

RESSALVA

Atendendo solicitação da autora, o texto completo desta dissertação será disponibilizado somente a partir de 08/01/2021.



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

Daniela Fernanda Taino

Um modelo baseado em algoritmo genético
para seleção de características e
classificação de padrões em imagens
médicas

São José do Rio Preto
2020

Daniela Fernanda Taino

Um modelo baseado em algoritmo genético
para seleção de características e
classificação de padrões em imagens
médicas

Dissertação apresentada como parte dos requisitos para obtenção do título do Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientador:

Prof. Dr. Leandro Alves Neves

São José do Rio Preto
2020

T134m	Taino, Daniela Fernanda Um modelo baseado em algoritmo genético para seleção de características e classificação de padrões em imagens médicas / Daniela Fernanda Taino. -- São José do Rio Preto, 2020 102 f. Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto Orientador: Leandro Alves Neves 1. Algoritmos genéticos. 2. Reconhecimento de padrões. 3. Inteligência artificial. 4. Imagens médicas. 5. Câncer. I. Título.
-------	---

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

Essa ficha não pode ser modificada.

Daniela Fernanda Taino

Um modelo baseado em algoritmo genético
para seleção de características e
classificação de padrões em imagens
médicas

Dissertação apresentada como parte dos requisitos para obtenção do título do Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Comissão Examinadora

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

Profa. Dra. Rogéria Cristiane Gratão de
Souza
Unesp - Câmpus São José do Rio Preto

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

São José do Rio Preto
08 de Julho de 2020

Agradecimentos

Primeiramente, agradeço aos meus pais, Aparecida e Lúcio, e à minha irmã, Denise, pelo apoio incondicional durante toda a minha vida. Agradeço também minha família, principalmente minhas primas Tamiris e Tainá. Tive muita sorte de ser tão próxima de vocês.

Agradeço ao Prof. Dr. Leandro Alves Neves pela orientação, paciência e pelo apoio desde a graduação. Sem a sua ajuda, este trabalho não teria saído do papel.

Agradeço às minhas amigas Isabella, Larissa e Daniela Pereira por sempre estarem ao meu lado. Nunca poderei agradecer suficientemente por tudo o que vocês já fizeram por mim.

Agradeço também aos amigos que o Ibilce me deu: Guilherme Moraes, Lais, Maycon, Matheus, André, João, Leandro, Hugo e João Paulo. Conhecer vocês foi uma das melhores coisas que já me aconteceu. Tive e, felizmente, ainda tenho o privilégio de acompanhar a trajetória de cada um de vocês. Aqueles jovens que sofriam com provas e tomavam café juntos quase todos os dias se tornaram adultos incríveis, bem-sucedidos e com um futuro brilhante pela frente. Talvez as nossas conversas na cantina deixaram uma certa dúvida sobre o quão promissor o futuro seria, mas *juro* que não duvidei de vocês por mais que um instante.

Por fim, agradeço também aos meus colegas de laboratório, Guilherme Freire, Guilherme Rozendo, Álvaro, Eduardo Pavarino, e novamente, ao Matheus. Nossas trocas de experiências e trabalhos foram fundamentais para que este trabalho fosse concluído.

Resumo

A análise de imagens histológicas é baseada na avaliação visual dos tecidos por especialistas, utilizando um microscópio óptico. Essa tarefa pode ser demorada e desafiadora, principalmente devido à complexidade das estruturas e doenças sob investigação. Estes fatos motivaram o desenvolvimento de métodos computacionais para apoiar especialistas em pesquisas e tomadas de decisões. Apesar das diferentes estratégias computacionais disponíveis na literatura, as soluções baseadas em programação genética não foram totalmente exploradas para fornecer a melhor combinação de recursos, algoritmos de seleção e classificadores. Uma abordagem baseada em algoritmo genético, capaz de avaliar um número significativo de características, métodos de seleção e classificadores, é descrita para fornecer uma combinação aceitável para o diagnóstico e reconhecimento de padrões de linfomas não-Hodgkin e câncer colorretal. A estrutura cromossômica foi representada por quatro genes. A avaliação e seleção dos indivíduos, bem como os processos de cruzamento e mutação, foram definidos para distinguir os grupos investigados, com o maior valor da AUC e o menor número de características. Os testes foram realizados considerando 1.512 características de imagens histológicas, diferentes tamanhos de população, taxas de mutação e número de iterações. Uma população inicial, de 50 indivíduos e 50 iterações, forneceu o melhor resultado para o câncer colorretal. Para os linfomas não-Hodgkin, a população inicial foi composta de 500 indivíduos e também foram necessárias 50 iterações para fornecer melhor resultado. Os valores da AUC foram de 0,984 e 0,947 para o câncer colorretal e os linfomas não-Hodgkin, respectivamente. A metodologia proposta, com informações detalhadas sobre os métodos, características e melhores combinações, é uma contribuição relevante para a comunidade interessada no estudo do reconhecimento de padrões em imagens histológicas de linfomas e câncer colorretal.

Palavras-chave: algoritmo genético, seleção de características, associação de técnicas, classificação de características, imagens médicas.

Abstract

The analysis of histological images is based on visual assessment of tissues by specialists using an optical microscopy. This task can be time-consuming and challenging, mainly due to the complexity of the structures and diseases under investigation. These facts have motivated the development of computational methods to support specialists in research and decision-making. Despite the different computational strategies available in the literature, the solutions based on genetic programming have not been fully explored to provide the best combination of features, selection algorithms and classifiers. An approach based on genetic algorithm able to evaluate a significant number of features, selection methods and classifiers is described in order to provide an acceptable association for the diagnosis and pattern recognition of Non-Hodgkin lymphomas and colorectal cancer. The chromosomal structure was represented with four genes. The evaluation and selection of individuals, as well as the crossover and mutation processes were defined to distinguish the groups under investigation, with the highest AUC value and the smallest number of features. The tests were performed considering 1,512 features from histological images, different population sizes, mutation rate and number of iterations. An initial population of 50 individuals and 50 iterations provided the best result for colorectal cancer. For non-Hodgkin's lymphomas, the initial population consisted of 500 individuals and 50 iterations was also necessary to provide the best result. The AUC values were 0.984 and 0.947 for the colorectal cancer and the Non-Hodgkin lymphomas, respectively. The proposed methodology with detailed information regarding the methods, features and best combinations are relevant contributions for the community interested into the study of pattern recognition of colorectal cancer and lymphomas.

Keywords: genetic algorithm, feature selection, combination of techniques, feature classification, medical images.

