



Trabalho de Conclusão de Curso

Classificação de Nódulos Pulmonares Utilizando Redes Neurais Convolucionais 3D e Otimização de Hiperparâmetros

Anthony Emanuel de Albuquerque Jatobá
aeaj@ic.ufal.br

Orientador:
Prof. Dr. Marcelo Costa Oliveira

Maceió, Abril de 2019

Anthony Emanuel de Albuquerque Jatobá

Classificação de Nódulos Pulmonares Utilizando Redes Neurais Convolucionais 3D e
Otimização de Hiperparâmetros

Monografia apresentada como requisito parcial
para obtenção do grau de Bacharel em Ciência da
Computação da Universidade Federal de Alagoas -
UFAL, Campus A.C. Simões.

Orientador: Prof. Dr. Marcelo Costa Oliveira

Maceió
2019

Anthony Emanuel de Albuquerque Jatobá

Classificação de Nódulos Pulmonares Utilizando Redes Neurais Convolucionais 3D e
Otimização de Hiperparâmetros

Monografia apresentada como requisito parcial
para obtenção do grau de Bacharel em Ciência da
Computação da Universidade Federal de Alagoas -
UFAL, Campus A.C. Simões.

Data de Aprovação: 23/04/2019

Banca Examinadora

Prof. Dr. Marcelo Costa Oliveira
Universidade Federal de Alagoas
UFAL - Instituto de Computação
Orientador

Prof. Dr. Thales Miranda Vieira
Universidade Federal de Alagoas
UFAL - Instituto de Matemática
Examinador

Prof. Dr. Tiago Figueiredo Vieira
Universidade Federal de Alagoas
UFAL - Instituto de Computação
Examinador

AGRADECIMENTOS

À minha família, em especial meu pai, Emanuel, e minha mãe, Anne, pelo suporte em todos esses anos.

À minha namorada Adélia, pelo carinho e companheirismo.

À galera do toco e toco++, especialmente ao Júlio, por ter sido um irmão nesses anos de curso e ao Douglas, por ter acompanhado meu trabalho de perto. Aos colegas da 2015.1 e chegados.

Ao meu orientador Marcelo, pela atenção, disponibilidade e por ter me ensinado tanto. Aos colegas de laboratório, em especial ao Lucas, que muito me ajudou desde o início do meu trabalho.

Aos professores que aceitaram fazer parte de minha banca, Thales e Tiago.

A todos que formam o Instituto de Computação.

A Deus por colocar todas essas pessoas no meu caminho.

The kernel has died, and the automatic restart has failed.

Jupyter Notebook

RESUMO

Câncer é um termo para um grupo de doenças caracterizadas pelo crescimento desordenado de células que invadem tecidos e órgãos. O câncer de pulmão é a forma mais comum desta doença, além de ser o tipo de câncer que mais mata, compreendendo 18,4% do total de mortes por câncer. O diagnóstico da doença em seus estágios iniciais é crucial para a sobrevivência do paciente, que pode ter taxas de sobrevivência até 90% contra 15% nos seus últimos estágios. A principal manifestação do câncer pulmonar é o nódulo pulmonar. O advento da tomografia computadorizada permitiu um diagnóstico mais preciso do câncer de pulmão, porém este tipo de exame ainda é uma tarefa desafiadora para os radiologistas e sujeita a fatores subjetivos como fadiga e experiência. Assim, se faz necessário o desenvolvimento de ferramentas computacionais para auxílio ao diagnóstico, de forma a tornar esta tarefa mais rápida e precisa. Técnicas tradicionais funcionam por meio da extração prévia de atributos descritores e posterior aplicação de técnicas de aprendizagem de máquina para classificação dos nódulos. Recentemente, a aprendizagem profunda ganhou ampla aceitação, por ser capaz de aprender atributos de forma automática. As Redes Neurais Convolucionais são uma modalidade de aprendizagem profunda muito usada para classificação de imagens e amplamente empregada em aplicações de diagnóstico médico. No entanto, estas redes geralmente operam sobre imagens bidimensionais, enquanto o exame de tomografia computadorizada gera informação tridimensional. Recentemente, passou-se a investir esforços em Redes Neurais Convolucionais 3D, capazes de utilizar a informação tridimensional no seu processo de aprendizagem. No entanto, estas redes ainda estão em sua infância em aplicações do tipo, necessitando de mais estudos acerca de questões como a forma como os nódulos são processados para serem usados pelas redes. Este trabalho propõe uma metodologia para avaliar uma série de estratégias para criação de volumes de nódulos pulmonares para Redes Neurais Convolucionais 3D, utilizando otimização de hiperparâmetros para a escolha da topologia adequada para as redes e validação cruzada para validação dos resultados. O melhor modelo avaliado atingiu AUC de 0,89, acurácia de 81,37% e sensibilidade de 84,83% em um conjunto de nódulos pulmonares sólidos de diâmetro entre 3 e 30mm. Os resultados apontam que, dentro das especificações deste trabalho, é preferencial alimentar a rede com os primeiros cortes de um nódulo; que 5 cortes são a quantidade ideal e quais as topologias mais adequadas a cada estratégia avaliada.

Palavras-chave: Câncer de Pulmão; Auxílio ao Diagnóstico por Computador; Aprendizagem Profunda; Redes Neurais Convolucionais 3D; Otimização de Hiperparâmetros.

ABSTRACT

Cancer is a generic term for a large group of diseases characterized by an abnormal growth of cells that can invade adjoining tissues and organs. Lung cancer is the most common form of this disease and is the leading cause of cancer death, with 18.4% of the total deaths. Early diagnosis is decisive to the patient survival chances, going from 90% of survival rate for patients in early stages to 15% in advanced stages. The most common symptom of lung cancer is the pulmonary nodule. Thanks to the popularization of the computerized tomography, lung cancer diagnosis became more accurate, however, the diagnosis still is a challenging task for radiologists and subjected to factors such as radiologists tiredness and experience. That motivates developing computer-aided diagnosis tools, making diagnosis faster and more accurate. Traditional techniques work with a previous stage of extracting features and later usage of machine learning techniques for nodule classification. In the past few years, deep learning techniques received wide acceptance, as they are able to extract features in an automatic fashion. Convolutional Neural Networks are a deep learning architecture frequently used for image classification and widely applied to medical diagnosis. Nevertheless, those networks usually operate through bidimensional images, despite the computerized tomography exam provides tridimensional information. Recently, efforts have been made in developing 3D Convolutional Neural Networks, able to take advantage of the tridimensional information in its learning process. However, this field is still in its infancy for applications of this kind, requiring more studies about questions such as the way nodules are prepared for the network usage. This work proposes a methodology for evaluating a series of different methods for composing pulmonary nodules volumes for 3D Convolutional Neural Networks, using hyperparameter optimization to reach the best network topology for each method and cross validation for result validation. Our best model achieved an AUC of 0.89, accuracy of 81.37% and sensitivity of 84.43% in a set of solid pulmonary nodules with diameter between 3 and 30mm. Our results show that, under this work specifications, it is preferential to feed the network with the first slices of an nodule; that using 5 slices is the best choice and what are the fittest topologies for each evaluated strategies.

Keywords: Lung cancer; Computer-aided Diagnosis; Deep Learning; 3D Convolutional Neural Networks; Hyperparameter Optimization.

LISTA DE FIGURAS

Figura 1 – Exemplos de nódulos sólido (a), semi-sólido (b) e não sólido (c).	17
Figura 2 – Esquema básico de funcionamento de um neurônio	18
Figura 3 – Representação de um neurônio artificial	19
Figura 4 – RNA feedforward	19
Figura 5 – Exemplo de funcionamento de uma rede AP	21
Figura 6 – Exemplo de <i>feature map</i> gerada por um filtro.	22
Figura 7 – Demonstração de uma operação de max-pooling	22
Figura 8 – Exemplo de busca de hiperparâmetros.	24
Figura 9 – Espaço de busca de hiperparâmetros com limite y^*	26
Figura 10 – Distribuição de probabilidade do TPE.	27
Figura 11 – Matriz de Confusão	28
Figura 12 – Espaço ROC	30
Figura 13 – Exemplo de curva ROC para 20 instâncias	31
Figura 14 – Particionamento de uma validação cruzada 5-fold	32
Figura 15 – Esquema geral da metodologia utilizada neste trabalho.	33
Figura 16 – Processo de segmentação de uma fatia.	34
Figura 17 – Exemplos de nódulos	35
Figura 18 – Demonstração do processo de cálculo do diâmetro do nódulo.	35
Figura 19 – Distribuição da contagem de cortes.	36
Figura 20 – Estratégias de montagem de volume.	38
Figura 21 – Resumo dos diferentes modelos avaliados.	42
Figura 22 – Curvas ROC para os melhores modelos de cada estratégia.	46

LISTA DE TABELAS

Tabela 1 – Distribuição dos nódulos sólidos de 3-30mm no BNP.	36
Tabela 2 – Balanceamento e aumento de base nas diferentes etapas.	39
Tabela 3 – Espaço de busca para a arquitetura RNC1.	40
Tabela 4 – Espaço de busca para a arquitetura RNC2.	40
Tabela 5 – Topologias fornecidas pela OH para redes de arquitetura CNN1.	43
Tabela 6 – Topologias fornecidas pela OH para redes de arquitetura CNN2.	43
Tabela 7 – Resultados para a classificação usando a estratégia de primeiros cortes.	44
Tabela 8 – Resultados para a classificação usando a estratégia de cortes alternados.	44
Tabela 9 – Resultados para a classificação usando a estratégia de corte principal centralizados.	45
Tabela 10 – Modelos com melhor desempenho para cada estratégia.	45
Tabela 11 – Comparação com outros trabalhos em classificação de nódulos pulmonares.	46

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizagem de Máquina
AP	Aprendizagem Profunda
AUC	do inglês <i>Area Under the Curve</i>
BNP	Banco de Nódulos Pulmonares
CADx	do inglês <i>Computer-aided Diagnosis</i>
EI	do inglês <i>Expected Improvement</i>
GPU	do inglês <i>Graphics Processing Unit</i>
IDRI	do inglês <i>Image Database Resource Initiative</i>
LIDC	do inglês <i>Lung Image Database Consortium</i>
NPS	Nódulo Pulmonar Solitário
OH	Otimização de Hiperparâmetros
TC	Tomografia Computadorizada
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
ROC	do inglês <i>Receiver Operating Characteristics</i>
SGBD	Sistema de Gerenciamento de Banco de Dados
SMBO	do inglês <i>Sequence model-based Optimization</i>
tfp	taxa de falsos positivos
TPE	do inglês <i>Tree-structured Parzen Estimator</i>
tpv	taxa de verdadeiros positivos

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Estrutura do Trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Nódulos Pulmonares em Imagens de Tomografia Computadorizada	17
2.2	Redes Neurais Artificiais	17
2.2.1	Aprendizagem profunda	20
2.2.2	Redes Neurais Convolucionais	21
2.2.2.1	Camadas Convolucionais	21
2.2.2.2	Camadas de Pooling	22
2.2.3	Redes Neurais Convolucionais 3D	22
2.3	Otimização de Hiperparâmetros	23
2.3.1	Otimização Bayesiana	24
2.3.2	Tree-structured Parzen Estimators	25
2.4	Avaliação de Desempenho de Modelos Preditivos	27
2.4.1	Matriz de Confusão	27
2.4.2	Métricas de Desempenho	28
2.4.3	Curva ROC e AUC	29
2.4.4	Validação Cruzada	31
3	MATERIAIS E MÉTODOS	33
3.1	Banco de Dados de Nódulos Pulmonares	33
3.2	Seleção dos Nódulos	35
3.3	Pré-processamento dos dados	37
3.3.1	Estratégias de Montagem dos Volumes	37
3.3.2	Balanceamento e Aumento de Base	38
3.3.2.1	Etapa de Otimização de Hiperparâmetros	39
3.3.2.2	Etapa de Classificação	39
3.4	Otimização de Hiperparâmetros	39
3.5	Classificação	40
3.5.1	Métricas avaliadas	40
4	RESULTADOS E DISCUSSÃO	42
4.1	Topologias	42
4.2	Desempenho da Classificação	43

4.2.1	Estratégia de primeiros cortes	44
4.2.2	Estratégia de cortes alternados	44
4.2.3	Estratégia de corte principal centralizado	45
4.3	Visão geral	45
4.4	Trabalhos relacionados	46
5	CONCLUSÃO	48
	REFERÊNCIAS	49

1 INTRODUÇÃO

Câncer é um termo genérico para um grupo de doenças caracterizadas pelo crescimento desordenado de células que invadem tecidos e órgãos. É a segunda maior causa de mortes no mundo, atrás de doenças cardiovasculares, com 9,6 milhão de mortes globais estimadas em 2018 (World Health Organisation, 2019). O câncer de pulmão é a forma mais frequentemente diagnosticada desta doença, contabilizando 1,8 milhão de novos casos em 2012, ou 13% do total de casos de câncer diagnosticados naquele ano (CHEN et al., 2016). É também o tipo de câncer que mais mata, compreendendo 18,4% do total de mortes por câncer (BRAY et al., 2018).

Quando diagnosticado em seu estágio inicial, a taxa de sobrevivência em 5 anos é de favoráveis 70-90% (KNIGHT et al., 2017). No entanto, a maioria dos pacientes ainda é diagnosticada em estágios avançados da doença, reduzindo as taxas de sobrevivência em 1 ano para apenas 15-19% (BANNISTER; BROGGIO, 2016).

