- Implementar algoritmos de homogeneização, com o objetivo de reduzir falsos positivos gerados na classificação dos pixels e possibilitar que a imagem digital, contendo o resultado do pré-diagnóstico, seja exibida de forma visualmente mais plástica, facilitando a interpretação dos usuários.

- Avaliar outros cenários. Ainda sobre os resultados apresentados no gráfico da Figura 31, mesmo havendo uma oscilação baixa entre os cenários 2 e 3, acreditamos que seja possível que em novas propostas de cenários os resultados possam convergir de forma diferente. Todavia, o objetivo nesta etapa do projeto foi o de obter uma acurácia aceitável e o índice de 94% foi suficiente para darmos andamento em todas as fases da pesquisa.

- Ampliar as funcionalidades da ferramenta PAIH incluindo recursos gráficos que permitirão ao patologista demarcar digitalmente na imagem hiperespectral novas ROIs em imagens candidatas a integrarem o classificador da RNA com arquitetura de Aprendizagem Profunda.

O trabalho aqui apresentado representa um passo a mais no estado da arte das técnicas de RNA com arquitetura de Aprendizagem Profunda aplicada em imagens hiperespectrais, e com isto contribui para a geração do conhecimento a partir da imensa quantidade de pesquisas relacionadas ao efetivo pré-diagnóstico informatizado de câncer.

Referências Bibliográficas

AREL, I; ROSE, D. and KARNOWSKI, T. Research frontier: Deep machine learning a new frontier in artificial intelligence research. *Comp. Intell. Mag.* , 5(4):13–18, November 2010.

ÁVILA, S. E. F. DE. Extended bag-of-words formalism for image Classification. Belo Horizonte: Universitté Pierre et Marie Curie/Universidade Federal de Minas Gerais, 2013. Tese de doutorado. Disponível em: <https://tinyurl.com/y46d2933>

AZEVEDO, L.; Métodos Estatísticos em Aprendizado de Máquinas para problemas de Classificação, Brasília: Instituto de Ciências Exatas, Departamento de Estatística, Universidade de Brasília, 2018. Dissertação de Mestrado em Estatística.

BALDI, P. and Sadowski, P.J. Understanding dropout. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *NIPS 26*, pages 2814–2822. Curran Associates, Inc., 2013.

BARRETO, J.M. Introdução às redes neurais artificiais. Laboratório de connexionismo e ciências cognitivas. Florianópolis: Departamento de Informática e Estatística, UFSC, 2002. 57p.

BENGIO, Y. Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127, January 2009.

BENGIO, Y.; LAMBLIN, P.; POPOVICI, D. and LAROCHELLE, H. Greedy layer-wise training of deep networks. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *NIPS 19*, pages 153–160. MIT Press, Cambridge, MA, 2007.

BENGIO, Y; COURVILLE, Y; AND P. VINCENT. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013

BENGIO, Y; COURVILLE, Y; and P. VINCENT. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013

BENZECRY, M.A. et al. Redes neurais artificiais: conceitos e aplicações. In: IX Encontro

de Engenharia de Produção, 9., 2005, Rio de Janeiro. Anais... Rio de Janeiro: UFRJ, 2005.

BEZERRA, E. Introdução à Aprendizagem Profunda. Tópicos em Gerenciamento de Dados e Informações, pages 57-86, edition 1, chapter 3, publisher SBC. CEFET/RJ. 2016

BIANCHI, M.; Extração de características de imagens de faces humanas através de wavelets, São Carlos: Escola de Engenharia de São Carlos, Universidade de São Paulo, 2006. Dissertação de Mestrado em Engenharia Elétrica.

BRAGA, A. DE P.; CARVALHO A. P. DE L. F.; LUDEMIR, T. B. Redes Neurais Artificiais: Teoria e Aplicações. LTC - Livros Técnicos e Científicos Editora S.A., 2000. 262p.

BRAGA, A.P.; CARVALHO, A.P.L.F.; LUDERMIR, T.B. Redes neurais artificiais: teoria e aplicações. 2.ed. São Paulo: LTC, 2007.

BYRNE, H. J. et al. Spectropathology for the next generation: Quo vadis? The Analyst, Royal Society of Chemistry (RSC), v. 140, n. 7, p. 2066-2073, 2015. ISSN 1364-5528. Available from Internet: <<http://dx.doi.org/10.1039/c4an02036g>>.