Lista de Figuras

2.1	Exemplo de uma curva ROC (Fonte: Adaptado de Hanley e McNeil, (1982)).	33
4.1	Estrutura cromossomial adotada no algoritmo proposto para representar os indivíduos da população, considerando a identificação do indivíduo, índice do método de seleção, índice de método de classificação e número de características.	46
4.2	Ilustração da formação de dois novos indivíduos C e D por meio da estratégia de <i>crossover</i> de dois pontos entre os pais A e B.	49
4.3	Ilustração do método proposto (Fonte: Elaborado pela autora).	51
4.4	Exemplos de imagens histológicas colorretais. Em (a), exemplo do grupo benigno e em (b), exemplo do grupo maligno. (Fonte: Sirinukunwattana et al., (2017)).	54
4.5	Exemplos de imagens histológicas dos grupos de linfomas. Em (a), exemplo do grupo MCL. Em (b), do grupo FL e por fim, em (c), exemplo do grupo CLL. (Fontes: Aging, (2020); Institute, (2020)).	54

Lista de Tabelas

2.1	Genótipos e Fenótipos (Fonte: Adaptado de Linden, (2006)).	24
2.2	Estrutura de uma matriz de confusão (Fonte: Elaborado pela autora).	31
2.3	Níveis de relevância de um teste medidos pela AUC (Fonte: Elaborado pela autora).	33
4.1	Parâmetros necessários para a aplicação do método proposto (Fonte: Elaborado pela autora).	47
4.2	Ilustração do vetor de características (Fonte: Adaptado de Ribeiro et al., (2019)).	56
5.1	Parâmetros e valores explorados no método proposto.	58
5.2	Imagens colorretais: resultados obtidos com $P = 5$ e iterações definidas como 5, 25, 50, 100 e 500, sendo que a melhor associação está em negrito. .	59
5.3	Imagens colorretais: resultados obtidos com $P = 25$ e iterações definidas como 5, 25, 50, 100 e 500, sendo que a melhor associação está em negrito. .	59
5.4	Imagens colorretais: resultados obtidos com $P = 50$ e iterações definidas como 5, 25, 50, 100 e 500, sendo que a melhor associação está em negrito. .	59
5.5	Imagens colorretais: resultados obtidos com $P = 500$ e iterações definidas como 5, 25, 50, 100 e 500, sendo que a melhor associação está em negrito. .	60
5.6	Imagens de linfomas NHL: resultados obtidos com $P = 5$ e $Iter$ definido como 5, 25, 50, 100 e 500, A melhor associação está em negrito.	61
5.7	Imagens de linfomas NHL: resultados obtidos com $P = 25$ e $Iter$ definido como 5, 25, 50, 100 e 500, A melhor associação está em negrito.	61
5.8	Imagens de linfomas NHL: resultados obtidos com $P = 50$ e $Iter$ definido como 5, 25, 50, 100 e 500, A melhor associação está em negrito.	61
5.9	Imagens de linfomas NHL: resultados obtidos com $P = 500$ e $Iter$ definido como 5, 25, 50, 100 e 500, A melhor associação está em negrito.	62
5.10	Câncer colorretal: taxas de AUC fornecidas por trabalhos correlatos e pela melhor associação do método proposto.	63
5.11	Linfomas NHL: taxas de AUC de trabalhos correlatos e da melhor combinação fornecida pelo método proposto.	63
5.12	Câncer Colorretal: Taxas de acurácias obtidas pelos métodos <i>Deep Learning 4j</i> (GIBSON; NICHOLSON; PATTERSON, 2016), <i>AlexNet</i> (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e <i>ResNet-50</i> (HE et al., 2016).	65
5.13	NHL: Taxas de acurácias obtidas pelos métodos <i>Deep Learning 4j</i> (GIBSON; NICHOLSON; PATTERSON, 2016), <i>AlexNet</i> (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e <i>ResNet-50</i> (HE et al., 2016).	66

5.14	Taxas de acurácias obtidas pelo método <i>Auto-WEKA 2.0</i> (KOTTHOFF et al., 2017), considerando os mesmos conjuntos de características do método proposto.	66
5.15	Resultados do teste estatístico de <i>Wilcoxon-Mann-Whitney</i> para os desempenhos envolvendo as imagens colorretais: método proposto versus <i>Deep Learning 4j</i> (GIBSON; NICHOLSON; PATTERSON, 2016), <i>AlexNet</i> (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), <i>ResNet-50</i> (HE et al., 2016) e <i>Auto-WEKA 2.0</i> (KOTTHOFF et al., 2017).	69
5.16	Resultados do teste estatístico de <i>Wilcoxon-Mann-Whitney</i> para os desempenhos envolvendo as imagens de linfomas NHL: método proposto versus <i>Deep Learning 4j</i> (GIBSON; NICHOLSON; PATTERSON, 2016), <i>AlexNet</i> (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), <i>ResNet-50</i> (HE et al., 2016) e <i>Auto-WEKA 2.0</i> (KOTTHOFF et al., 2017).	69
5.17	Discriminação das características selecionadas no melhor resultado para distinguir o conjunto colorretal.	70
5.18	Discriminação das características selecionadas no melhor resultado para distinguir os linfomas NHL.	70
A.1	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 5$ e iterações definidas como 5, 25, 50, 100 e 500.	93
A.2	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 25$ e iterações definidas como 5, 25, 50, 100 e 500.	94
A.3	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 50$ e iterações definidas como 5, 25, 50, 100 e 500.	94
A.4	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 500$ e iterações definidas como 5, 25, 50, 100 e 500.	95
A.5	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 5$ e iterações definidas como 5, 25, 50, 100 e 500,	95
A.6	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 25$ e iterações definidas como 5, 25, 50, 100 e 500.	96
A.7	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 50$ e iterações definidas como 5, 25, 50, 100 e 500.	96
A.8	Colorretal: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 500$ e iterações definidas como 5, 25, 50, 100 e 500.	97
A.9	Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 5$ e iterações definidas como 5, 25, 5, 100 e 500.	97

A.10 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 25$ e iterações definidas como 5, 25, 50, 100 e 500.	98
A.11 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 50$ e iterações definidas como 5, 25, 50, 100 e 500.	98
A.12 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 500$ e iterações definidas como 5, 25, 50, 100 e 500.	99
A.13 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 5$ e iterações definidas como 5, 25, 50, 100 e 500.	99
A.14 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 25$ e iterações definidas como 5, 25, 50, 100 e 500.	100
A.15 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 50$ e iterações definidas como 5, 25, 50, 100 e 500.	100
A.16 Linfoma: melhor combinação (método de seleção, classificador, número de características e AUC média) obtida com $P = 500$ e iterações definidas como 5, 25, 50, 100 e 500.	101
A.17 Valores de Acurácia utilizados no teste de <i>Wilcoxon-Mann-Whitney</i> para o câncer colorretal.	102
A.18 Valores de Acurácia utilizados no teste de <i>Wilcoxon-Mann-Whitney</i> para os linfomas NHL.	102