A tomografia computadorizada (TC) é o método mais indicado para o diagnóstico do câncer de pulmão em seus estágios iniciais (KNIGHT et al., 2017; WANG et al., 2016). Este exame produz um volume de fatias com alta resolução e contraste, permitindo uma melhor distinção de características dos tumores, tais como sua forma (esfericidade ou espiculação) e sua textura, como presença de calcificação e gordura (HUA et al., 2015).

Apesar deste avanço, a tarefa de diagnosticar um paciente ainda apresenta desafios: o radiologista deve examinar cautelosamente cada imagem dentro de um grande conjunto de imagens gerado pelo exame (KANG et al., 2017). Além disso, o fator subjetivo é muito presente nessa tarefa, o que faz com que um mesmo radiologista possa diagnosticar um mesmo exame de forma distinta em ocasiões diferentes, além da discordância entre radiologistas (CHUQUICUSMA et al., 2018; JÚNIOR et al., 2015).

As ferramentas de auxílio computadorizado ao processo de diagnóstico (do inglês Computer-Aided Diagnosis - CADx) surgem para auxiliar neste trabalho. Sistemas CADx empregam técnicas de identificação e classificação de imagens para auxiliar o diagnóstico do radiologista, tornando esta tarefa mais rápida e precisa (HUA et al., 2015; KANG et al., 2017).

Técnicas tradicionais empregadas para o diagnóstico do câncer de pulmão em sistemas CADx funcionam por meio da extração prévia de atributos descritores, como histogramas (ADETIBA; OLUGBARA, 2015) ou de atributos desenvolvidos manualmente para descrever a geometria ou textura de um nódulo pulmonar (FERREIRA et al., 2018; LUCENA et al., 2016; HAN et al., 2015). Em seguida, técnicas de Aprendizagem de Máquina (AM) são empregadas

para dar um diagnóstico baseado nos padrões encontrados nestes atributos.

Nos últimos anos houve um crescimento na aplicação de Aprendizagem Profunda (AP), técnicas de AM que se propõem a aprender representações dos dados envolvidos no problema, dispensando a etapa de extração de atributos (LITJENS et al., 2017). Uma Rede Neural Convolutiva (RNC) é uma arquitetura de AP rápida e escalável que revolucionou diversas áreas de visão computacional, como classificação de imagens (SZEGEDY et al., 2015), detecção de objetos (REN et al., 2015) e segmentação semântica (CHEN et al., 2014).

Motivado pelo sucesso das RNC no campo de classificação de imagens, foi desenvolvida uma série de trabalhos que empregam a técnica para diagnóstico médico, incluindo a classificação de nódulos pulmonares em TC (RAVÌ et al., 2017). (GINNEKEN et al., 2015) e (DING et al., 2017) utilizaram os atributos gerados por RNC para auxiliar na detecção de nódulos em imagens de TC. (KIM et al., 2016) combinou os atributos gerados por autoencoders com atributos gerados manualmente, obtendo um melhor desempenho na classificação. (KUMAR et al., 2015) usou auto-encoders e RNC para o diagnóstico de nódulos. Contudo, estes trabalhos utilizaram apenas RNC 2D, que ignoram a informação a natureza tridimensional do exame de TC.

Nos últimos anos, surgiram trabalhos que utilizam a terceira dimensão. (KANG et al., 2017) usa redes 3D multi-view para classificação binária (benignos e malignos) e ternária (benignos, malignos e malignos metastáticos) de nódulos, atingindo acurácia de 95,68% e AUC de 0.99. (ONISHI et al., 2019) emprega Redes Adversárias Geradoras para gerar novos casos e melhorar o desempenho de classificadores, atingindo acurácia de 81,7%. (ANIRUDH et al., 2016) usa esta técnica para a detecção de nódulos em exames de TC, atingindo 80% de sensibilidade.

O uso dessa arquitetura ainda está no seu início, mas os resultados apontam ganhos no desempenho das tarefas de detecção e classificação em relação às RNC 2D. No entanto, pela escassez destes trabalhos, não existe consenso sobre a forma mais adequada para construir os volumes 3D para essas redes.

O desenvolvimento de redes neurais esbarra no desafio de que são modelos muito sensíveis às suas configurações (MONTAVON et al., 2012). Trabalhos recentes sugerem que mais esforço deve ser investido nesta configuração do que tem sido feito, como forma de se obter melhores resultados (BERGSTRA et al., 2011). Estes problemas podem ser mitigado por meio de técnicas automatizadas de Otimização de Hiperparâmetros (OH), que têm apresentado resultados competitivos com a configuração manual feita por especialistas (BERGSTRA et al., 2011; BERGSTRA; BENGIO, 2012).

O objetivo principal deste trabalho foi desenvolver um modelo computacional para

classificação binária de nódulos pulmonares utilizando Redes Neurais Convolucionais 3D e otimização hiperparamétrica. Como objetivo secundário, propomos investigar o impacto de diferentes estratégias de construção de volumes no desempenho da classificação por RNC 3D.

1.1 ESTRUTURA DO TRABALHO

- **Capítulo 2 - Fundamentação Teórica:** apresenta os principais conceitos que fundamentam este trabalho;
- **Capítulo 3 - Materiais e Métodos:** descreve a metodologia empregada na preparação e execução dos experimentos;
- **Capítulo 4 - Resultados e Discussão:** apresenta e discute os resultados obtidos pelos experimentos e os compara com a literatura;
- **Capítulo 5 - Conclusão:** apresenta as conclusões obtidas pelo trabalho.

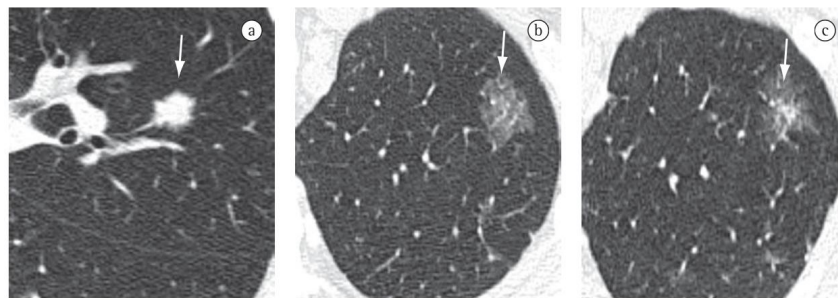
2 FUNDAMENTAÇÃO TEÓRICA

2.1 NÓDULOS PULMONARES EM IMAGENS DE TOMOGRAFIA COMPUTADORIZADA

Em imagens de Tomografia Computadorizada (TC), um nódulo pulmonar solitário (NPS) pode ser definido como uma opacidade arredondada com diâmetro menor que 3 cm, geralmente apresentando tecido de densidade mole ou calcificado (SILVA et al., 2010). Estes nódulos devem ser acompanhados, pois podem se tratar de tumores malignos (SHEN et al., 2011).

NPS podem ser classificados como sólidos (Figura 1a), quando obscurecem completamente o parênquima (tecido que constitui o pulmão); como não-sólido, quando não obscurecem as marcas vasculares e paredes brônquicas (Figura 1b) e semissólido (vidro-fosco) quando obscurece os tecidos parcialmente (Figura 1c) (SILVA et al., 2010).

Figura 1 – Exemplos de nódulos sólido (a), semi-sólido (b) e não sólido (c).



Fonte: Disponível em: (SILVA et al., 2010)

A *Fleischner Society*, sociedade internacional de radiologia torácica, propõe um protocolo para lidar com NPS, que inclui ressecções cirúrgicas, biópsias e acompanhamento com radiografias (TEAM, 2011). Com a popularização do uso de TC e sua maior sensibilidade, o número de NPS detectados em pacientes assintomáticos cresceu drasticamente, mas apenas uma pequena fração são tumores malignos (OST et al., 2003). Isto torna necessário que os NPS sejam prontamente classificados quanto a sua malignância por meio da TC, poupando o paciente de biópsias e tratamentos invasivos (TEAM, 2011).

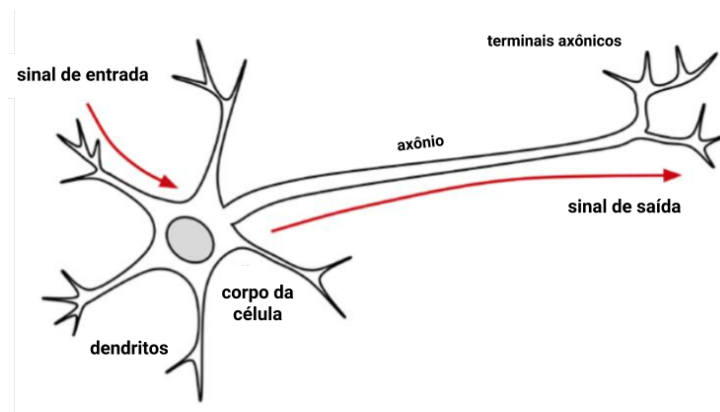
2.2 REDES NEURAIIS ARTIFICIAIS

Redes Neurais Artificiais (RNA) são modelos computacionais para predição ou regressão inspirados pelas redes neurais biológicas. O poder destas redes vem de sua capacidade de

relacionar suas variáveis de entrada às saídas que devem ser previstas. Neste sentido, pode-se dizer que uma RNA aprende mapeamentos entre estes conjuntos de variáveis. Foi matematicamente provado que uma RNA com uma camada oculta é capaz de aproximar uma grande quantidade de funções contínuas (SONODA; MURATA, 2017).

O elemento básico de uma RNA é o neurônio, em alusão às células nervosas. A figura 2 ilustra o funcionamento de um neurônio biológico. Sinais de entrada são recebidos pelos dendritos da célula por meio de um processo bioquímico. Este processo permite que seja atribuído um peso ao sinal de acordo com sua relevância ou frequência. Conforme o corpo da célula acumula os sinais de entrada, é atingido um limite no qual a célula dispara um sinal eletroquímico pelo axônio, que é convertido em um sinal químico e passado aos neurônios vizinhos por meio das sinapses. (LANTZ, 2013)

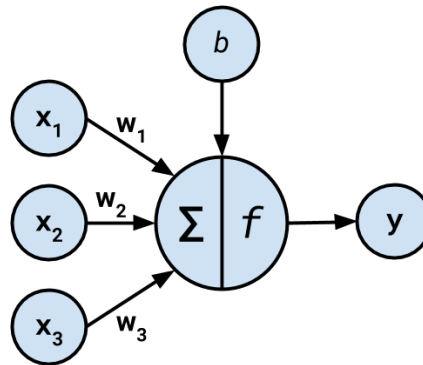
Figura 2 – Esquema básico de funcionamento de um neurônio



Fonte: (LANTZ, 2013)

O funcionamento de um neurônio artificial remete ao biológico. A Figura 3 ilustra os elementos desta estrutura. O neurônio mapeia um relacionamento entre os sinais de entrada (variáveis x) e um sinal de saída (variável y). Assim como no neurônio biológico, cada sinal de entrada tem um peso associado (valores w). Os sinais de entrada são somados e a esta soma é aplicada uma função de ativação denotada por f . Por fim, permite-se que um viés b seja somado ao resultado.

Figura 3 – Representação de um neurônio artificial



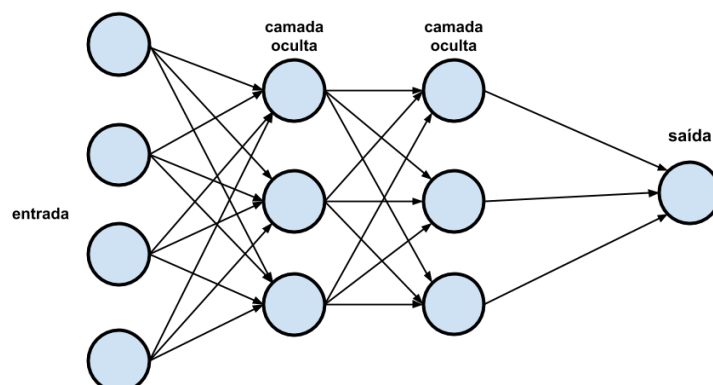
Um neurônio artificial típico com n entradas pode ser representado pela equação 1. Os pesos w permitem que cada entrada x contribua de forma diferenciada com a soma dos sinais de entrada. O somatório dos sinais é somado a um viés b e este valor é dado como entrada para a função de ativação $f(x)$, gerando um sinal resultante $y(x)$ como saída do neurônio.

$$y(x) = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (1)$$

Estes neurônios podem ser usados para construir redes capazes de mapear funções complexas. Normalmente são organizados em *camadas* e as saídas dos neurônios de uma camada são dadas como entrada para os neurônios das camadas seguintes. Questões como o número de neurônios, número de camadas e direção de propagação da informação dizem respeito à *topologia da rede*. Uma topologia em camadas na qual os dados são propagados da camada de entrada para a de saída é chamada *feedforward*. A Figura 4 ilustra uma rede do tipo (FRIEDMAN et al., 2001).

Uma vez definida a topologia da rede, esta deve ser treinada para desempenhar sua

Figura 4 – RNA feedforward



tarefa. O treino tem como função ajustar os pesos da rede de forma a refletir os padrões que são observados durante esta etapa. É uma tarefa exigente computacionalmente, o que fez com que as RNA raramente fossem usadas até os anos 80, quando foi desenvolvida uma forma eficiente de treinar RNAs: o algoritmo **backpropagation** (RUMELHART et al., 1988).