CALSAVERINI, R. S. Aprendizado profundo: Princípios e uso. Tecnologia de Crédito, v. 13, n. 94, p. 6-15, março 2016. Disponível em: <<https://tinyurl.com/y43g547y>>

CASTELANO, C.R. Estudo comparativo da transformada Wavelet no Reconhecimento de Padrões da íris humana. Dissertação (Mestrado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo, 2006.

CHAO, K. et. al. Chicken heart disease characterization by multi-spectral imaging. Applied Engineering in Agriculture, v.17, p.99-106, 2001.

CONNEL, S. D.; JAIN, A. K. (2001). Template-based online character recognition. Pattern Recognition Society. Published by Elsevier Science Ltd (2001), 1-14.

D. JIANG AND C. TANG AND A. ZHANG. Cluster Analysis for Gene Expression Data: A Survey. IEEE Trans. on Knowledge and Data Engineering, 16(11):1370-1386, 2004.

D. VAN DYK and X. L. Meng. The art of data augmentation (with discussion). Journal of Computational and Graphical Statistics, 10:1-111, 2001.

DE JUAN, A; MAEDER, M; HANCEWICZ, T; DUPONCHEL, L; TAULER, R. Chemo-metrics tools for image analysis. In: H. W. SALZER, R.; SIESLER (Ed.); Infrared and Raman Spectroscopic Imaging. p.65-106, 2009. Inglaterra: WILEY-VCH.

DE LIMA, F. A. Microespectroscopia infravermelha de processos inflamatórios e tumores de cólon. Tese (Doutorado - Programa de Pós-graduação em Física aplicada à Medicina e Biologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo. São Paulo, p. 119. 2016.

- DEHAVEN, V. R. Machine learning: Future capabilities and their implications. Tucson: University of Arizona, 2017. Disponível em: <https://bityli.com/gOfWx>
- DENG, L.; YU, D. Deep learning: Methods and applications. *Foundations and Trends R in Signal Processing*, v. 7, n. 3–4, p. 197–387, 2014. ISSN 1932-8346. Disponível em: <http://dx.doi.org/10.1561/20000000039>.
- DICARLO, J; ZOCCOLAN, D. and RUST, N. How does the brain solve visual object recognition? *Neuron*, 73(3):415 – 434, 2012
- DOUGHERTY, G. Digital image processing for medical applications. [S.l.]: Cambridge University Press, 2009.
- DUMITRU, E. and et al. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, 11:625–660, 2010.
- ELMASRY, G. et al. Early Detection of Apple Bruises on Different Background Colors Using Hyperspectral Imaging. *LWT-Food Science and Technology*, v.41, p. 337-345, 2008.
- EVANS, E.; FOWLER, M. Domain-driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley, 2004. Disponível em: <https://bityli.com/MHONE>.
- FACON, J., “Morfologia Matemática: Teoria e Exemplos”, Editor Jacques Facon, 1996.
- FU, K. S. (1983)] FU, K. S. (1983). A step towards unification of syntactic and statistical pattern recognition. *IEEE Trans Pattern Analysis and Machine Intelligence* 5,2 ,March of 1983, 200-205.
- GLOROT, X.; BORDES, A. and BENGIO, Y. Deep sparse rectifier neural networks. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, volume 15, pages 315–323. *Journal of Machine Learning Research - Workshop and Conference Proceedings*, 2011.
- GONZALES, R.C.; WOODS, R. E. Digital Image Processing. Addison-Wesley Pub. Company, 3ed, pg. 976, 2007.
- GOODFELLOW, I; BENGIO, Y. AND COURVILLE, A. Deep learning. Book in preparation for MIT Press, 2016.
- GRAVES, A. Generating sequences with recurrent neural networks. *CoRR*, abs / 1308.0850, 2013.
- GRAVES, A. and JAITLEY, N. Towards end-to-end speech recognition with recurrent neural networks. In *ICML-14*, pages 1764–1772, 2014.
- GRAVES, A.; MOHAMED, A. and HINTON, G. Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6645–6649. IEEE, 2013.