Lista de Abreviaturas e Siglas

G_{clf}	Índice do Método de Classificação
G_{id}	Identificador Único do Indivíduo
G_{num}	Número de Características na Iteração
G_{sel}	Índice do Método de Seleção
$Iter$	Iterações
$MaxF$	Número máximo de características a serem consideradas
$MeanAuc$	AUC média
Acc	Acurácia
ADABOOST	<i>Adaptive Boosting</i>
AG	Algoritmo Genético
AGFS	<i>Adaptive Genetic Fuzzy System</i>
AUC	Área sob a Curva ROC
BPNN	<i>Back Propagation Neural Network</i>
CLL	<i>Chronic Lymphocytic Leukemia</i>
CNN	<i>Convolutional Neural Network</i>
CSSVM	<i>Cost Sensitive Support Vector Machine</i>
Curva ROC	<i>Receiver Operating Characteristic Curve</i>
DL	<i>Deep Learning</i>
DT	<i>Decision Tree</i>
FDT	<i>Functional Decision Tree</i>
FL	<i>Follicular Lymphoma</i>
FP	Falsos Positivos
KNN	<i>K-Nearest Neighbors</i>

LOO *Leave-One-Out*
LWL *Locally Weighted Learning*
MCL *Mantle Cell Lymphoma*
MI *Mutual Information*
MLP *Multilayer Perceptron*
NHL *Non-Hodkin Lymphoma*
PID *Pima Indian Diabetes*
PSO *Particle Swarm Optimization*
SGD *Stochastic Gradient Descent*
SL *Simple Logistics*
SVM *Support Vector Machine*
VP *Verdadeiros Positivos*

Lista de Algoritmos

1	Estrutura Básica de um Algoritmo Genético (Fonte: Adaptado de Furtado e Lorena, (1998)).	26
---	--	----

Sumário

1	Introdução	14
1.1	Motivação e Justificativas	19
1.2	Objetivos	21
1.3	Organização do Trabalho	22
2	Fundamentação Teórica	23
2.1	Algoritmos Genéticos (AG)	23
2.2	Seleção de Características	26
2.3	Processo de Classificação de Dados	28
2.3.1	Tipos de Classificadores	28
2.3.2	Métricas para Avaliação de Desempenho	31
2.4	Considerações sobre a Fundamentação Teórica	34
3	Trabalhos Relacionados	36
3.1	Aplicação do AG em Contextos Diversificados	36
3.2	Aplicação do AG em Contextos de Imagens Médicas	38
3.3	Considerações sobre os Trabalhos Relacionados	43
4	Metodologia	45
4.1	Estrutura cromossômica e população inicial	45
4.2	Avaliação e Seleção da População	47
4.3	Operações de <i>Crossover</i> e Mutação	49
4.4	Especificações técnicas para implementação do modelo	52
4.5	Aplicação: Casos de Estudo	52
4.5.1	Composição do Vetor de Características	55
5	Resultados e Discussão	57
5.1	Parâmetros e Variações	57
5.2	Visão Geral do Desempenho de Classificação	62
6	Conclusão	72
6.1	Contribuições obtidas	73
6.2	Trabalhos futuros	74
	Referências	75
	APÊNDICE A	93

A.1	Resultados considerando variações na taxa de mutação (m)	93
A.1.1	Taxa de mutação $m = 0,3$ - Colorretal	93
A.1.2	Taxa de mutação $m = 0,5$ - Colorretal	95
A.1.3	Taxa de mutação $m = 0,3$ - Linfomas	97
A.1.4	Taxa de mutação $m = 0,5$ - Linfomas	99
A.2	Valores utilizados no teste de <i>Wilcoxon-Mann-Whitney</i>	101

Capítulo 1

Introdução

As áreas de radiologia e patologia são especialidades médicas que compreendem o uso de tecnologias de imagens para a investigação de doenças e indicações de diagnósticos, principalmente para identificar doenças com alta mortalidade. Essas especialidades se tornaram cada vez mais comuns, especialmente por possibilitarem diagnósticos via técnicas menos invasivas (PARIKH; TICKMAN, 2005).

A principal dificuldade para um diagnóstico médico está na avaliação da gravidade de achados anormais quando diferentes opiniões podem ocorrer interobservador e intraobservadores. Além disso, o processo de análise de imagens médicas para o diagnóstico consome tempo e pode ser influenciada pela experiência do especialista (VAN GINNEKEN; ROMENY; VIERGEVER, 2001; CHAN et al., 1990; DOI; HUANG, 2007). Esses fatos têm motivado o desenvolvimento e a aplicação de sistemas conhecidos como *Computer-Aided Diagnosis* (CAD) para apoiar os especialistas no estudo e nas tomadas de decisão (YU; GUAN, 2000). Esses sistemas consideram diferentes etapas, tais como extração, análise e seleção de características das imagens, bem como a interpretação e o reconhecimento de possíveis padrões. Um desafio comumente observado em propostas de sistemas CAD é o de indicar as melhores combinações entre algoritmos de seleção e classificação para alcançar as maiores taxas de acerto com o menor número de características (GURCAN et al., 2002; DOI; HUANG, 2007).

Os sistemas CAD estão amplamente relacionados com a área do conhecimento definida

como aprendizado de máquina, do inglês, *machine learning*) (ESTEVA et al., 2017; KOOI et al., 2017; BEJNORDI et al., 2017; ZEZOS et al., 2018). O aprendizado de máquina tem como objetivo a construção de estratégias computacionais que podem aprender a partir de seus erros e definir possíveis previsões sobre dados, aplicando, por exemplo, o raciocínio indutivo (AMORIM; BARONE; MANSUR, 2008). Os sistemas pautados em aprendizado de máquina podem ser observados a partir das etapas de levantamento de dados para treinamento e testes (usados para treinar e medir o desempenho do método), seleção dos características, escolha de classificadores, execução do treinamento e avaliação dos resultados obtidos (AMORIM; BARONE; MANSUR, 2008).

Segundo Janssen, Mourão-Miranda e Schnack ((2018)), a eficácia de técnicas baseadas em aprendizado de máquina tem sido alcançada com o desenvolvimento de algoritmos para atender demandas específicas, bem como pela maior disponibilidade de recursos computacionais e conjuntos de dados necessários para testes e validações. Além disso, o sucesso das abordagens baseadas em aprendizado de máquina deve-se ao enfoque na generalização, que é a manutenção do desempenho do modelo a partir da inserção progressiva de novos dados. É importante destacar que a capacidade de generalização pode ser afetada pelos fatores nomeados como: *underfitting*, no qual o modelo não está generalizado o suficiente; *overfitting*, no qual o modelo fornece resultados relevantes sobre o conjunto de dados utilizado para sua construção, porém ineficaz para prever novos resultados; e, por fim, o erro irreduzível, quando há presença de ruído nos dados e que pode determinar o limite teórico do desempenho do algoritmo (JANSSEN; MOURÃO-MIRANDA; SCHNACK, 2018; DIETTERICH, 1995). Portanto, o desenvolvimento de sistemas CAD combinando as melhores estratégias para contornar esses problemas é um tema desafiador, principalmente para assegurar que a capacidade de generalização não seja afetada (SAHINER; CHAN; HADJIISKI, 2008; JANSSEN; MOURÃO-MIRANDA; SCHNACK, 2018).