De forma geral, o algoritmo funciona em ciclos de dois processos. Cada ciclo é chamado **época**. Os processos são (LANTZ, 2013; FRIEDMAN et al., 2001):

- **Fase *forward***: os neurônios são ativados sequencialmente da camada de entrada para a camada de saída, aplicando os pesos e funções de ativação de cada neurônio. Ao chegar na última camada, é produzido um sinal de saída.
- **Fase *backward***: o sinal produzido na etapa anterior é comparado ao objetivo. A diferença entre estes valores dá-se o nome de **erro**. Este erro é usado para ajustar os pesos da rede e reduzir os valores de erro nas próximas épocas.

Para definir como os pesos devem ser ajustados, usa-se o método do **gradiente descendente** (GOODFELLOW et al., 2016). Esta técnica usa a derivada das funções de ativação dos neurônios para determinar como o valor de um peso deve ser ajustado de forma a reduzir o erro. A quantidade modificada é denominada **taxa de aprendizado**.

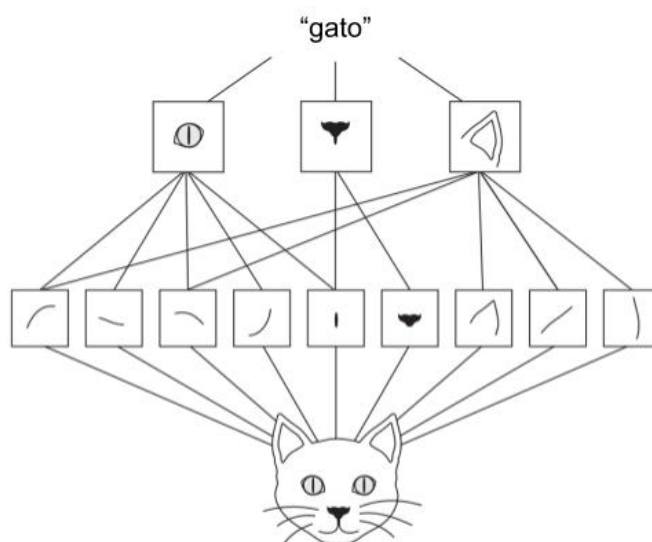
2.2.1 Aprendizagem profunda

Durante décadas, desenvolver sistemas de AM exigia uma etapa prévia de processamento das entradas em representações adequadas para que o sistema de aprendizado pudesse extrair padrões úteis para a tarefa. Esta etapa exigia um trabalho cuidadoso e um conhecimento considerável do domínio da aplicação (LITJENS et al., 2017).

Aprendizagem profunda (AP) é um conjunto de métodos para aprendizagem de representações a partir dos dados puros (como uma imagem ou vídeo). Técnicas de AP funcionam compondo transformações não-lineares simples de forma hierárquica, chegando a representações mais abstratas dos dados (LECUN et al., 2015).

A Figura 5 ilustra de forma simples o conceito de aprendizagem de representações. Na primeira camada desta rede, verifica-se a existência ou ausência de formas simples na imagem, como linhas e cantos. Estas formas simples compõem formas mais complexas, como os olhos, orelhas ou focinho. A presença destas três formas, por fim, indica que a imagem se trata de um gato.

Figura 5 – Exemplo de funcionamento de uma rede AP



Fonte: Disponível em: (CHOLLET, 2018)

Alguns fatores possibilitaram a ampla adoção de AP na última década, dentre eles a popularização de Unidades de Processamento Gráfico (do inglês *Graphics Processing Unit - GPU*), bem como o menor custo de hardware e os avanços recentes em AM e processamento de sinais (DENG, 2014). GPUs, inicialmente pensadas para processamento gráfico intenso, são muito eficientes em operações com matrizes e vetores, o que as permite acelerar o treinamento de RNAs em fatores de 50 vezes ou mais (SCHMIDHUBER, 2015).

2.2.2 Redes Neurais Convolucionais

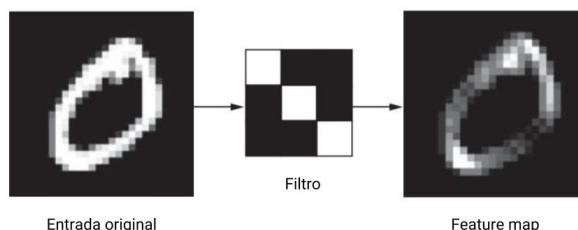
Redes Neurais Convolucionais (RNC) são uma classe de redes RNA para AP usada para processar dados em formatos de vetores, como vetores 1D para análise de texto, 2D para imagens e 3D para vídeos. Seu uso mais comum é para processamento de imagens. A arquitetura de uma RNC comum é composta por dois tipos de camadas: as *convolucionais* e de *pooling*.

2.2.2.1 Camadas Convolucionais

O elemento básico para a operação de convolução são os *filtros*. Assim como os neurônios, filtros também possuem recebem entradas e geram uma saída. Na camada de convolução, cada filtro é aplicado sobre a entrada e movido um pixel por vez, montando o chamado *feature map*. Estes *feature maps* servem de entrada para as camadas seguintes da rede, que podem inclusive aplicar novas convoluções sobre estes dados.

A figura 6 ilustra o mapa gerado pela aplicação de um filtro sobre uma imagem. Nesse caso em particular, o filtro detecta o padrão de *pixels* brancos em diagonal e o *feature map* reflete a ocorrência desse padrão na imagem.

Figura 6 – Exemplo de *feature map* gerada por um filtro.

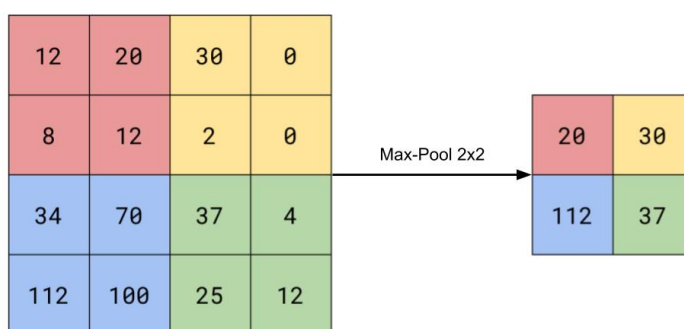


Fonte: Disponível em: (CHOLLET, 2018)

2.2.2.2 Camadas de Pooling

A operação de *pooling* é similar à de convolução, no sentido em que janelas são aplicadas à entrada e é gerada um *feature map* contendo o resultado. A diferença é que as operações de *pooling* têm por objetivo reduzir a dimensão de um *feature map*. Um tipo comum desta operação é o *max-pooling*, que seleciona o maior valor dentro de sua janela para constituir a saída. A figura 7 demonstra a operação do *max-pooling* com uma janela de dimensão 2x2 sobre a entrada.

Figura 7 – Demonstração de uma operação de *max-pooling*



Fonte: Disponível em: <https://computersciencewiki.org/index.php/Max-pooling/_Pooling> . Acesso em 16/01/2019.

2.2.3 Redes Neurais Convolucionais 3D

As RNC 3D aplicam as mesmas ideias das RNC 2D. A diferença é que agora as operações de convolução e *max-pooling* operam sobre volumes. De maneira similar, uma operação de

convolução vai buscar padrões de três dimensões nas entradas, gerando *feature maps* também tridimensionais.

2.3 OTIMIZAÇÃO DE HIPERPARÂMETROS

Diversas técnicas de AM podem ser configuradas por hiperparâmetros, que, diferente dos parâmetros do modelo (como os pesos de uma RNA), são definidos antes do treino. Estes hiperparâmetros podem modificar diversos aspectos do algoritmo de aprendizagem e ter grande impacto no modelo resultante e em seu desempenho (CLAESEN; MOOR, 2015).

O objetivo da otimização de hiperparâmetros (OH) é encontrar o conjunto de hiperparâmetros para um algoritmo de AM que leva a um melhor desempenho em um conjunto de validação (BERGSTRA et al., 2011). OH pode ser formalizada pela equação 2:

$$x^* = \underset{x \in X}{\operatorname{arg\,min}} f(x) \quad (2)$$

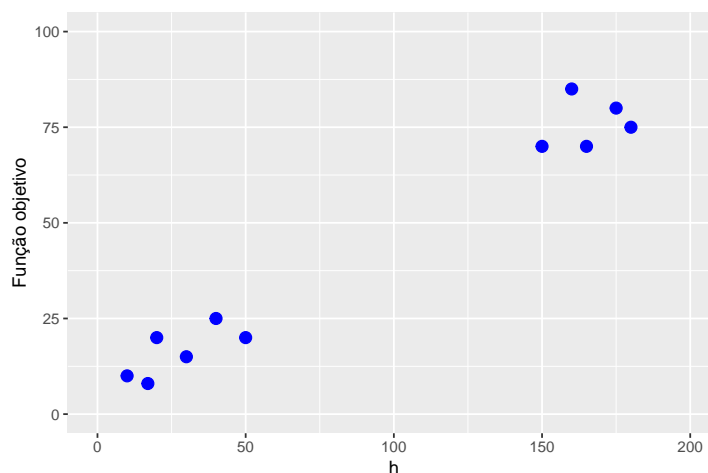
Onde $f(x)$ é a função objetivo cujo valor deve ser minimizado. Exemplos de função objetivo são a taxa de erro para modelos de classificação ou erro quadrático médio em modelos de regressão. x^* é o conjunto de valores de hiperparâmetros que leva ao valor mínimo da função objetivo $f(x)$ e x pode assumir qualquer valor em um espaço de hiperparâmetros X .

As estratégias tradicionais de OH são a busca manual, onde o pesquisador define e avalia manualmente as combinações de hiperparâmetros, guiado por sua intuição e experiência; a busca em grade, onde um conjunto de valores é definido para cada hiperparâmetro e todas as combinações destes parâmetros são avaliadas automaticamente e a busca aleatória, na qual as combinações de hiperparâmetros são escolhidas aleatoriamente dentro de um domínio (BERGSTRA; BENGIO, 2012).

A busca manual exige do pesquisador um esforço que poderia ser melhor investido em outras tarefas da modelagem. A busca em grade e aleatória resolvem este problema ao escolher e avaliar de forma automática combinações de hiperparâmetros. A busca aleatória leva vantagem sobre a busca em grade por avaliar hiperparâmetros em todo o domínio, e não apenas em valores predefinidos (BERGSTRA; BENGIO, 2012).

Contudo, estas técnicas deixam a desejar nos quesitos de reprodutibilidade e são impraticáveis quando o número de hiperparâmetros é grande. A figura 8 ajuda a entender outro problema destas técnicas.

Figura 8 – Exemplo de busca de hiperparâmetros.



No gráfico temos o resultado obtido pela função objetivo para diferentes valores de um hiperparâmetro h . Como deseja-se minimizar a função objetivo, logo percebe-se que a provável solução está nos valores de h menores que 50. As técnicas automáticas são incapazes de tirar proveito desse padrão, avaliando igualmente valores por todo o domínio. Isto é especialmente problemático porque a avaliação da função objetivo $f(x)$ é, com frequência, uma operação custosa, como é o caso das RNA.

2.3.1 Otimização Bayesiana

Diferente das técnicas de busca em grade e aleatória, a otimização Bayesiana usa o histórico de avaliações para criar um modelo probabilístico da função objetivo em função dos hiperparâmetros. Este modelo, chamado de substituto para a função objetivo, é representado por $p(y|x)$, denotando a probabilidade de se atingir um valor y para função objetivo dado um conjunto de hiperparâmetros x . A otimização Bayesiana usa o modelo substituto para escolher o melhor conjunto de hiperparâmetros, reduzindo a necessidade de avaliar a função objetivo verdadeira (BERGSTRA et al., 2011).

A otimização sequencial baseada em modelo (do inglês Sequence model-based Optimization - SMBO) é uma formalização da otimização Bayesiana. O algoritmo de uma técnica

genérica de SMBO é apresentado em 1:

Algoritmo 1: Pseudocódigo de uma técnica SMBO genérica.

Entrada: f, M_0, T, S

Saída: \mathcal{H}

$\mathcal{H} \leftarrow \emptyset$;

para $t \in T$ **faça**

$x^* \leftarrow \arg \min_x S(x, M_{t-1})$;

Avaliar $f(x^*)$;

$\mathcal{H} \leftarrow \mathcal{H} \cup (x^*, f(x^*))$;

Gerar um novo modelo M_t para \mathcal{H} ;

fim

Fonte: (BERGSTRA et al., 2011)

O algoritmo consiste em buscar o conjunto de hiperparâmetros x^* que minimize a função substituta S . Em seguida, avalia-se a função objetivo verdadeira f para x^* . O histórico \mathcal{H} é atualizado com os hiperparâmetros x^* e o score $f(x^*)$. Por fim, é gerado um modelo M_t para o histórico atualizado \mathcal{H} . Este processo é repetido T vezes, retornando o histórico da busca (BERGSTRA et al., 2011).