- GREGOR, K.; DANIHELKA, I.; GRAVES, A.; REZENDE, D. and WIERSTRA, D. Draw: A recurrent neural network for image generation. In David Blei and Francis Bach, editors, ICML-15, pages 1462–1471. JMLR Workshop and Conference Proceedings, 2015.
- HAGAN, M. T.; MENHAJ M. B. Training feedforward networks with the Marquardt algorithm. IEE Transactions on Neural Networks, v. 5, n. 6, 1994. p. 989-993.
- HAYKIN, S. (1999). Redes Neurais Artificiais, princípios e prática. 2ª edição, Bookman 1999.
- HAYKIN, S. (2001) Redes Neurais: Princípios e Prática. trad. Paulo M. Engel. Porto Alegre, Bookman. 900 p.
- HAYKIN, S. Neural Networks: A Comprehensive Foundation (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2007.
- HINTON G. E. and SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, July 2006.
- HINTON, G. E. To recognize shapes first learn to generate images. In Computational Neuroscience: Theoretical Insights into Brain Function. Elsevier, 2007.
- HINTON, G.; OSINDERO, S. and YEE-WHYE, T. A fast learning algorithm for deep belief nets. Neural Computation, 18(7):1527–1554, 2006.
- HINTON, G.E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I. and SALAKHUTDINOV, R. Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580, 2012.
- HOCHREITER, S. and JÜRGEN, S. Long short-term memory. Neural Comput., 9(8):1735–1780, November 1997.
- HOCHREITER, S.; BENGIO, Y.; FRASCONI, P. and SCHMIDHUBER, J. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In Kremer and Kolen, editors, A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press, 2001.
- HORNIK, K. Approximation capabilities of multilayer feedforward networks. Neural Networks, 4(2):251 – 257, 1991.
- IME-USP. Instituto de Matemática e Estatística da Universidade de São Paulo. Disponível em: <<https://sites.icmc.usp.br/andre/research/neural/>>. Acesso em: 18 de outubro de 2020.
- INCA. Instituto Nacional do Câncer. Disponível em: <<https://tinyurl.com/y6aq8uh7>>. Acesso em: 29 de agosto de 2020.
- JAIN, A. K.; DUIN, R.P.W. and MAO, J. (2000). Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,1 , January

of 2000, 4-37.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. An Introduction to Statistical Learning: with Applications in R. New York: Springer Velarg. 2013.

JANSSENS, O.; WALLE, R.; LOCCUFIER, M; HOECKE, S. Deep Learning For Infrared Thermal Image Based Machine Health Monitoring. IEEE/ASME TRANSACTIONS ON MECHATRONICS, 2017.

JARRETT, K.; KAVUKCUOGLU, K.; RANZATO, M. and LECUN, Y. What is the best multi-stage architecture for object recognition? In ICCV 2009, pages 2146–2153. IEEE, 2009.

JENSEN, J. R.; JACKSON, M. W. Introductory digital remote sensing image processing. v. 3. modulo 1: The remote sensing process. Disponível em www.cla.sc.edu/geog/rsrab/. Acesso em 21 novembro. 2018.

JOST, I. Aplicação de Deep Learning em dados refinados para mineração de opiniões. São Leopoldo: Unisinos, 2016. Disponível em: <https://tinyurl.com/y2eq26nv>

JUNYOUNG, C.; CAGLAR, G.; KYUNGHYUN, C. and YOSHUA, B. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014.

JÜRGEN, S. Deep Learning in neural networks: An overview. Neural Networks, 61:85 – 117, 2015.

KAIMING, H.; XIANGYU, Z.; SHAOQING, R. and JIAN, S. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, pages 1026–1034, 2015.

KIERSZENBAUM, A. L. Histologia e Biologia celular, Uma introdução à patologia. 3ª edição. Elsevier, 2012

KONDA, K.R.; BOUTHILLIER, X.; MEMISEVIC, R. and VINCENT, P. Dropout as data augmentation. CoRR, abs/1506.08700, 2015.

KRIZHEVSKY, A.; SUTSKEVER, I. and HINTON, G. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, NIPS 25, pages 1097–1105. Curran Associates, Inc., 2012.

KRUG, A. B; et. al. Análise e reconhecimento de padrões usando processamento de imagens e inteligência artificial. Revista de Iniciação Científica da ULBRA – Universidade Luterana do Brasil. Rio Grande do Sul, 2008.

KRÖSE, B.; SMAGT P. V. D. An introduction to Neural Networks. The University of Amsterdam. 8nd, 1996. 136p.

KUMAR, S.; MITTAL, G. S. Rapid Detection of Microorganisms Using Image Processing