Nesse contexto, desde o desenvolvimento das primeiras técnicas de seleção de características e classificação de dados, muito se fez na área do aprendizado de máquina aplicada em medicina. Por exemplo, existem estudos direcionados para a utilização de sistemas CAD em conjunto com os exames comumente encontrados na prática clínica, tais como raios X

(MELENDEZ et al., 2015; WANG, X. et al., 2017; RAJPURKAR et al., 2017; APOSTOLOPOULOS; MPESIANA, 2020), tomografia computadorizada (GURCAN et al., 2002; HUA et al., 2015; CHENG et al., 2016; JALALIAN et al., 2017; EL-ANWAR et al., 2020) e ressonância magnética (ARIMURA et al., 2009; LEITE et al., 2015; SINGH et al., 2015; GIANNINI et al., 2017; LEE et al., 2020), todos com o objetivo de fornecer uma segunda opinião e apurar ainda mais o diagnóstico. As doenças de maior interesse são definidas a partir das suas taxas de mortalidade, letalidade e incidência, além do fato de o diagnóstico precoce contribuir significativamente para a melhoria do prognóstico. Alguns exemplos são o câncer de mama (HO; LAM, 2003; JALALIAN et al., 2017; KOOI et al., 2017; BEJNORDI et al., 2017; WANG, P. et al., 2020), o câncer de pulmão (TAGUCHI et al., 1999; HUA et al., 2015; HUANG; PARK et al., 2017; EL-REGAILY et al., 2018; SILVA et al., 2020), os tumores cerebrais (VRJI; JAYAKUMARI, 2011; PRAVEEN; AGRAWAL, 2015; ANITHA; MURUGAVALLI, 2016; TIWARI; SACHDEVA et al., 2017; TIWARI; SRIVASTAVA; PANT, 2020), o câncer de próstata (DOYLE et al., 2006; KOIZUMI et al., 2017; AFEF et al., 2018; ALDOJ et al., 2020), as doenças cardíacas (ANBARASI; ANUPRIYA; IYENGAR, 2010; SHARMA et al., 2015; PATIDAR; PACHORI; ACHARYA, 2015; EL-ANWAR et al., 2020), os linfomas (HERMINE et al., 2016; CODELLA et al., 2016; ABAS et al., 2017; TCHRAKIAN et al., 2018; RIBEIRO, Matheus Gonçalves et al., 2019; ROBERTO et al., 2017; SWERDLOW; COOK, 2020) e o câncer colorretal (KATHER et al., 2016; KOMINAMI et al., 2016; MISAWA et al., 2016; RIBEIRO, Matheus Gonçalves et al., 2019; JAYACHANDRAN; GHOSH, 2020).

Para investigar as melhores combinações entre características, algoritmos de seleção e classificadores, estudos disponíveis na literatura foram desenvolvidos para maximizar a acurácia de classificações por meio de novas metodologias para a seleção de características, muitas vezes influenciadas por estratégias baseadas em modelos metaheurísticos e *deep learning* (DL). Como exemplo de métodos baseados em DL, podemos citar As Redes Neurais Convolucionais (BRANCATI et al., 2019) e as Redes Neurais Recorrentes (LI et al., 2018). As técnicas inspiradas em analogias encontradas na natureza ou em processos evolutivos (EIBEN; SMITH et al., 2003) merecem destaque, tais como os métodos baseados em colônia

de abelhas (*Bee Colony*) (ZHANG et al., 2015), colônia de formigas (*Ant Colony*) (KARNAN; LOGHESHWARI, 2010) e *Particle Swarm Optimization* ou PSO (REMAMANY et al., 2015) para estudos de tumores cerebrais, bem como o uso de *Differential Evolution* para investigação de nódulos pulmonares (JAFFAR; SIDDIQUI; MUSHTAQ, 2018) e os modelos baseados em Algoritmos Genéticos (AG) (MUNI; PAL; DAS, 2006; WELIKALA et al., 2015; PAUL et al., 2017; JIANG et al., 2017; KEČO; SUBASI; KEVRIC, 2018; FARID; SELIM; KHATER, 2020). É importante mencionar que as metaheurísticas e os métodos baseados em DL apresentam limitações bem-conhecidas. Por exemplo, a colônia de formigas requer mais tempo de execução para convergência e esse modelo apresenta desempenho ruim quando aplicado a problemas caracterizados por grandes espaços de busca (AB WAHAB; NEFTI-MEZIANI; ATYABI, 2015). O método PSO apresenta uma tendência de convergência prematura em pontos médios ótimos, assim como uma fraca busca local e com convergência lenta (AB WAHAB; NEFTI-MEZIANI; ATYABI, 2015). O fraco desempenho em relação à capacidade de busca local também é observado na abordagem *Differential Evolution*, com problemas de convergência prematura. Além disso, nessa estratégia, o desempenho pode diminuir quando a dimensão do espaço de busca cresce expressivamente (MOHAMMED; SABRY; KHORSHID, 2012). O método *Cuckoo Search* também tem o problema de facilmente encontrar soluções locais ótimas e possui uma lenta taxa de convergência (LIU et al., 2018; YANG; DEB, 2009). As desvantagens da estratégia *Bee Colony* estão relacionadas à necessidade de um número expressivo de avaliações na função objetiva e novos testes de aptidão para cada conjunto de novos parâmetros, a fim de melhorar o desempenho (AB WAHAB; NEFTI-MEZIANI; ATYABI, 2015). Os métodos baseados em DL estão associados a um custo computacional extremamente alto quando as imagens de entrada são de alta resolução. Uma solução comumente usada é a redução dessas imagens, mas informações relevantes acabam sendo perdidas no processo (LOPES, 2017). Outras desvantagens são que esses métodos precisam de um número expressivo de dados rotulados para treinamento e teste (RODRIGUES, 2018), além de não ser possível obter um entendimento detalhado sobre as relevâncias das características utilizadas para definir a melhor solução.