As diferentes SMBO variam, normalmente, a forma como a função substituta é modelada e a **função de seleção** utilizada para escolher os hiperparâmetros. A função de seleção mais comum é a *Expected Improvement* (EI), que tem equação:

$$EI_{y^x}(x) = \int_{-\infty}^{\infty} \max(y^* - y, 0) p_M(y|x) dy \quad (3)$$

Onde y^* é um valor observado da função objetivo, x é o conjunto de hiperparâmetros proposto, y é o valor da função objetivo para x e $p_M(y|x)$ é o modelo substituto, que denota a probabilidade de y dado x . De forma geral, EI é a expectativa, dado um modelo M de f , de que $f(x)$ vai exceder negativamente uma observação anterior y^* .

2.3.2 Tree-structured Parzen Estimators

A abordagem Tree-structured Parzen Estimator (TPE) é uma forma proposta por (BERGSTRA et al., 2011) de construir o modelo substituto. TPE utiliza o Teorema de Bayes no lugar de

$p(y|x)$:

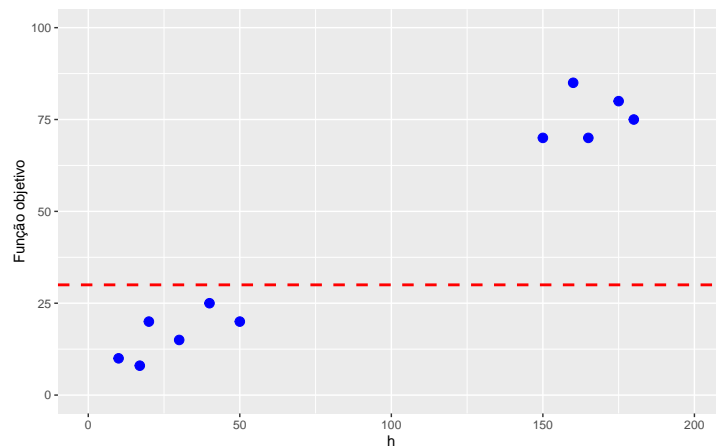
$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (4)$$

Por sua vez, o termo $p(x|y)$, que expressa a probabilidade dos hiperparâmetros x dado um score y é expresso por:

$$p(x|y) = \begin{cases} l(x) & \text{se } y < y^* \\ g(x) & \text{se } y \geq y^* \end{cases} \quad (5)$$

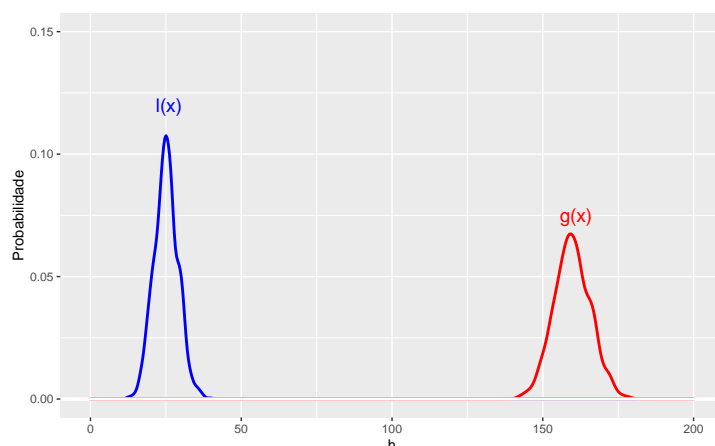
Onde $l(x)$ é a distribuição de probabilidade dos hiperparâmetros onde a função objetivo é menor que o limite e $g(x)$ onde a função objetivo é maior que o limite. Por exemplo, se traçarmos uma linha em nosso gráfico para y^* , como na figura 9

Figura 9 – Espaço de busca de hiperparâmetros com limite y^* .



E criarmos duas distribuições de probabilidade para o nosso hiperparâmetro h , uma usando os parâmetros que levaram a valores abaixo do limite e outra que levaram a valores superiores, temos algo como a figura 10:

Figura 10 – Distribuição de probabilidade do TPE.



Isto faz com que TPE avalie mais frequentemente valores abaixo do limite y^* , onde espera atingir os valores mínimos da função objetivo. No contexto do SMBO, $l(x)$ e $g(x)$ são construídos com base no histórico \mathcal{H} , melhorando as estimativas do modelo substituto.

Foi demonstrado que a técnica TPE atinge taxas de erros menores que a busca manual e aleatória (BERGSTRA et al., 2011). A biblioteca de otimização Hyperopt (BERGSTRA et al., 2013) contém uma implementação do TPE para python, e é usada pela biblioteca Hyperas (PUMPERLA, 2019) para simplificar a sua aplicação para otimização de RNA.

2.4 AVALIAÇÃO DE DESEMPENHO DE MODELOS PREDITIVOS

Uma tarefa essencial ao desenvolvimento de modelos preditivos é a avaliação de seu desempenho. Esta avaliação deve ser capaz de fornecer indicativos do comportamento do modelo em diferentes aspectos de uma tarefa e como este modelo se compara a outros (POWERS, 2011).

2.4.1 Matriz de Confusão

Dada uma tarefa de classificação binária, isto é, o modelo deve ser capaz de distinguir casos de duas categorias distintas, é convenção resumir os resultados no formato de uma tabela de contingência: a **matriz de confusão**. A figura 11 apresenta uma matriz de confusão.

Figura 11 – Matriz de Confusão

		Classe predita	
		Positivo	Negativo
Classe real	Positivo	Verdadeiros positivos (VP)	Falsos negativos (FN)
	Negativo	Falsos positivos (FP)	Verdadeiros negativos (VN)

As colunas da matriz indicam os valores preditos pelo classificador, enquanto as linhas dizem respeito aos valores reais das classes. As células da tabela contém a interseção entre estes conjuntos, podendo ser apresentada a contagem de casos ou a proporção. Esta matriz é especialmente útil para ter uma visão do tipo de acerto e erro que um classificador comete, contendo:

- **Verdadeiro positivo (VP):** a amostra era positiva e foi classificada como tal;
- **Falso negativo (FN):** a amostra era positiva e foi classificada como negativa;
- **Falso Positivo (FP):** a amostra era negativa e foi classificada como positiva;
- **Verdadeiro negativo (VN):** a amostra era negativa e foi classificada como tal.

2.4.2 Métricas de Desempenho

A matriz de confusão permite elaborar métricas que descrevam de forma sucinta características do modelo classificador.

$$Acurácia = \frac{VP + VN}{VP + FN + FP + VN} \quad (6)$$

$$Sensibilidade = cobertura = \frac{VP}{VP + FN} \quad (7)$$

$$Especificidade = \frac{VN}{VN + FP} \quad (8)$$

$$Precisão = \frac{VP}{VP + FP} \quad (9)$$

$$F_1score = \frac{2 * precisão * cobertura}{precisão + cobertura} \quad (10)$$

- **Acurácia:** taxa de acerto geral do classificador.
- **Sensibilidade ou cobertura:** taxa de acerto de casos positivos do classificador. Indica quão bem o classificador identifica casos positivos.
- **Especificidade:** taxa de acerto de casos negativos do classificador. Indica o quão bem o classificador identifica casos negativos.
- **Precisão:** proporção dos casos classificados como positivos que realmente são positivos. É uma medida da confiabilidade do preditor para classificações positivas.
- **F1 score:** média harmônica da precisão e cobertura.

2.4.3 Curva ROC e AUC

O gráfico de Característica de Operação do Receptor (do inglês Receiver Operating Characteristics - ROC) é uma técnica para visualizar, organizar e escolher classificadores com base em seu desempenho (FAWCETT, 2006). A comunidade médica tem uma extensa literatura sobre o uso de curvas ROC para analisar o comportamento de sistemas de auxílio ao diagnóstico (ZOU, 2019). Nos últimos anos, curvas ROC têm sido cada vez mais usadas em AP, pois métricas simples como acurácia podem ser insuficientes na avaliação de um modelo (POWERS, 2011).

Consideremos um problema de classificação binária, ou seja, as instâncias devem ser classificadas em positivo ou negativo. Um modelo de classificação mapeia instâncias a classes. Alguns classificadores fazem esse mapeamento de forma discreta, ou seja, cada instância é mapeada a uma classe. RNA são capazes de produzir uma saída contínua, como uma probabilidade ou confiança de que uma instância pertence a uma classe. Isso permite que estabeleçamos um limite nesse valor a partir do qual as classes são diferenciadas (FAWCETT, 2006).

Em um problema de classificação binária, temos 4 saídas possíveis, como mostra a figura 11. Podemos definir os valores de **taxa de verdadeiros positivos** (tvp) (equação 11), cujo valor é o mesmo da sensibilidade ou cobertura, e a **taxa de falsos positivos** (tfp) (equação 12).

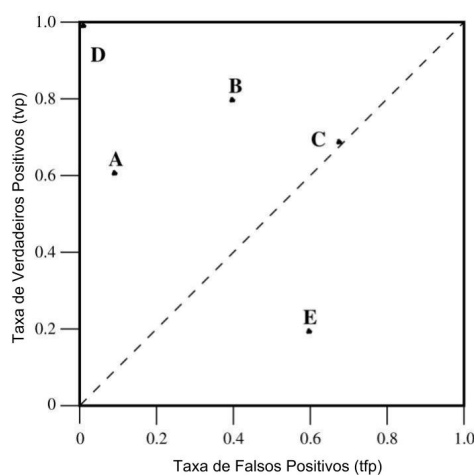
$$\text{taxa de VP (tvp)} = \frac{\text{positivos corretamente classificados}}{\text{Todos os positivos}} \quad (11)$$

$$\text{taxa de FP (tfp)} = \frac{\text{negativos incorretamente classificados}}{\text{Todos os negativos}} \quad (12)$$

O **espaço ROC** é um gráfico de duas posições onde a taxa de fp é o eixo X e a taxa de vp é o eixo Y. A figura 12 mostra um espaço do tipo contendo 5 instâncias de classificadores,

nomeados de A a E. Estes classificadores são classificadores discretos, atribuindo cada instância a apenas uma classe.

Figura 12 – Espaço ROC



Fonte: (FAWCETT, 2006)

Alguns pontos são dignos de nota. O ponto (0,0) é um modelo que nunca classifica uma instância como negativa, logo, tem ambas tvp e tfp nulas. A estratégia oposta se situa no ponto (1,1), onde todas as classes são classificadas como positivo. O ponto D (0,1) representa o classificador perfeito (POWERS, 2011).

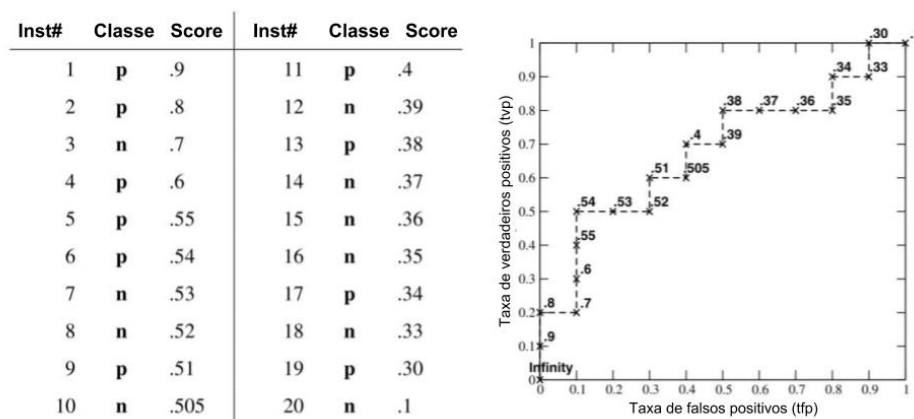
A linha tracejada $y = x$ representa um classificador aleatório. Por exemplo, se a chance de classificar uma instância como positiva for de 50%, obtemos o ponto (0.5, 0.5) no espaço. Se a chance for de 90%, a tvp é de 0.9, mas a tfp cresce na mesma proporção, levando ao ponto (0.9, 0.9), e assim por diante. O ponto C tem o desempenho condizente com esse tipo de classificador (POWERS, 2011).

De forma geral, um classificador é melhor que outro caso se encontre a noroeste deste. Classificadores mais à esquerda são "conservadores", fazendo classificações positivas apenas com evidências suficientes, cometendo menos falsos positivos (mas geralmente com poucos verdadeiros positivos). Classificadores à direita são "liberais", classificando positivos com poucas evidências, logo, com altas tvp e tfp (FAWCETT, 2006).

Em um classificador probabilístico, ou seja, cuja saída é uma probabilidade de uma entrada corresponder a uma classe, podemos variar o limite a partir do qual uma instância é classificada como positiva, chegando a um conjunto de pontos no espaço ROC que ilustra o comportamento do classificador na distinção das classes em diferentes configurações (FAWCETT, 2006). Estes pontos formam a *curva ROC*. A figura 13 ilustra esse procedimento para 20

instâncias.

Figura 13 – Exemplo de curva ROC para 20 instâncias



Fonte: (FAWCETT, 2006)

Essa característica faz da curva ROC uma boa forma de visualizar e comparar o desempenho de modelos preditivos, sobretudo os probabilísticos. Outra vantagem é que o gráfico permite o ajuste dos limites de classificação de forma a se chegar a um modelo com tvp e tnp adequadas ao problema (FAWCETT, 2006).