- Parameters and Neural Network. Food and Bioprocess Technology, v3, n.5, p. 741-751, 2009. L. C. JUNQUEIRA E CARNEIRO, J. Histologia básica I. [12 . ed]. - Rio de Janeiro: Guanabara Koogan, 2013.
- LAROCHELLE, H; BENGIO, Y; LOURADO, J AND LAMBLIN, P. Exploring strategies for training deep neural networks. J. Mach. Learn. Res., 10:1-40, June 2009
- LASCH, P.; NAUMANN, D. Spatial resolution in infrared microspectroscopic imaging of tissues. Biochimica et Biophysica Acta (BBA) - Biomembranes, Elsevier BV, v. 1758, n. 7, p. 814-829, jul 2006. Disponível em: <<http://dx.doi.org/10.1016/j.bbamem.2006.06.008>>.
- LECUN, Y.; BENGIO, Y. and HINTON, G. Deep learning. Nature, 521(7553):436-444, 2015.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y. and HAFNER, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278-2324, November 1998.
- LECUN, Y; BENGIO, Y. HINTON, G. Deep learning. Nature, 521(7553):436-444, 2015.
- LEE, S.U; CHUNG S.Y; PARK, R.H. A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation. Computer Vision, Graphics, and Image Processing, (52), p 171 -190, 1990.
- LIJUN, X.; QIANG, M; PENGHAO, L.; XIAOCHUN Z.; ZHIWEI L. A Fault Diagnosis Method Of Insulator String Based On Infrared Image Feature Extraction And Probabilistic Neural Network. 10th International Conference on Intelligent Computation Technology and Automation. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. 2017.
- LOPES, F. Introdução ao Reconhecimento de Padrões e aplicações em problemas de Bioinformática. Grupo de Pesquisa em Bioinformática e Reconhecimento de Padrões, UTFPR-CP, 2012.
- MATHERS, C. D. et al. Global burden of disease in 2002: data sources, methods and results. [Geneve]: World Health Organization, 2003. (Global programme on evidence for health policy discussion paper, v. 54).
- MATTHEW D. ZEILER and Rob Fergus. Visualizing and Understanding Convolutional Networks, pages 818-833. Springer International Publishing, Cham, 2014.
- MCCULLOCH, W.S.; PITTS, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 5:115
- MELO, D. H. C. T. B. Uso de dados Ikonos II na análise urbana: testes operacionais na zona leste de São Paulo / D. H. C. T. B. Melo. - São José dos Campos: INPE, 2002. 146p. - (INPE-9865-TDI/870).

- MURRAY, Dan. Tuning neural networks with genetic algorithms. *AI Expert*, v. 9, n. 6, p. 26-32, 1994.
- NAIR, V and HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In Johannes Fürnkranz and Thorsten Joachims, editors, *ICML-10*, pages 807–814. Omnipress, 2010.
- NEARY, P. Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning. Department of Computer Science, Utah State University. Logan, UT, USA. 2018.
- NEVES, S. R. Algoritmos para Segmentação de Imagens Infravermelhas [Rio de Janeiro] 2003 X, 121 p, 29,7cm (COPPE/UFRJ, D.Sc., Engenharia Elétrica, 2003). Tese - Universidade Federal do Rio de Janeiro, COPPE
- NIE, S.; LI, Y.; REDD, D.C.B. and YU, N.T. Pattern recognition algorithms for tissue diagnosis by near-infrared ft-raman spectroscopy. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 13, No, 4, 1991. Review. E, Statistical, Nonlinear, and Soft Matter Physics. 76.
- NOVO, E. M. L. de M. Sensoriamento Remoto: Princípios e Aplicações. Editora Edgar Blücher Ltda. São José dos Campos, 1989. 308p.
- NUBIATO, K.E.Z.; Sistema de Imagem Hiperespectral para Classificação de Características qualitativas da Carne de Bovinos Nelore. Tese (Doutorado em Zootecnia) - Universidade de São Paulo, São Paulo, 2016.
- NWANKPA, C.; IJOMAH, W.; GACHAGAN, A.; MARSHALL, S. Activation functions: Comparison of trends in practice and research for deep learning. *Corr*, Glasgow, v.1811 p. 1–20 2018.
- ORTEGA, S.; CALLICÓ, G.M.; PLAZA, M.L.; CAMACHO, R.; FABELLO, H. and SARMIENT, R. Hyperspectral Database of Pathological In-Vitro Human Brain Samples to Detect Carcinogenic Tissues. Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Spain, 2011.
- OTSU, N. "A Threshold Selection Method from Gray-Level Histograms," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
- PAHO. Organização Pan-americana de Saúde – Câncer - <<https://www.paho.org/bra>>. Acesso em: 28 de agosto de 2020.
- PHAM, D.T. E KARABOGA, D. "Intelligent Optimisation Techniques", SpringerVerlag, 238p, 2000.
- PRATS-MONTALBÁN, J. M.; DE JUAN, A.; FERRER, A. Multivariate image analysis:

A review with applications. *Chemometrics and Intelligent Laboratory Systems*, v. 107, n. 1, p. 1–23, 2011.