A principal desvantagem de um AG é a convergência não rápida, já que os processos

de cruzamento e mutação são aleatórios (AB WAHAB; NEFTI-MEZIANI; ATYABI, 2015). Mesmo considerando essas limitações, a principal vantagem de um AG, em comparação a outras estratégias descritas anteriormente, é ter uma estrutura que permita representar novas formas organizacionais (indivíduos) aceitáveis a partir de uma população anterior bem-sucedida (cruzamento) (BRUDERER; SINGH, 1996) sem perder informações críticas do problema (MUNI; PAL; DAS, 2006) (WHITLEY, 1994). Apresentados por John Holland em 1975 (HOLLAND, 1992), o AG é uma técnica heurística de otimização global, desenvolvida a partir de princípios evolutivos, ou o chamado darwinismo. Essa técnica é empregada na resolução de problemas que geralmente demandam a avaliação de um número significativo de possíveis soluções ou combinações, ou seja, aqueles considerados como problemas de otimização complexos, como o do “caixeiro viajante”, por exemplo (MIRANDA, 2007).

Diversos são os trabalhos que propuseram modelos CAD combinando AG com diferentes classificadores (ANBARASI; ANUPRIYA; IYENGAR, 2010; ÖZÇİFT; GÜLTEN, 2013; LU; ZHU; GU, 2014). Anbarasi, Anupriya e Iyengar ((2010)) exploraram AG com os classificadores *Naive Bayes*, *Clustering* e árvore de decisão para auxiliar no diagnóstico de doenças cardíacas. Özçift e Gülten ((2013)) propuseram em um novo método para o diagnóstico diferencial de doenças eritêmato-escamosas. O método, baseado em um AG e uma *Bayesian Network*, apresentou acurácia de 99,2% e foi superior quando comparado aos desempenhos obtidos com outros quatro classificadores conhecidos: *Support Vector Machine* (98,36%), *Multi-Layer Perceptron* (97,00%), *Simple Logistic* (98,36%) e *Functional Decision Tree* (97,81%). Um sistema inteligente para o diagnóstico de câncer de pulmão, baseado em AG e em *Mutual Information*, foi proposto com a finalidade de selecionar as características mais relevantes em um conjunto e, posteriormente, classificá-los (LU; ZHU; GU, 2014). Os classificadores aplicados foram *Support Vector Machine* (SVM), *Back Propagation Neural Network* (BPNN) e o *K-Nearest Neighbors* (KNN).

As abordagens baseadas em AG também foram definidas para reduzir o número de características. Por exemplo, Velu e Kashwan ((2013)) apresentou uma proposta para analisar características de pacientes diabéticos. AG, *Expectation-maximization Algorithm* e *h-means+* foram combinados e utilizados para definir agrupamentos e encontrar similaridades entre as

características. A combinação forneceu um coeficiente de correlação de 0,96. No trabalho proposto por Kumar, Ramachandra e Nagamani ((2014)), foi descrita uma combinação entre os AGs, como método para selecionar as características, e SVM como classificador. O método foi capaz de indicar um ótimo subconjunto de características e fornecer as melhores taxas de desempenho quando comparadas aos trabalhos disponíveis na literatura. Outro exemplo foi a proposta descrita por Gokulnath e Shantharajah ((2019)), em que o AG foi utilizado como meio de encontrar as características mais relevantes e reduzir o conjunto inicial para obter melhores resultados na classificação de doenças cardíacas. Os classificadores utilizados foram BPNN, KNN e *Cost Sensitive Support Vector Machine* (CSSVM). Os autores concluíram que o método superou os desempenhos de trabalhos disponíveis na literatura.

1.1 Motivação e Justificativas

Nos trabalhos disponíveis na literatura (ANBARASI; ANUPRIYA; IYENGAR, 2010; ÖZÇİFT; GÜLTEN, 2013; VELU; KASHWAN, 2013; LU; ZHU; GU, 2014; KUMAR; RAMACHANDRA; NAGAMANI, 2014; GOKULNATH; SHANTHARAJAH, 2019), foram propostas diferentes metodologias para identificar as melhores combinações de técnicas e características. Um ponto em comum entre esses trabalhos foi a utilização de um AG para alcançar os melhores resultados. Além disso, o AG foi associado com diferentes classificadores de padrões, tais como SVM (LU; ZHU; GU, 2014) (GOKULNATH; SHANTHARAJAH, 2019), *Decision Tree* (DT) (ANBARASI; ANUPRIYA; IYENGAR, 2010), KNN, BPNN (LU; ZHU; GU, 2014), *Clustering* (ANBARASI; ANUPRIYA; IYENGAR, 2010), *Multilayer Perceptron* (MLP) (ÖZÇİFT; GÜLTEN, 2013), *Simple Logistics* (SL) (ÖZÇİFT; GÜLTEN, 2013), FDT (ÖZÇİFT; GÜLTEN, 2013) e *Naive Bayes* (ANBARASI; ANUPRIYA; IYENGAR, 2010). Nesses trabalhos, os autores também identificaram as características mais relevantes para cada contexto de aplicação, com um número reduzido de características.

Contudo, os estudos encontrados na literatura não envolveram uma combinação mais abrangente de técnicas (características, métodos de seleção e classificadores) com a flexibi-

lidade necessária para diferentes tipos de características e contextos. Por exemplo, Anbarasi, Anupriya e Yengar ((2010)) utilizaram o AG como seletor único para reduzir o conjunto de características e melhorar o diagnóstico em doenças cardíacas, assim como Gokulnath e Shantharajah ((2019)). Neste, os autores utilizaram três métodos de classificação para avaliar o desempenho: BPNN, KNN e CSSVM. Özçift E Gülten ((2013)) e Lu, Zhu e Gu ((2014)) associaram o AG a somente outro método, uma rede Bayesiana e MI (*Mutual Information*), respectivamente. Assim, não é possível afirmar que os bons resultados obtidos podem ser alcançados em outros contextos de interesse. Portanto, uma investigação direcionada para determinar um método capaz de combinar um número significativo de características, métodos de seleção e de classificação, usando-os como parâmetros para um AG, pode contribuir significativamente com o desenvolvimento de métodos de aprendizado de máquina e, consequentemente, sistemas CAD.

É importante destacar que são vários os contextos de imagens médicas que podem ser explorados para avaliar os sistemas computacionais que apoiam o diagnóstico médico, tais como os de imagens de linfomas e de câncer colorretal. A importância do estudo dessas doenças deve-se ao número de casos e à mortalidade. O método mais comum para o diagnóstico de ambos os tipos de câncer é a partir de biópsias dos tecidos que são corados com hematoxilina e eosina (H&E) e avaliados microscopicamente por médicos especialistas, uma tarefa que, além de desafiadora, pode consumir muito tempo do patologista (ORLOV et al., 2010).