2.4.4 Validação Cruzada

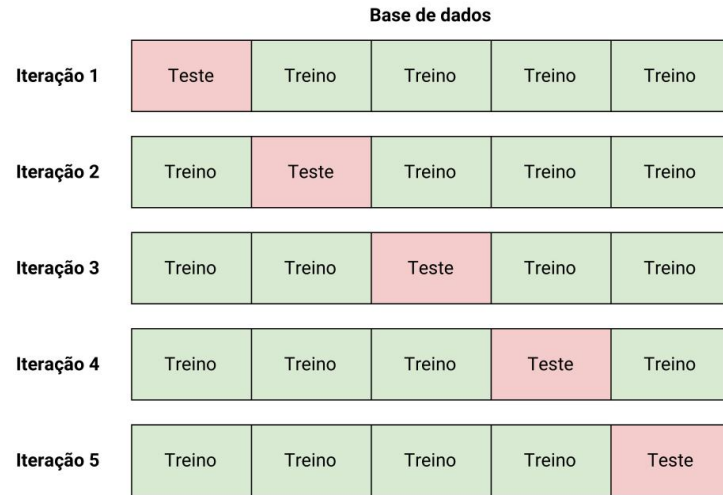
A avaliação de um modelo preditivo tem duas etapas: na de treino, o modelo usa os dados para encontrar padrões relevantes à classificação; na de avaliação, verifica-se o desempenho do modelo treinado. Sabe-se que treinar e avaliar um modelo no mesmo conjunto de dados leva a resultados muito otimistas. A validação cruzada soluciona esse problema baseando-se no princípio de que avaliar o modelo com dados não vistos leva a uma boa estimativa da capacidade real desse modelo (ARLOT et al., 2010).

Na maioria das aplicações em saúde tem-se uma quantidade limitada de dados. É comum dividir parte dos dados para a etapa de treino (base de treino) e outra parte para a avaliação (base de teste). Um particionamento dos dados leva a uma estimativa do desempenho de um modelo. Efetuar múltiplos particionamentos é chamado de *validação cruzada* (ARLOT et al., 2010).

Uma forma popular desse de validação é a validação cruzada k -fold. Neste procedimento, os dados são divididos em k partições de tamanho aproximadamente igual e cada partição é usada para testar o modelo, enquanto as $k - 1$ restantes são usadas para o treino do modelo. As métricas de desempenho de cada iteração são medidas e usa-se a média destas para estimar o

desempenho do modelo (WONG, 2015). A figura 14 ilustra o procedimento para 5 partições.

Figura 14 – Particionamento de uma validação cruzada 5-fold

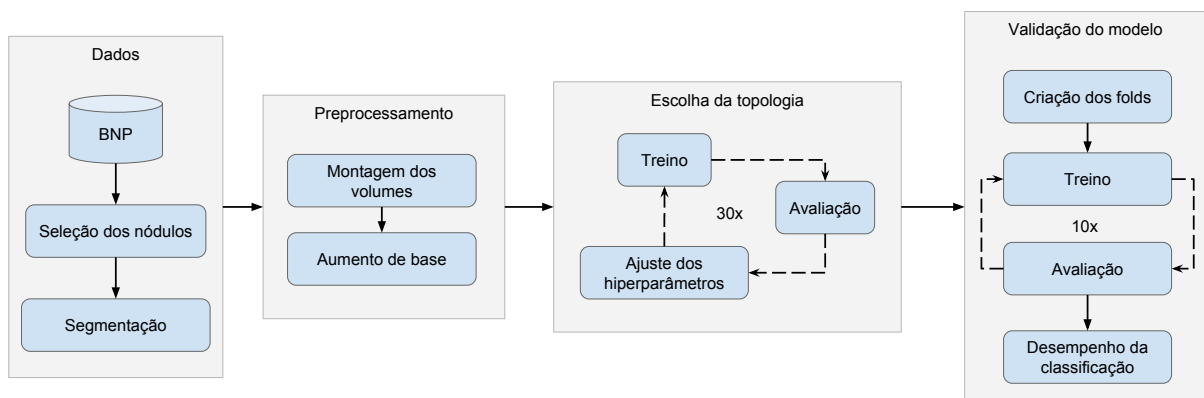


Esta estimativa está sujeita à aleatoriedade da escolha dos elementos que compõem as partições, o que faz com que este valor esteja sujeito a um viés e variância. O viés é a distância entre o resultado observado e o resultado real. A variância é o quanto esta estimativa varia durante as iterações.

3 MATERIAIS E MÉTODOS

A figura 15 apresenta uma visão geral da metodologia proposta. Neste trabalho foi utilizado um banco de dados de nódulos pulmonares desenvolvido em nosso laboratório (JUNIOR et al., 2016). As imagens foram selecionadas e segmentadas (seção 3.2) para a montagem dos volumes, empregando diferentes estratégias (seção 3.3.1). O modelo foi escolhido por meio de OH (seção 3.4) e os resultados foram validados por meio de validação cruzada (seção 3.5).

Figura 15 – Esquema geral da metodologia utilizada neste trabalho.



Todos os experimentos deste trabalho foram realizados em uma GPU Nvidia Titan X com 12 GB de RAM, 3.584 núcleos, velocidade de 1,5 GHz e arquitetura Pascal. Os frameworks utilizados foram: Hyperas (PUMPERLA, 2019) (versão 0.4) com o Hyperopt (BERGSTRA et al., 2013) (versão 0.1.1) como backend e Keras (CHOLLET et al., 2017) (versão 2.2.4) com o Tensorflow (ABADI et al., 2016) (versão 1.11.0).

3.1 BANCO DE DADOS DE NÓDULOS PULMONARES

Neste trabalho foi utilizado o repositório público LIDC-IDRI (ARMATO et al., 2011), atualmente contendo 1.018 exames de 1010 pacientes e 244.529 imagens de TC de tórax. As lesões foram identificadas e classificadas por quatro radiologistas experientes, que avaliaram os nódulos acerca de algumas características subjetivas, dentre as quais a probabilidade de malignância, em uma escala de 1 a 5:

- **Malignância 1:** alta probabilidade de ser benigno;
- **Malignância 2:** probabilidade moderada de ser benigno;
- **Malignância 3:** malignância indefinida;

- **Malignância 4:** probabilidade moderada de ser maligno;
- **Malignância 5:** alta probabilidade de ser maligno.

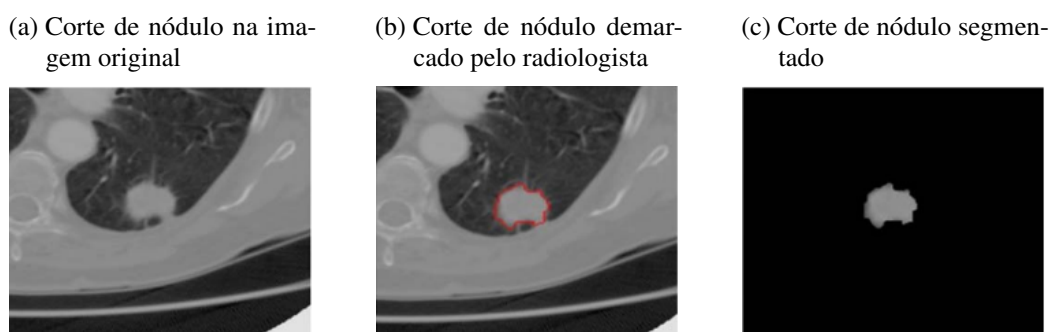
Apesar de ser uma das bases de dados mais completas em exames de TC disponíveis (ARMATO et al., 2011), o LIDC-IDRI é organizado em uma série de arquivos em formatos distintos e não relacionados, o que dificulta a manipulação das informações contidas na base. Tendo em vista este problema, optamos por utilizar o Banco de Nódulos Pulmonares (BNP) desenvolvido em nosso laboratório por (JUNIOR et al., 2016).

O BNP foi desenvolvido em MongoDB, um Sistema de Gerenciamento de Banco de Dados (SGBD) de abordagem não relacional (do inglês Not only Structured Query Language - NoSQL) orientado a documentos. Os exames e imagens foram armazenados de forma organizada e de fácil acesso. (JUNIOR et al., 2016) escolheu as anotações de apenas um dos quatro radiologistas do LIDC-IDRI para evitar redundâncias. Foram selecionados os nódulos do radiologista que identificou o maior número de lesões (JUNIOR et al., 2016).

Atualmente o BNP contém 752 exames e 1.944 nódulos segmentados corte a corte no formato PNG. As imagens originais de TC do tórax foram armazenadas no padrão de imagens médicas DICOM. De maneira a uniformizar as imagens da base do BNP, aplicamos o janelamento de escalas de cinza do pulmão fixando a janela em 1600 e o nível em -600.

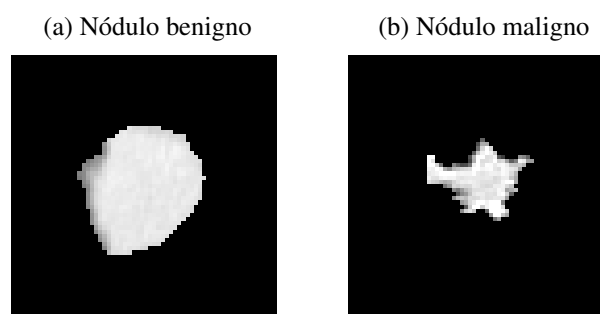
Os nódulos foram segmentados de acordo com as coordenadas das marcação feita pelo radiologista no LIDC-IDRI e em seguida as imagens foram cortadas na dimensão 64x64 pixel e armazenadas em disco. O processo de segmentação é apresentado na figura 16. A figura 16a mostra a imagem original; a figura 16b traz a demarcação feita pelo radiologista em vermelho; a figura 16c mostra o resultado da segmentação.

Figura 16 – Processo de segmentação de uma fatia.



Para os fins deste trabalho, os nódulos com malignância 3 foram desconsiderados, restando 1.171 nódulos. Os nódulos com malignância 1 e 2 foram considerados benignos e os nódulos com malignância 4 e 5 foram considerados malignos. A figura 17 mostra exemplos de nódulos benigno (17a) e maligno (17b).

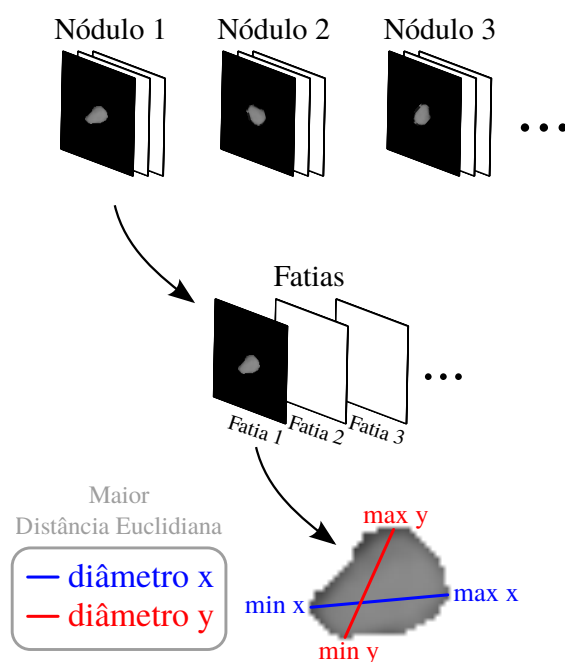
Figura 17 – Exemplos de nódulos



3.2 SELEÇÃO DOS NÓDULOS

Os nódulos presentes no BNP tiveram suas dimensões medidas por Lima Filho (FILHO et al., 2016). O cálculo foi feito medindo as distâncias euclidianas entre as coordenadas dos pontos mínimo e máximo ao longo dos eixos x e y de todos os cortes de cada nódulo (Figura 18). A distância de maior valor foi armazenada no BNP.

Figura 18 – Demonstração do processo de cálculo do diâmetro do nódulo.



Fonte: Imagem extraída de (FILHO et al., 2016).

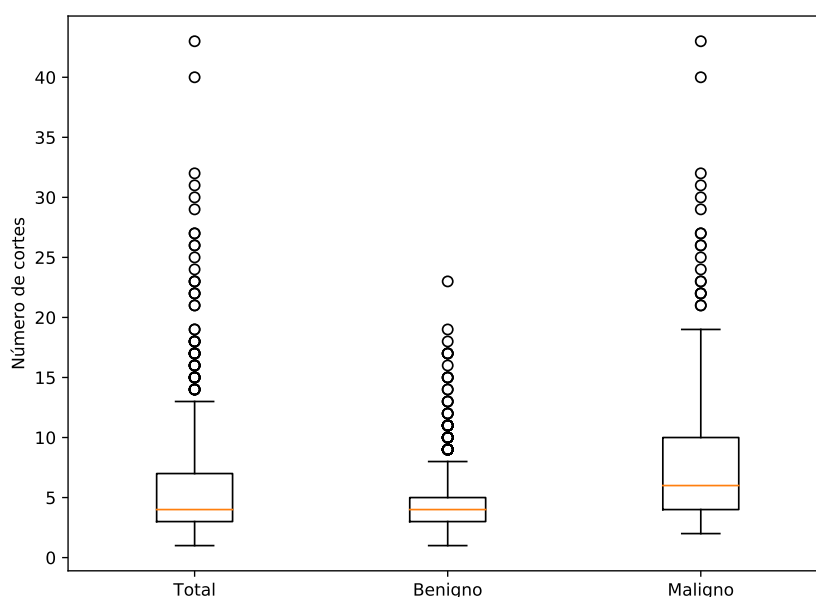
Um nódulo pulmonar é definido como uma opacidade focal arredondada de diâmetro entre 3mm e 30mm (SILVA et al., 2010). Para este trabalho foram considerados os nódulos destas dimensões. Além disso, foram selecionados apenas os nódulos sólidos, pois sua demarcação é feita com maior precisão pelos radiologistas. Ao fim desta etapa de seleção, obtivemos um total de 1006 nódulos, distribuídos como mostrado na Tabela 1.