QUEVEDO, R. A.; AGULERA, J. M.; PEDRESCHI, F. Color of Salmon Fillets by Computer Vision and Sensory Panel. *Food and Bioprocess Technology*, v.3, n.5, p. 637–643, 2010.

RANJIT, M; Ganapathy, G; Sridhar, K; Arumugham, V. Efficient Deep Learning Hyperparameter Tuning using Cloud Infrastructure. Microsoft Corporation, India. 2019

RANZATO, M.; POULTNEY, C.; CHOPRA, S. AND LECUN, Y. Efficient learning of sparse representations with an energy-based model. In Bernhard Schölkopf, John C. Platt, and Thomas Hofmann, editors, *NIPS 19*, pages 1137– 1144. MIT Press, 2006.

SAHOO, P.K.; SOLTANI,S.; WONG, A.K.C. A Survey of Thresholding Techniques. *Computer Vision, Graphics, and Image Processing*, (41), p233-260, 1988.

SAMMOUDA, R.; WANG, X.; BASILION, J. Hopfield Neural Network For The Segmentation Of Near Infrared Fluorescent Images For Diagnosing Prostate Cancer. 6th International Conference on Information and Communication Systems (ICICS), 2015.

SCHOWENGERDT, R. A. Remote Sensing: models and methods for image processing. v.3. London: Academic Press, 2006.

SIGNORONI, A.; SAVARDI, M.; PEZZONI, M.; GUERRINI, F.; ARRIGONI, F.; TURRA, G. Combining the use of CNN classification and strength-driven compression for the robust identification of bacterial species on hyperspectral culture plate images. Information Engineering Department, University of Brescia, Brescia, Italy. 2017

SILVA, C. S. Uso de imagens hiperespectrais na região do infravermelho próximo para identificar fraudes de documentos. Dissertação (Mestrado) - Universidade Federal de Pernambuco. Recife, p. 80. 2013.

SOILLE, P. Morphological Image Analysis - Principle and Applications. Springer, 1999.

SOUSA, M. F. M. Caracterização semântica de sinais acústicos. 134f. 2002. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores: Telecomunicações) - Faculdade de Engenharia, Universidade do Porto, Porto, Portugal. 2002.

SRIVASTAVA, N, Hinton, G.; Krizhevsky, A.; Sutskever, I. and Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014.

SRIVASTAVA, N.; MANSIMOV, E. and SALAKHUDINOV, R. Unsupervised learning of video representations using lstms. In David Blei and Francis Bach, editors, *ICML-15*, pages 843–852. JMLR Workshop and Conference Proceedings, 2015.

STUART, B. H. Infrared Spectroscopy: fundamentals and applications. [S.l.]: John Wiley

Sons Ltda, 2004. 244 p. ISBN 978-0-470-85428-0.

SUTSKEVER, I.; MARTENS, J. and HINTON, G. Generating text with recurrent neural networks. In L. Getoor and T. Scheffer, editors, ICML-11, ICML '11, pages 1017–1024, New York, NY, USA, June 2011. ACM.

SUTSKEVER, I.; VINYALS, O. and QUOC V. LE. Sequence to sequence learning with neural networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14, pages 3104–3112, Cambridge, MA, USA, 2014. MIT Press.

THEODORIDIS, S. AND KOUTROUMBAS, K. Pattern Recognition. Academic Press. 1999.

VANDERBEEK, R.; RUSSELL, K. An Analog Hardware Detection Scheme For Passive Remote Infrared Interferometers. Department of the Army, ERDEC, 1994.

WANI, M.; Afzal, S. A New Framework for Fine Tuning of Deep Networks. Postgraduate Department of Computer Science University of Kashmir, India, 2017.

WINDROW, B.; HOFF M. E. Adaptative switching circuits. Institute of Radio Engineers, Western Electronic Show and Convention, 1960 p. 96-104.

XIANG, Y; TIAN, J; ZHANG, Z; DAI, Y. Diagnosis Of Endometrial Cancer Based On Back-Propagation Neural Network And Near-Infrared Spectroscopy Of Tissue. Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.

XU, K.; BA, J.; KIROS, R.; CHO, K.; COURVILLE, A.; SALAKHUDINOV, R.; ZEMEL, R. and BENGIO, Y. Show, attend and tell: Neural image caption generation with visual attention. In David Blei and Francis Bach, editors, ICML-15, pages 2048–2057. JMLR Workshop and Conference Proceedings, 2015.

YANG, X. et al. Canonical correlation analysis networks for two-view image recognition. Information Sciences, v. 385–386, p. 338–352, 2017.