O câncer colorretal é um tumor maligno que se desenvolve na parede interna do intestino grosso (cólon) ou reto (ALTERI; KRAMER; S, 2014). As principais motivações para o estudo são o número de casos e a mortalidade. Esse é o terceiro câncer mais comum em homens (746 mil casos) e o segundo mais comum em mulheres (614 mil casos) (SIEGEL; MILLER; JEMAL, 2020). O número de mortes foi 694 mil, com a maior incidência nas regiões menos desenvolvidas do mundo, a uma taxa de 52%. No Brasil, estima-se que, para cada ano do triênio de 2020-2022, haverá 20.520 casos de câncer de cólon e reto em homens e 20.470 em mulheres (INCA, 2019). Em 2020, projeta-se que 78.300 novos casos para homens e 69.650 novos para mulheres ocorram nos Estados Unidos: as mortes estimadas

são 28.630 para homens e 24.570 para mulheres (SIEGEL; MILLER; JEMAL, 2020).

O linfoma é um tipo de câncer que se origina no sistema linfático e, a partir do padrão de crescimento e das características citológicas das células, pode ser classificado em linfomas de Hodgkin (HL) e linfomas não-Hodgkin (NHL) (SIEGEL; MILLER; JEMAL, 2020; XERRI et al., 2016). Em 2020, estima-se que 8.848 novos casos do tipo HL e 77.240 casos do NHL sejam diagnosticados nos Estados Unidos, com aproximadamente 970 e 19.940 mortes, respectivamente (SIEGEL; MILLER; JEMAL, 2020). No Brasil, são esperados, para cada ano do triênio 2020-2020, 6.580 casos em homens e 5.450 em mulheres (INCA, 2019). Considerando sua maior incidência, o NHL é o principal foco de pesquisa. Os linfomas NHL são divididos em três grupos: MCL (*mantle cell lymphoma*), FL (*follicular lymphoma*) e CLL (*chronic lymphocytic leukemia*).

1.2 Objetivos

O objetivo principal deste trabalho foi explorar técnicas de aprendizado de máquina para desenvolver um método capaz de fornecer as melhores combinações de técnicas e características, a partir de um AG, para auxiliar na classificação e no reconhecimento de padrões de imagens médicas H&E. Especificamente, buscou-se definir:

1. Um modelo baseado em AG capaz de definir combinações entre características, métodos de seleção e classificadores para o reconhecimento de padrões em linfomas Não-Hodgkin e câncer colorretal;
2. Associações de métodos capazes de distinguir os grupos sob investigação, com desempenho relevante e o menor número de características, além do detalhamento sobre as características selecionadas;
3. Aplicar o modelo proposto em imagens H&E de câncer colorretal e linfomas, que apresentam altas taxas de mortalidade.
4. Diferentes testes com os parâmetros mais relevantes, tais como tamanhos de populações

e número de iterações, a fim de classificar os grupos benignos e malignos de imagens colorretais, bem como os grupos MCL, FL e CLL (linfomas Não-Hodgkin).

1.3 Organização do Trabalho

O presente trabalho foi estruturado em cinco capítulos, além da Introdução. No Capítulo 2, é apresentada a Fundamentação Teórica sobre alguns assuntos relacionados ao tema, tais como informações sobre os algoritmos genéticos, técnicas de seleção de características, classificadores e métricas para avaliação de desempenho. No Capítulo 3, encontra-se um detalhamento sobre os trabalhos relacionados, visando obter as informações necessárias para a estruturação do modelo. No Capítulo 4, estão descritas as etapas que embasam o método proposto. No Capítulo 5, os resultados são apresentados. Por fim, no Capítulo 6, as conclusões são apresentadas, assim como uma perspectiva sobre os trabalhos futuros.

Capítulo 6

Conclusão

Neste trabalho, foi desenvolvido um método baseado em Algoritmo Genético (AG) para encontrar a melhor combinação entre características, métodos de seleção e de classificação. A metodologia foi desenvolvida a partir do princípio de evolução apresentado por Charles Darwin, agregando um conjunto de indivíduos (chamado de população) que foi submetido a três etapas: definição da estrutura cromossômica e população inicial, avaliação e seleção da população e, por fim *crossover* e mutação. O método foi testado em imagens histológicas de câncer colorretal e linfomas NHL, considerando diferentes combinações de parâmetros. Foram analisadas variações da taxa de mutação (m), população P e número máximo de iterações $Iter$, permitindo encontrar soluções aceitáveis e com um número reduzido de características.

Para o conjunto de dados colorretal, a melhor solução foi obtida com 50 indivíduos e 50 iterações, com apenas 24 características, algoritmo de seleção *Relief* e classificador *Random Forest*. A área sob a curva ROC (AUC) foi de 0,984 e a taxa de mutação associada foi $m = 0,7$. Em relação ao conjunto de dados de linfomas, a melhor solução foi determinada com 500 indivíduos e 50 iterações, também com taxa de mutação $m = 0,7$. Nesse caso, o valor da AUC foi de 0,947. O modelo também conseguiu reduzir o conjunto de dados de 1.512 características para 73, aplicando o algoritmo de seleção *Relief* e o classificador *K-Nearest Neighbors*.

Ao considerar uma visão geral em relação aos estudos disponíveis na literatura, o método

proposto apresenta um importante diferencial: discrimina e detalha possíveis padrões de características indicadas para a classificação dos grupos de câncer sob investigação. O padrão identificado envolveu os descritores de Haralick, lacunaridade e percolação, todos obtidos de sub-imagens *curvelets*. O uso do método estatístico *Wilcoxon-Mann-Whitney* também foi aplicado para comparar os desempenhos do método proposto e os fornecidos por técnicas importantes e amplamente exploradas na área de aprendizado de máquina. Os resultados obtidos foram comparados com os fornecidos por redes DL (GIBSON; NICHOLSON; PATTERSON, 2016; KRIZHEVSKY; SUTSKEVER; HINTON, 2012; HE et al., 2016) e com o método descrito por Kotthoff et al. ((2017)), que utiliza a otimização Bayesiana para indicar as melhores soluções. Para o conjunto de dados de linfomas, o método proposto foi superior em todas comparações e, *Computer-Aided Diagnosis* (CAD) para o conjunto de dados colorretal, superior ao método (GIBSON; NICHOLSON; PATTERSON, 2016). Uma vantagem do método proposto está em indicar uma solução aceitável com a identificação das características mais relevantes para o processo de classificação. Dessa forma, acredita-se que os resultados são relevantes para o estudo e reconhecimento de padrões de linfomas não Hodgkin e câncer colorretal, além de contribuir para o desenvolvimento de sistemas *Computer-Aided Diagnosis* (CAD).