Tabela 1 – Distribuição dos nódulos sólidos de 3-30mm no BNP.

Probabilidade de malignância	Benignos		Malignos		Total
	1	2	4	5	
Número de nódulos	304	394	171	137	1.006
Por categoria	698		308		

Estes 1006 nódulos têm, ao todo, 6.633 cortes. O número de cortes pode variar de acordo com seu comprimento no plano perpendicular às imagens de TC. A figura 19 mostra a distribuição do número de cortes para todos os nódulos e para nódulos benignos e malignos. Os nódulos benignos têm em média menos cortes que os malignos, com medianas de 4 e 6 cortes, respectivamente. Os benignos também têm cortes com menor variância, apresentando um desvio padrão de 2,9 cortes, contra 6,2 dos malignos. O conjunto de todos os nódulos tem mediana de 4 cortes e desvio padrão de 4,539 cortes. Esta variação no número de cortes é um empecilho para o treinamento de RNC, pois estes modelos precisam que suas entradas tenham formato conhecido. Isto motivou o estudo de diferentes métodos de composição de volumes, descritos na próxima seção.

Figura 19 – Distribuição da contagem de cortes.



3.3 PRÉ-PROCESSAMENTO DOS DADOS

Nesta seção descrevemos como os dados foram preparados para as tarefas de OH e Classificação. Os dois passos foram a montagem dos volumes utilizando algumas estratégias e o balaceamento e aumento da base por meio de rotações dos volumes:

3.3.1 Estratégias de Montagem dos Volumes

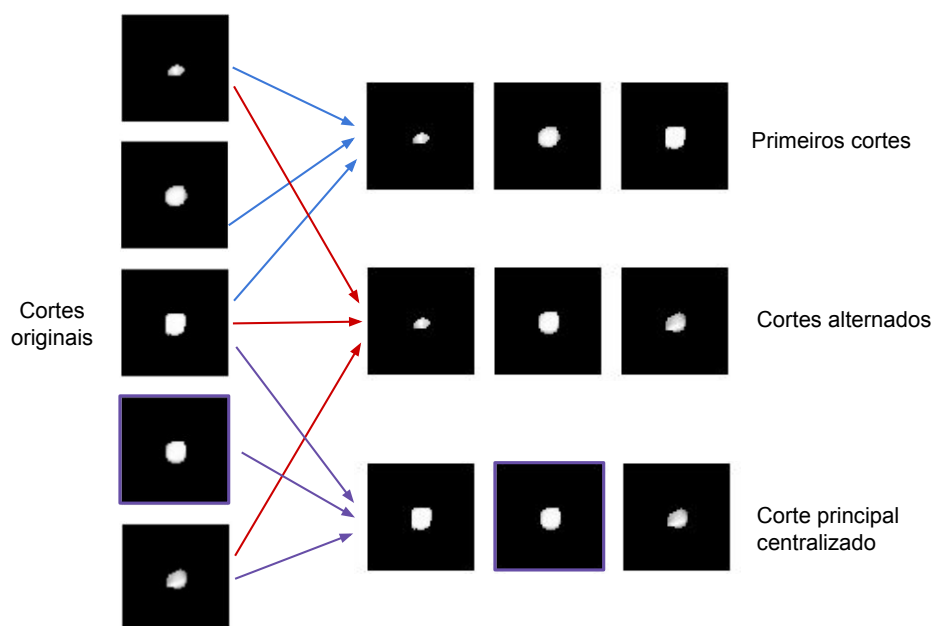
RNAs, assim como outras técnicas de AM, precisam que suas entradas sejam fornecidas em um formato conhecido (CHOLLET, 2018). Isto significa que uma RNC que lide com imagens deve receber estes dados em dimensões de altura e largura predeterminadas. O mesmo se aplica aos volumes em RNC 3D, onde além das dimensões de altura e largura, deve ter sua profundidade em um formato determinado.

No caso específico deste trabalho, como foi mostrado na Figura 19, temos um número variável de cortes por nódulo. Para tratar este problema empregamos algumas estratégias para normalizar o número de cortes de cada nódulo de forma a criar volumes de dimensões uniformes:

- **Primeiros cortes:** consiste em selecionar k cortes de um nódulo com n cortes na ordem em que estão dispostos.
- **Cortes alternados:** para selecionar k cortes em um nódulo com n cortes, são selecionados os cortes 1, corte n e $1 + \frac{n-1}{k-1} \cdot i$. Onde $i = \{1, 2, \dots, k-2\}$ e $\lceil \cdot \rceil$ é a função de arredondamento inteiro. Este cálculo fornece um volume onde os cortes são alternados. Esta estratégia é usada em (KANG et al., 2017).
- **Corte principal centralizado:** o corte de maior diâmetro, como calculado na figura 18, é selecionado, junto com os $\lceil \frac{k-1}{2} \rceil$ cortes anteriores e os $\lfloor \frac{k-1}{2} \rfloor$ cortes posteriores. Esta estratégia é equivalente à usada em (ONISHI et al., 2019).

Uma vez selecionados os cortes, estes são organizados num volume de dimensão $64 \times 64 \times k$ voxels. As estratégias descritas são ilustradas na figura 20. No caso em questão, temos $n = 5$ e $k = 3$:

Figura 20 – Estratégias de montagem de volume.



Nas situações em que $n < k$, ou seja, quando o número de cortes é insuficiente para compor o volume os k cortes são atingidos adicionando cortes completamente pretos no final dos volumes.

Neste trabalho, os valores de k avaliados são 5, 6 e 7 cortes. Escolhemos o número de 6 cortes, por se tratar da mediana do número de cortes dos nódulos malignos. Valores menores podem descartar informação de muitos nódulos e valores muito maiores podem gerar um grande número de volumes com cortes na cor preta. O valor de 6 cortes foi usado também em (KANG et al., 2017). Tomamos esse valor como ponto de partida e experimentamos também executar o procedimento com 5 e 7 cortes.

3.3.2 Balanceamento e Aumento de Base

Como mostra a tabela 1, os dados se encontram desbalanceados, com 698 nódulos benignos para 308 malignos. Usando os dados da forma como se encontram, é provável que a RNC tenha uma inclinação a classificar os nódulos como benignos, que são maioria com aproximadamente 70% da base. Esse problema foi solucionado efetuando rotações nos volumes, de forma a balancear a base e aumentá-la. Esta estratégia é empregada em (KANG et al., 2017; ONISHI et al., 2019).

Esta operação foi feita de forma distinta na etapa de OH (seção 3.4) e de classificação (seção 3.5), por usarem partições diferentes dos dados. A tabela 2 resume as operações:

Tabela 2 – Balanceamento e aumento de base nas diferentes etapas.

Etapa	Nódulos	Teste	Restante	Rotações	Treino
OH	698 benignos	50	648	90°	2592 nódulos
	308 malignos	50	258	36°	2580 nódulos
Classificação	698 benignos	30-31	667-668	72°	3335 nódulos
	308 malignos	30-31	277-278	30°	3324 nódulos

3.3.2.1 Etapa de Otimização de Hiperparâmetros

Para esta etapa, foram selecionados 50 nódulos malignos e 50 nódulos benignos para constituir o conjunto de validação, enquanto o restante foi usado como conjunto de treino. Na base de treino, aos 648 nódulos benignos foram efetuadas rotações em intervalos de 90°, totalizando 2.592 nódulos benignos; aos 258 malignos foram feitas rotações em intervalos de 36°, totalizando 2580 nódulos.

3.3.2.2 Etapa de Classificação

A validação foi feita por meio de validação cruzada 10-fold com nódulos balanceados, gerando conjuntos de teste com 30-31 nódulos de cada classe. Para um conjunto de treino de 31 nódulos de cada, restam 667 nódulos benignos e 277 nódulos malignos para compor o conjunto de treino. Como estas são quantidades distintas das usadas na etapa de OH, as rotações foram realizadas de forma diferente para manter os dados balanceados: os 667 nódulos benignos são rotacionados em intervalos de 72°, totalizando 3335 nódulos; os 277 malignos são rotacionados em intervalos de 30°, totalizando 3324 nódulos.

3.4 OTIMIZAÇÃO DE HIPERPARÂMETROS

A primeira etapa do processo de classificação é a escolha da topologia da rede. A escolha da topologia foi feita usando a biblioteca Hyperas, usando o algoritmo TPE descrito na seção 2.3.2. Como temos um grande número de modelos a ser avaliado, optamos por trabalhar com um espaço de busca reduzido, com valores predefinidos de número de filtros. Desta forma podemos executar menos passos da busca e atingir bons resultados mais rapidamente. Estes valores foram tomados com base em experimentos manuais com topologias de rede.

A tabela 3 descreve os parâmetros utilizados para a escolha da melhor topologia de rede com uma camada convolucional (RNC1).

Na arquitetura com duas camadas convolucionais, o espaço de busca é apresentado na tabela 4. Neste caso, foi aplicado um *padding* nas camadas convolucionais, para evitar que os

Tabela 3 – Espaço de busca para a arquitetura RNC1.

Camada	Hiperparâmetro	Espaço de busca
Convolutacional	Número de filtros	{32, 48, 64, 96}
	Tamanho do Kernel	(3x3x3)
Max-pooling	Tamanho do Kernel	(2x2x2)
Densa 1	Número de unidades	{64, 96, 128}
	Dropout	[0, 0.5]
Densa 2	Número de unidades	{16, 24, 32}
	Dropout	[0, 0.5]

feature maps tivessem suas dimensões muito reduzidas para as operações das camadas seguintes.

Tabela 4 – Espaço de busca para a arquitetura RNC2.

Camada	Hiperparâmetro	Espaço de busca
Convolutacional 1	Número de filtros	{16, 32, 48}
	Tamanho do Kernel	(3x3x3)
Convolutacional 2	Número de filtros	{64, 96, 128}
	Tamanho do Kernel	(3x3x3)
Max-pooling	Tamanho do Kernel	(2x2x2)
Densa 1	Número de unidades	{64, 96, 128}
	Dropout	[0, 0.5]
Densa 2	Número de unidades	{16, 24, 32}
	Dropout	[0, 0.5]

Em ambas as arquiteturas, o algoritmo de otimização de aprendizagem escolhido foi o *RMSprop* e a taxa de aprendizagem foi fixada em 10^{-4} . As funções de ativação escolhidas foram *ReLU* para as unidades das camadas convolucionais e densas e sigmoide na camada de saída. Esses valores apresentaram melhores resultados nos estudos iniciais e são amplamente usados na literatura (KANG et al., 2017). A função de perda utilizada foi a *binary crossentropy*.

A busca foi configurada para executar 30 avaliações de modelos e o critério de aptidão usado foi a acurácia do modelo na classificação dos exemplos da base de validação.

3.5 CLASSIFICAÇÃO

A validação dos modelos foi feita usando validação cruzada 10-fold. Os conjuntos de treino e teste foram os mesmos para todas as estratégias, permitindo comparações mais justas entre os modelos. As configurações da rede foram as encontradas na etapa de OH (seção 3.4).

3.5.1 Métricas avaliadas

Para cada modelo foram calculadas 5 métricas: acurácia, especificidade, sensibilidade, F1-score e AUC. Este conjunto de métricas é frequente em trabalhos da área (KANG et al.,

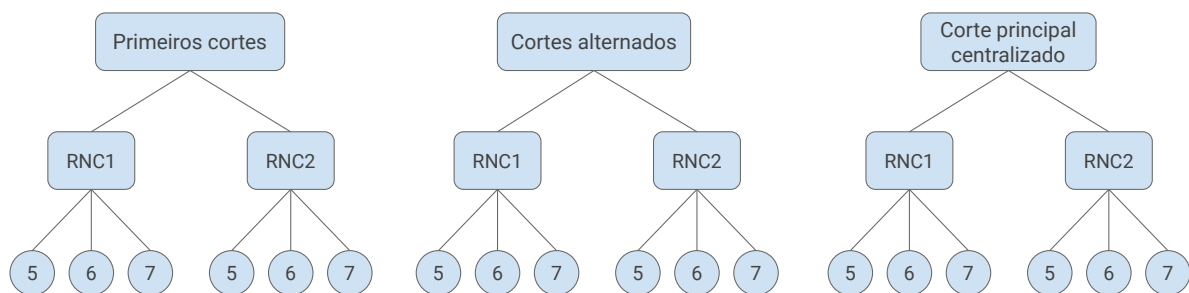
2017; DEY et al., 2018) e são bastante descritivas do comportamento da rede. Estas métricas foram calculadas para cada iteração da validação cruzada e no fim consideramos a média de cada métrica, computando também o desvio padrão destes valores.

Além disso, geramos as curvas ROC para cada iteração da validação e calculamos uma curva média que represente o modelo, também apresentando o desvio padrão dessa curva. A curva ROC e a AUC são técnicas bem estabelecidas na literatura de sistemas CADx, sendo portanto a métrica que usaremos na escolha do melhor modelo.