6.1 Contribuições obtidas

As principais contribuições com o desenvolvimento deste trabalho foram:

1. Um modelo que permite fornecer soluções aceitáveis para o reconhecimento de padrões de linfomas NHL e câncer colorretal, mesmo envolvendo inúmeras combinações de parâmetros, características, métodos de seleção e classificadores;
2. Identificação dos parâmetros mais relevantes, como tamanho da população e número de iterações, para classificar os grupos benigno e maligno de imagens colorretais, bem como os grupos MCL, FL e CLL de linfomas NHL;
3. Identificação de quantas e quais foram as características mais relevantes, bem como

suas associações, para investigar os grupos de câncer colorretal e os de linfomas NHL;

4. Indicações de combinações de técnicas que forneceram classificações com taxas melhores do que as obtidas via modelos *DL* e em otimização Bayesiana para definir as melhores soluções.

6.2 Trabalhos futuros

Em trabalhos futuros, pretende-se: 1) explorar o modelo para o reconhecimento de padrões em imagens de H&E de câncer de mama, com ou sem normalização dos corantes presentes nas lâminas. 2) realizar um levantamento das principais características que constituem as melhores soluções; 3) explorar mais variações dos parâmetros utilizados como entrada para o método, tais como t e $MaxF$, e identificar quais deles afetam significativamente os resultados. O parâmetro t permite aumentar o número de indivíduos que serão selecionados e passarão por cruzamento, criando novos indivíduos para avaliação. O parâmetro $MaxF$ possibilita verificar se um número inicial reduzido de características pode causar uma otimização mais rápida no método; 4) alterar a estrutura cromossômica do método para investigar o ajuste de hiperparâmetros dos classificadores. Essa alteração pode fornecer taxas de distinções ainda mais relevantes.

Referências

- AB WAHAB, Mohd Nadhir; NEFTI-MEZIANI, Samia; ATYABI, Adham. A comprehensive review of swarm optimization algorithms. **PloS one**, Public Library of Science, v. 10, n. 5, e0122827, 2015.
- ABAS, Fazly S et al. Computer-assisted quantification of CD3+ T cells in follicular lymphoma. **Cytometry Part A**, Wiley Online Library, v. 91, n. 6, p. 609–621, 2017.
- ABINASH, MJ; VASUDEVAN, V. A Hybrid Forward Selection Based LASSO Technique for Liver Cancer Classification. In: **NANOELECTRONICS, Circuits and Communication Systems**. [S.l.]: Springer, 2019. p. 185–193.
- AFEF, Latrach et al. Comparison study for computer assisted detection and diagnosis CADsystems dedicated to prostate cancer detection using MRImp modalities. In: **IEEE. ADVANCED Technologies for Signal and Image Processing (ATSIP)**, 2018 4th International Conference on. [S.l.: s.n.], 2018. p. 1–6.
- AGING, National Institute on. **National Institute on Aging**. [S.l.: s.n.], 2020. [⟨https://www.nia.nih.gov/⟩](https://www.nia.nih.gov/). (Accessed on 04/17/2020).
- AIN, Qurrat Ul et al. Genetic programming for feature selection and feature construction in skin cancer image classification. In: **SPRINGER. PACIFIC Rim International Conference on Artificial Intelligence**. [S.l.: s.n.], 2018. p. 732–745.
- ALDOJ, Nader et al. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. **European radiology**, Springer, v. 30, n. 2, p. 1243–1253, 2020.

ALTERI, R; KRAMER, J; S, Simpson. Colorectal Cancer Facts and Figures 2014-2016. **Atlanta: American Cancer Society**, p. 1–30, 2014.

AMORIM, Mauricio JV; BARONE, Dante; MANSUR, André Uebe. Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica. In: 1. BRAZILIAN Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). [S.l.: s.n.], 2008. v. 1, p. 666–674.

ANBARASI, M; ANUPRIYA, E; IYENGAR, NCSN. Enhanced prediction of heart disease with feature subset selection using genetic algorithm. **International Journal of Engineering Science and Technology**, v. 2, n. 10, p. 5370–5376, 2010.

ANITHA, V; MURUGAVALLI, S. Brain tumour classification using two-tier classifier with adaptive segmentation technique. **IET computer vision**, IET, v. 10, n. 1, p. 9–17, 2016.

EL-ANWAR, Mohammad Waheed et al. Anterior Ethmoidal Artery: A Computed Tomography Analysis and New Classifications. **Journal of Neurological Surgery Part B: Skull Base**, Georg Thieme Verlag KG, 2020.

APOSTOLOPOULOS, Ioannis D; MPESIANA, Tzani A. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. **Physical and Engineering Sciences in Medicine**, Nature Publishing Group, p. 1, 2020.

ARIMURA, Hidetaka et al. Computer-aided diagnosis systems for brain diseases in magnetic resonance images. **Algorithms**, Molecular Diversity Preservation International, v. 2, n. 3, p. 925–952, 2009.

ARLOT, Sylvain; CELISSE, Alain 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.

ATKESON, Christopher G; MOORE, Andrew W; SCHAAL, Stefan. Locally weighted learning for control. In: LAZY learning. [S.l.]: Springer, 1997. p. 75–113.

BABATUNDE, Oluleye H et al. A genetic algorithm-based feature selection. **IJECCE**, 2014.

BAI, Jie et al. NHL Pathological Image Classification Based on Hierarchical Local Information and GoogLeNet-Based Representations. **BioMed research international**, Hindawi, v. 2019, 2019.

BEJNORDI, Babak Ehteshami et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. **Jama**, American Medical Association, v. 318, n. 22, p. 2199–2210, 2017.

BHARDWAJ, Harshit et al. Breast Cancer Diagnosis using Simultaneous Feature Selection and Classification: A Genetic Programming Approach. In: IEEE. 2018 IEEE Symposium Series on Computational Intelligence (SSCI). [S.l.: s.n.], 2018. p. 2186–2192.

BRANCATI, Nadia et al. A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. **IEEE Access**, IEEE, v. 7, p. 44709–44720, 2019.

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

BRUDERER, Erhard; SINGH, Jitendra V. Organizational evolution, learning, and selection: A genetic-algorithm-based model. **Academy of management journal**, Academy of Management Briarcliff Manor, NY 10510, v. 39, n. 5, p. 1322–1349, 1996.

CANDÈS, Emmanuel J; DONOHO, David L. New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities. **Communications on pure and applied mathematics**, Wiley Online Library, v. 57, n. 2, p. 219–266, 2004.

CHAN, Heang-Ping et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms. **Arbor**, v. 1001, p. 48109–0326, 1990.

CHANDRASHEKAR, Girish; SAHIN, Ferat. A survey on feature selection methods. **Computers & Electrical Engineering**, Elsevier, v. 40, n. 1, p. 16–28, 2014.

CHENG, Jie-Zhi et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. **Scientific reports**, Nature Publishing Group, v. 6, p. 24454, 2016.