4 RESULTADOS E DISCUSSÃO

Os modelos avaliados são apresentados na figura 21. No nível mais alto temos as três estratégias de montagem dos volumes: primeiros cortes, cortes alternados e corte principal centralizado. Um nível abaixo, temos as duas arquiteturas utilizadas: com uma camada convolucional/max-pooling (RNC1) e com duas uma camadas convolucional/max-pooling. Por fim, para cada estratégia e arquitetura são utilizados volumes compostos por 5, 6 e 7 cortes.

Figura 21 – Resumo dos diferentes modelos avaliados.



Nas próximas seções são apresentados os resultados obtidos na etapa de OH descrita na seção 3.4. Em seguida, o desempenho da classificação dos modelos de cada estratégia de montagem de volume são apresentados e comparados, bem como os modelos de melhor desempenho em cada estratégia. Por fim, os valores obtidos são comparados com a literatura.

4.1 TOPOLOGIAS

O primeiro resultado são as topologias de rede fornecidas pela OH. A tabela 5 apresenta os resultados da busca para as arquiteturas de uma camada convolucional. A coluna *Estratégia* indica o método empregado na construção dos volumes e a coluna *Cortes*, o número de cortes usados nestes volumes. A coluna *Conv1* indica o número de unidades na camada convolucional, enquanto as colunas *Densa1* e *Densa2* dizem respeito ao número de unidades nas camadas densas. As colunas *Drop1* e *Drop2* mostram a taxa de Dropout empregada após as camadas *Densa1* e *Densa2*, respectivamente. A coluna *Param* indica o total de parâmetros treináveis, que contabiliza os pesos e vieses da rede.

A tabela 6 apresenta os resultados para a arquitetura com duas camadas convolucionais. Os parâmetros listados são os mesmos apresentados na tabela 5, com a adição da coluna *Conv2* para o número de unidades presente na segunda camada convolucional da rede.

Apesar da topologia mais complexa, as redes de arquitetura RNC2 tiveram em média

Tabela 5 – Topologias fornecidas pela OH para redes de arquitetura CNN1.

Estratégia	Cortes	Conv1	Densa1	Drop1	Densa2	Drop2	Param
Primeiros cortes	5	48	128	0,258	16	0,213	5.907.937
	6	32	96	0,367	32	0,346	3.939.329
	7	48	64	0,435	24	0,160	2.954.385
Cortes alternados	5	16	128	0,489	16	0,419	3.189.953
	6	16	128	0,489	16	0,419	3.189.953
	7	16	128	0,489	16	0,419	3.189.953
Corte principal centralizado	5	48	128	0,258	16	0,213	5.907.937
	6	48	64	0,418	32	0,498	2.954.657
	7	48	64	0,202	16	0,127	2.954.113

Tabela 6 – Topologias fornecidas pela OH para redes de arquitetura CNN2.

Estratégia	Cortes	Conv1	Conv2	Densa1	Drop1	Densa2	Drop2	Param
Primeiros cortes	5	32	128	128	0,477	16	0,069	4.308.129
	6	16	64	128	0,021	24	0,270	2.128.561
	7	16	96	128	0,488	16	0,418	3.189.953
Cortes alternados	5	16	96	128	0,488	16	0,418	3.189.953
	6	16	96	128	0,488	16	0,418	3.189.953
	7	16	96	128	0,488	16	0,418	3.189.953
Corte principal centralizado	5	16	96	128	0,488	16	0,418	3.189.953
	6	16	96	128	0,488	16	0,418	3.189.953
	7	16	96	128	0,488	16	0,418	3.189.953

menos parâmetros treináveis (3.196.262 parâmetros) que as redes RNC1 (3.798.691 parâmetros). O número de parâmetros é um dos fatores que influenciam o tempo de treino de uma rede, assim como o número de camadas. A maioria das redes do tipo RNC2 tiveram uma mesma topologia como resultado da busca de parâmetros.

4.2 DESEMPENHO DA CLASSIFICAÇÃO

Nesta seção são apresentados os resultados para cada topologia apresentada na seção 4.1. A validação foi feita como descrito na seção 3.5.

O critério para escolha do melhor modelo é o valor de *AUC*, por ser tradicional em aplicações de diagnóstico e se tratar de uma métrica bastante descritiva da capacidade de um modelo. Em casos de empate, é levado em conta o *F1-score*, por sintetizar a precisão e cobertura - ambas características desejáveis nos modelos em questão. Os resultados são apresentados para as arquiteturas RNC1 e RNC2. Os melhores valores de cada métrica estão destacados em negrito.

4.2.1 Estratégia de primeiros cortes

Os resultados do experimento para a estratégia de primeiros cortes são apresentados na tabela 7.

Tabela 7 – Resultados para a classificação usando a estratégia de primeiros cortes.

Arquitetura	Cortes	Acurácia	Sensibilidade	Especificidade	F1-score	AUC
RNC1	5	81,37%	84,83%	77,91%	81,65	0,89
	6	80,73%	77,69%	83,77%	79,77	0,89
	7	79,59%	81,19%	77,98%	79,44	0,88
RNC2	5	81,03%	82,22%	79,85%	80,71	0,88
	6	79,90%	77,67%	82,13%	79,01	0,88
	7	79,40%	80,56%	78,25%	79,15	0,88

As redes de arquitetura RNC1 alcançaram os melhores resultados em todas as 5 métricas. Também obtiveram o melhor desempenho sob os critérios deste trabalho, atingindo a AUC de 0,89 ao usar volumes de 5 e 6 cortes e F1-score de 81,65 em volumes de 5 cortes. A rede RNC1 de 5 cortes apresentou uma sensibilidade de 84,83%, tendo maior capacidade de classificação de positivos (nódulos malignos) e a rede RNC1 de 6 cortes apresentou especificidade de 83,77%, sendo um melhor classificador de negativos (nódulos benignos). Para esta estratégia, o modelo da arquitetura RNC1 e volume de 5 cortes foi escolhido como o melhor.

4.2.2 Estratégia de cortes alternados

A tabela 8 apresenta os resultados obtidos pelos modelos no uso dos volumes de cortes alternados.

Tabela 8 – Resultados para a classificação usando a estratégia de cortes alternados.

Arquitetura	Cortes	Acurácia	Sensibilidade	Especificidade	F1-score	AUC
RNC1	5	81,20%	81,22%	81,18%	81,13	0,88
	6	80,41%	84,49%	76,33%	81,01	0,87
	7	79,91%	79,28%	80,54%	79,67	0,88
RNC2	5	80,55%	82,19%	78,90%	80,46	0,89
	6	79,72%	80,88%	78,56%	79,66	0,89
	7	79,72%	80,88%	78,56%	79,50	0,89

As redes de arquitetura RNC2 obtiveram um melhor desempenho, atingindo AUC de 0,89 em todos os casos. Apesar disso, as redes de arquitetura RNC1 obtiveram os valores mais altos nas demais métricas. Dentre as redes de arquitetura RNC2, aquela que usou volumes de 5 cortes obteve o melhor desempenho, com F1-score de 80,46 e uma sensibilidade de 82,19%, indicando uma boa capacidade de classificação de casos positivos (nódulos malignos).

4.2.3 Estratégia de corte principal centralizado

Os resultados para a classificação usando volumes de corte principal centralizado são apresentados na tabela 9.

Tabela 9 – Resultados para a classificação usando a estratégia de corte principal centralizados.

Arquitetura	Cortes	Acurácia	Sensibilidade	Especificidade	F1-score	AUC
RNC1	5	79,27%	80,58%	77,97%	79,45	0,88
	6	81,05%	85,13%	76,97%	81,74	0,88
	7	77,81%	76,74%	78,87%	76,66	0,87
RNC2	5	80,05%	79,90%	80,19%	79,64	0,89
	6	79,73%	83,49%	75,96%	80,12	0,89
	7	79,40%	80,56%	78,25%	79,13	0,89

As redes de arquitetura RNC2 obtiveram um melhor desempenho, atingindo AUC de 0,89 em todos os casos. No entanto, as redes de arquitetura RNC1 obtiveram os valores mais altos em 3 das 5 métricas. Dentre as redes de arquitetura RNC2, aquela que usou volumes de 6 cortes obteve o melhor desempenho, com F1-score de 80,12 e uma sensibilidade de 83,49%, indicando uma boa capacidade de classificação de casos positivos (nódulos malignos).

4.3 VISÃO GERAL

Nesta seção apresentamos uma visão geral dos resultados obtidos na classificação. Os modelos com melhor desempenho em cada estratégia são listados na tabela 10, bem como os valores obtidos em cada uma das métricas.

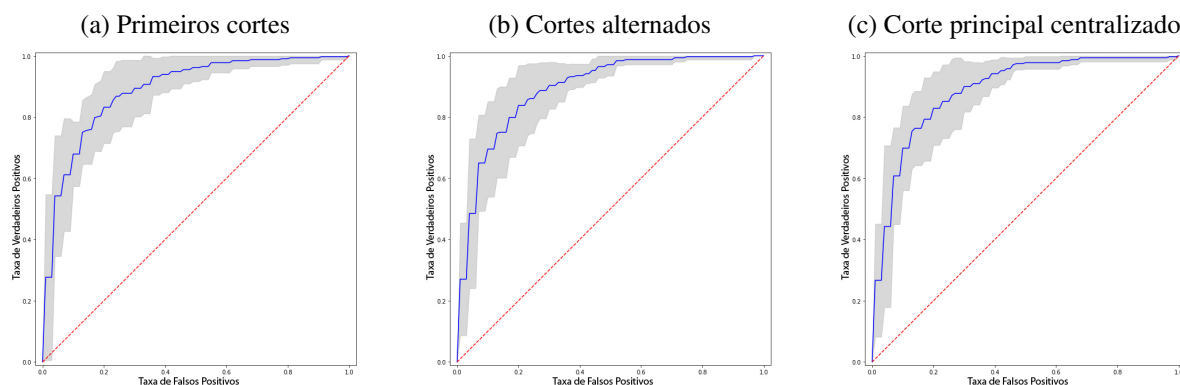
Tabela 10 – Modelos com melhor desempenho para cada estratégia.

Estratégia	Arq.	Cortes	Acurácia	Sensibilidade	Especificidade	F1-score	AUC
Primeiros cortes	RNC1	5	81,37%	84,83%	77,91%	81,65	0,89
Cortes alternados	RNC2	5	80,55%	82,19%	78,90%	80,46	0,89
Corte principal centralizado	RNC2	6	79,73%	83,49%	75,96%	80,12	0,89

De acordo com os critérios propostos, o modelo de melhor desempenho é o que utiliza a estratégia de volumes com os primeiros 5 cortes e rede de arquitetura RNC1. Este modelo apresentou o maior valor de F1-score (81,65%) acurácia com (81,37%) e sensibilidade (84,83%), além da AUC equivalente aos demais. As curvas ROC para cada um destes modelos é mostrada na figura 22.

A linha azul é a curva média das curvas geradas por cada iteração da validação cruzada. A área em cinza representa o desvio padrão destas curvas, sendo um indicativo da estabilidade do modelo.

Figura 22 – Curvas ROC para os melhores modelos de cada estratégia.



As curvas têm um comportamento muito semelhante, o que confere aos modelos valores próximos em suas métricas. A curva da figura 22a tem um acento ligeiramente mais acentuado, sinal de que distingue melhor os casos positivos (nódulos malignos), mas tarda a chegar a 100% de verdadeiros positivos, como ocorre mais rapidamente na figura 22b e 22c.

4.4 TRABALHOS RELACIONADOS

A tabela 11 apresenta os resultados do melhor modelo obtido pela metodologia proposta neste trabalho em relação a trabalhos relacionados existentes na literatura.

Tabela 11 – Comparação com outros trabalhos em classificação de nódulos pulmonares.

Trabalho	Modelo utilizado	Acc.	Sens.	Spec.	AUC	Validação
(HUA et al., 2015)	Deep Belief Network (DBN)	-	73,40%	82,2%	-	Leave-one-out
(KUMAR et al., 2015)	Autoencoder + Binary Decision Tree	75,01%	83,35%	-	-	VC 10-fold
(SHEN et al., 2015)	Multi-scale CNN (MCNN)	86,84%	-	-	-	VC 5-fold
(KUMAR et al., 2017)	Autoencoder + Binary Decision Tree	77,52%	79,06%	77,52%	-	VC 10 fold
(SHEN et al., 2017)	Multi-crop CNN (MC-CNN)	87,14%	77,00%	93,00%	0,93	VC 5-fold
(KANG et al., 2017)	Multi-view CNN 3D (MV-CNN)	95,41%	95,68%	94,51%	0,99	5 x VC 10-fold
(DEY et al., 2018)	3D Multi-Output DenseNet	86,84%	-	-	0,90	VC 3-fold
(LIMA, 2019)	CNN 2D	83,70%	86,96%	80,43%	-	20% holdout
Proposto	CNN 3D	81,37%	84,83%	77,91%	0,89	VC 10-fold

O modelo proposto por este trabalho tem desempenho comparável aos trabalhos relacionados, embora distante dos trabalhos com melhores resultados. No entanto, existem alguns pontos que dificultam a comparação entre estes trabalhos.

Poucos trabalhos reportaram a AUC. Em alguns casos não foi possível gerar a curva ROC, por usar modelos nos quais a classificação não é associada a uma probabilidade, como Binary Decision Trees (KUMAR et al., 2015; KUMAR et al., 2017) e estratégias de votação (LIMA, 2019).

Outro ponto importante é o uso de dados distintos. Todos os trabalhos, com exceção de

(DEY et al., 2018), que usa uma base própria, utilizam o LIDC-IDRI. No entanto, o conjunto de nódulos selecionados, bem como a forma como estes foram rotulados enquanto benignos e malignos é, com frequência, diferente. Este trabalho utiliza o subconjunto de nódulos descrito em 3.1, que é o mesmo utilizado em (LIMA, 2019).

A forma de validação também é um fator importante. A validação do tipo *leave-one-out* (HUA et al., 2015) traz resultados mais confiáveis, mas o custo desta técnica é muitas vezes proibitivo. A maioria dos trabalhos chega a um meio termo entre confiabilidade e custo de execução utilizando a validação cruzada de 5 ou 10 folds. O trabalho (KANG et al., 2017) repete a validação cruzada 5 vezes, no entanto, os dados aumentados são usados na base de teste, o que pode ser considerada uma má prática. (LIMA, 2019) valida seu modelo utilizando holdout, o que pode levar a resultados menos confiáveis que as outras técnicas de validação.

5 CONCLUSÃO

O câncer de pulmão é a forma mais frequente de câncer e também a que mais mata. O diagnóstico da doença em seus estados iniciais é crucial para a sobrevivência de seus pacientes, o que motiva o desenvolvimento de ferramentas computacionais capazes de auxiliar neste diagnóstico.

O uso de técnicas de aprendizagem profunda trouxe grandes avanços no diagnóstico automático de câncer de pulmão. Mais recentemente, o uso de modelos capazes de utilizar a informação tridimensional dos nódulos tem apresentado resultados superiores às técnicas existentes, porém, mesmo diante destes avanços, a classificação de nódulos pulmonares ainda é um problema em aberto na literatura.

Este trabalho apresentou um modelo de classificação de nódulos pulmonares utilizando redes neurais convolucionais 3D. A metodologia proposta englobou desde o processamento dos dados para a rede (estratégia de montagem dos volumes e aumento de base), à escolha da topologia da rede utilizando otimização de hiperparâmetros e validação dos resultados. O modelo proposto nesse trabalho obteve resultados comparáveis à literatura, atingindo AUC de 0,89, acurácia de 81,37% e sensibilidade de 84,83%

Evidenciamos que dentre as estratégias avaliadas a que obteve melhor desempenho foi a montagem do volume com os primeiros cortes dos nódulos, embora a vantagem seja tênue. Observamos também que volumes de 5 cortes obtiveram desempenho igual ou superior aos volumes de 6 e 7 cortes, o que indica que, na metodologia proposta, 5 cortes são o suficiente para a obtenção de bons resultados, com a vantagem de gerar modelos menos exigentes computacionalmente.

REFERÊNCIAS

- ABADI, M.; BARHAM, P.; CHEN, J.; CHEN, Z.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; IRVING, G.; ISARD, M. et al. Tensorflow: A system for large-scale machine learning. In: **12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)**. [S.l.: s.n.], 2016. p. 265–283.
- ADETIBA, E.; OLUGBARA, O. O. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. **The Scientific World Journal**, Hindawi, v. 2015, 2015.
- ANIRUDH, R.; THIAGARAJAN, J. J.; BREMER, T.; KIM, H. Lung nodule detection using 3d convolutional neural networks trained on weakly labeled data. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging 2016: Computer-Aided Diagnosis**. [S.l.], 2016. v. 9785, p. 978532.
- ARLOT, S.; CELISSE, A. et al. A survey of cross-validation procedures for model selection. **Statistics surveys**, The author, under a Creative Commons Attribution License, v. 4, p. 40–79, 2010.
- ARMATO, S. G.; MCLENNAN, G.; BIDAUT, L.; MCNITT-GRAY, M. F.; MEYER, C. R.; REEVES, A. P.; ZHAO, B.; ABERLE, D. R.; HENSCHKE, C. I.; HOFFMAN, E. A. et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. **Medical physics**, Wiley Online Library, v. 38, n. 2, p. 915–931, 2011.
- BANNISTER, N.; BROGGIO, J. Cancer survival by stage at diagnosis for england (experimental statistics): Adults diagnosed 2012, 2013, 2014 and followed up to 2015. **Produced in collaboration with Public Health England**, 2016.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, n. Feb, p. 281–305, 2012.
- BERGSTRA, J.; YAMINS, D.; COX, D. D. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In: CITESEER. **Proceedings of the 12th Python in science conference**. [S.l.], 2013. p. 13–20.
- BERGSTRA, J. S.; BARDENET, R.; BENGIO, Y.; KÉGL, B. Algorithms for hyper-parameter optimization. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2011. p. 2546–2554.
- BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 68, n. 6, p. 394–424, 2018.
- CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Semantic image segmentation with deep convolutional nets and fully connected crfs. **arXiv preprint arXiv:1412.7062**, 2014.
- CHEN, W.; ZHENG, R.; BAADE, P. D.; ZHANG, S.; ZENG, H.; BRAY, F.; JEMAL, A.; YU, X. Q.; HE, J. Cancer statistics in china, 2015. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 66, n. 2, p. 115–132, 2016.

CHOLLET, F. Deep learning with python. Manning Publications, 2018.

CHOLLET, F. et al. **Keras (2015)**. 2017.

CHUQUICUSMA, M. J.; HUSSEIN, S.; BURT, J.; BAGCI, U. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In: IEEE. **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.], 2018. p. 240–244.

CLAESEN, M.; MOOR, B. D. Hyperparameter search in machine learning. **arXiv preprint arXiv:1502.02127**, 2015.

DENG, L. A tutorial survey of architectures, algorithms, and applications for deep learning. **APSIPA Transactions on Signal and Information Processing**, Cambridge University Press, v. 3, 2014.

DEY, R.; LU, Z.; HONG, Y. Diagnostic classification of lung nodules using 3d neural networks. In: IEEE. **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.], 2018. p. 774–778.

DING, J.; LI, A.; HU, Z.; WANG, L. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2017. p. 559–567.

FAWCETT, T. An introduction to roc analysis. **Pattern recognition letters**, Elsevier, v. 27, n. 8, p. 861–874, 2006.

FERREIRA, J. R.; OLIVEIRA, M. C.; AZEVEDO-MARQUES, P. M. de. Characterization of pulmonary nodules based on features of margin sharpness and texture. **Journal of digital imaging**, Springer, v. 31, n. 4, p. 451–463, 2018.

FILHO, A. L.; MACHADO, A. P.; OLIVEIRA, M. **Modelo para Classificação de Nódulos Pulmonares Pequenos usando Descritores Radiomics**. Dissertação (Mestrado) — University of Alagoas - (UFAL), 2016.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. [S.l.]: Springer series in statistics New York, 2001. v. 1.

GINNEKEN, B. V.; SETIO, A. A.; JACOBS, C.; CIOMPI, F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: IEEE. **2015 IEEE 12th International symposium on biomedical imaging (ISBI)**. [S.l.], 2015. p. 286–289.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

HAN, F.; WANG, H.; ZHANG, G.; HAN, H.; SONG, B.; LI, L.; MOORE, W.; LU, H.; ZHAO, H.; LIANG, Z. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. **Journal of digital imaging**, Springer, v. 28, n. 1, p. 99–115, 2015.

HUA, K.-L.; HSU, C.-H.; HIDAYATI, S. C.; CHENG, W.-H.; CHEN, Y.-J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. **OncoTargets and therapy**, Dove Press, v. 8, 2015.

JUNIOR, J. R. F.; OLIVEIRA, M. C.; AZEVEDO-MARQUES, P. M. de. Cloud-based nosql open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. **Journal of digital imaging**, Springer, v. 29, n. 6, p. 716–729, 2016.

JÚNIOR, J. R. F. et al. Auxílio computadorizado ao diagnóstico do câncer de pulmão otimizado por gpu. Universidade Federal de Alagoas, 2015.

KANG, G.; LIU, K.; HOU, B.; ZHANG, N. 3d multi-view convolutional neural networks for lung nodule classification. **PloS one**, Public Library of Science, v. 12, n. 11, p. e0188290, 2017.

KIM, B.-C.; SUNG, Y. S.; SUK, H.-I. Deep feature learning for pulmonary nodule classification in a lung ct. In: IEEE. **2016 4th International Winter Conference on Brain-Computer Interface (BCI)**. [S.l.], 2016. p. 1–3.

KNIGHT, S. B.; CROSBIE, P. A.; BALATA, H.; CHUDZIAK, J.; HUSSELL, T.; DIVE, C. Progress and prospects of early detection in lung cancer. **Open biology**, The Royal Society, v. 7, n. 9, p. 170070, 2017.

KUMAR, D.; CHUNG, A. G.; SHAFEE, M. J.; KHALVATI, F.; HAIDER, M. A.; WONG, A. Discovery radiomics for pathologically-proven computed tomography lung cancer prediction. In: SPRINGER. **International Conference Image Analysis and Recognition**. [S.l.], 2017. p. 54–62.

KUMAR, D.; WONG, A.; CLAUSI, D. A. Lung nodule classification using deep features in ct images. In: IEEE. **2015 12th Conference on Computer and Robot Vision**. [S.l.], 2015. p. 133–138.

LANTZ, B. **Machine learning with R**. [S.l.]: Packt Publishing Ltd, 2013.

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

LIMA, L. L. de. **Modelo Computacional para Classificação de Nódulos Pulmonares Utilizando Redes Neurais Convolucionais**. Dissertação (Mestrado) — Universidade Federal de Alagoas, 2019.

LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. **Medical image analysis**, Elsevier, v. 42, p. 60–88, 2017.

LUCENA, D. J. F. de; JUNIOR, J. R. F.; MACHADO, A. P.; OLIVEIRA, M. C. Automatic weighing attribute to retrieve similar lung cancer nodules. **BMC medical informatics and decision making**, BioMed Central, v. 16, n. 2, p. 79, 2016.

MONTAVON, G.; ORR, G.; MÜLLER, K.-R. **Neural networks: tricks of the trade**. [S.l.]: springer, 2012. v. 7700.

ONISHI, Y.; TERAMOTO, A.; TSUJIMOTO, M.; TSUKAMOTO, T.; SAITO, K.; TOYAMA, H.; IMAIZUMI, K.; FUJITA, H. Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. **BioMed research international**, Hindawi, v. 2019, 2019.

OST, D.; FEIN, A. M.; FEINSILVER, S. H. The solitary pulmonary nodule. **New England Journal of Medicine**, Mass Medical Soc, v. 348, n. 25, p. 2535–2542, 2003.

POWERS, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. Bioinfo Publications, 2011.

PUMPERLA, M. **Hypersa**. 2019. Disponível em: <<http://maxpumperla.com/hyperas/>>.

RAVÌ, D.; WONG, C.; DELIGIANNI, F.; BERTHELOT, M.; ANDREU-PEREZ, J.; LO, B.; YANG, G.-Z. Deep learning for health informatics. **IEEE journal of biomedical and health informatics**, IEEE, v. 21, n. 1, p. 4–21, 2017.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2015. p. 91–99.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. et al. Learning representations by back-propagating errors. **Cognitive modeling**, v. 5, n. 3, p. 1, 1988.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural networks**, Elsevier, v. 61, p. 85–117, 2015.

SHEN, J.; LIU, Z.; TODD, N. W.; ZHANG, H.; LIAO, J.; YU, L.; GUARNERA, M. A.; LI, R.; CAI, L.; ZHAN, M. et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microrna biomarkers. **BMC cancer**, BioMed Central, v. 11, n. 1, p. 374, 2011.

SHEN, W.; ZHOU, M.; YANG, F.; YANG, C.; TIAN, J. Multi-scale convolutional neural networks for lung nodule classification. In: SPRINGER. **International Conference on Information Processing in Medical Imaging**. [S.l.], 2015. p. 588–599.

SHEN, W.; ZHOU, M.; YANG, F.; YU, D.; DONG, D.; YANG, C.; ZANG, Y.; TIAN, J. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. **Pattern Recognition**, Elsevier, v. 61, p. 663–673, 2017.

SILVA, C. I. S.; MARCHIORI, E.; JÚNIOR, A. S. S.; MÜLLER, N. L. Illustrated brazilian consensus of terms and fundamental patterns in chest ct scans. **Jornal Brasileiro de Pneumologia**, SciELO Brasil, v. 36, n. 1, p. 99–123, 2010.

SONODA, S.; MURATA, N. Neural network with unbounded activation functions is universal approximator. **Applied and Computational Harmonic Analysis**, Elsevier, v. 43, n. 2, p. 233–268, 2017.

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

TEAM, N. L. S. T. R. The national lung screening trial: overview and study design. **Radiology**, Radiological Society of North America, Inc., v. 258, n. 1, p. 243–253, 2011.

WANG, J.; LIU, X.; DONG, D.; SONG, J.; XU, M.; ZANG, Y.; TIAN, J. Prediction of malignant and benign of lung tumor using a quantitative radiomic method. In: IEEE. **2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.], 2016. p. 1272–1275.

WONG, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. **Pattern Recognition**, Elsevier, v. 48, n. 9, p. 2839–2846, 2015.

World Health Organisation. **Vision impairment and blindness, Fact Sheet N°282**. 2019. <<http://www.who.int/mediacentre/factsheets/fs282/fr/>>, Last accessed on 2019-02-14.

ZOU, K. H. **Receiver Operating Characteristic (ROC) Literature Research**. 2019. Disponível em: <<https://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html>>.