- CHOUAIB, Hassan et al. Feature selection combining genetic algorithm and adaboost classifiers. In: IEEE. PATTERN Recognition, 2008. ICPR 2008. 19th International Conference on. [S.l.: s.n.], 2008. p. 1–4.
- CLEARY, John G; TRIGG, Leonard E. K*: An instance-based learner using an entropic distance measure. In: MACHINE Learning Proceedings 1995. [S.l.]: Elsevier, 1995. p. 108–114.
- CODELLA, Noel et al. Lymphoma diagnosis in histopathology using a multi-stage visual learning approach. In: INTERNATIONAL SOCIETY FOR OPTICS e PHOTONICS. MEDICAL Imaging 2016: Digital Pathology. [S.l.: s.n.], 2016. v. 9791, 97910h.
- CONCI, Aura; AZEVEDO, Eduardo; LETA, Fabiana R. **Computação gráfica**. [S.l.]: Elsevier, 2008.
- CRISTIANO, Mariana Vitória de Menezes Bordalo. **Sensibilidade e especificidade na curva roc: um caso de estudo**. 2017. Tese (Doutorado).
- DAINESE, Renata Cilene. Sensoriamento remoto e geoprocessamento aplicado ao estudo temporal do uso da terra e na comparação entre classificação não supervisionada e análise visual. Universidade Estadual Paulista (UNESP), 2001.
- DENNIS, Binu; MUTHUKRISHNAN, Selvi. AGFS: Adaptive Genetic Fuzzy System for medical data classification. **Applied Soft Computing**, Elsevier, v. 25, p. 242–252, 2014.
- DIETTERICH, Tom. Overfitting and undercomputing in machine learning. **ACM computing surveys (CSUR)**, ACM, v. 27, n. 3, p. 326–327, 1995.
- DOI, Kunio; HUANG, HK. Computer-aided diagnosis (CAD) and image-guided decision support. **Computerized Medical Imaging and Graphics**, v. 4, n. 31, p. 195–197, 2007.
- DOYLE, Scott et al. A boosting cascade for automated detection of prostate cancer from digitized histology. In: SPRINGER. INTERNATIONAL Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.: s.n.], 2006. p. 504–511.
- EIBEN, Agoston E; SMITH, James E et al. **Introduction to evolutionary computing**. [S.l.]: Springer, 2003. v. 53.

ELTOUKHY, Mohamed Meselhy; FAYE, Ibrahima; SAMIR, Brahim Belhaouari. A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. **Computers in biology and medicine**, Elsevier, v. 42, n. 1, p. 123–128, 2012.

ESTEVA, Andre et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, Nature Publishing Group, v. 542, n. 7639, p. 115, 2017.

FARID, Ahmed Abdullah; SELIM, Gamal; KHATER, Hatem. A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm For Breast Cancer Detection, 2020.

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

FAY, Michael P; PROSCHAN, Michael A. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. **Statistics surveys**, NIH Public Access, v. 4, p. 1, 2010.

FUREY, Terrence S et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. **Bioinformatics**, Oxford University Press, v. 16, n. 10, p. 906–914, 2000.

FURTADO, Joao Carlos; LORENA, Luiz Antonio Nogueira. Algoritmo genético construtivo na otimização de problemas combinatoriais de agrupamentos. **São José dos Campos. Tese (Doutorado em Computação Aplicada)–Instituto Nacional de Pesquisas Espaciais**, 1998.

GABRYS, Bogdan; RUTA, Dymitr. Genetic algorithms in classifier fusion. **Applied soft computing**, Elsevier, v. 6, n. 4, p. 337–347, 2006.

GARDNER, Matt W; DORLING, SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. **Atmospheric environment**, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998.

GIANNINI, Valentina et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. **European radiology**, Springer, v. 27, n. 10, p. 4200–4208, 2017.

GIBSON, Adam; NICHOLSON, Chris; PATTERSON, Josh. **Eclipse Deeplearning4j Development Team. Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0**. [S.l.: s.n.], 2016.

<http://deeplearning4j.org>. Accessed: 2019-12-10.

GNANA, D Asir Antony; APPAVU, S; LEAVLINE, E Jebamalar. Literature review on feature selection methods for high-dimensional data. **methods**, v. 136, n. 1, 2016.

GOKULNATH, Chandra Babu; SHANTHARAJAH, SP. An optimized feature selection based on genetic approach and support vector machine for heart disease. **Cluster Computing**, Springer, v. 22, n. 6, p. 14777–14787, 2019.

GOU, Jianping et al. A generalized mean distance-based k-nearest neighbor classifier. **Expert Systems with Applications**, Elsevier, v. 115, p. 356–372, 2019.

GU, Jiuxiang et al. Recent advances in convolutional neural networks. **Pattern Recognition**, Elsevier, v. 77, p. 354–377, 2018.

GUIMARÃES, M Carolina S. Exames de laboratório: sensibilidade, especificidade, valor preditivo positivo. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 18, n. 2, p. 117–120, 1985.

GURCAN, Metin N et al. Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system. **Medical Physics**, Wiley Online Library, v. 29, n. 11, p. 2552–2558, 2002.

GUYON, Isabelle; ELISSEEFF, André. An introduction to variable and feature selection. **Journal of machine learning research**, v. 3, Mar, p. 1157–1182, 2003.

HALL, Mark et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, ACM, v. 11, n. 1, p. 10–18, 2009.

HANLEY, James A; MCNEIL, Barbara J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. **Radiology**, v. 143, n. 1, p. 29–36, 1982.

HARALICK, Robert M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, IEEE, v. 67, n. 5, p. 786–804, 1979.

HE, Kaiming et al. Deep residual learning for image recognition. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. p. 770–778.

HERMINE, Olivier et al. Addition of high-dose cytarabine to immunochemotherapy before autologous stem-cell transplantation in patients aged 65 years or younger with mantle cell lymphoma (MCL Younger): a randomised, open-label, phase 3 trial of the European Mantle Cell Lymphoma Network. **The Lancet**, Elsevier, v. 388, n. 10044, p. 565–575, 2016.

HO, WT; LAM, PWT. Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. **Clinical Radiology**, Elsevier, v. 58, n. 2, p. 133–136, 2003.

HOLLAND, John H. Genetic algorithms. **Scientific american**, JSTOR, v. 267, n. 1, p. 66–73, 1992.

HUA, Kai-Lung et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. **OncoTargets and therapy**, Dove Press, v. 8, 2015.

HUANG, Jin; LING, Charles X. Using AUC and accuracy in evaluating learning algorithms. **IEEE Transactions on knowledge and Data Engineering**, IEEE, v. 17, n. 3, p. 299–310, 2005.

HUANG, Peng; PARK, Seyoun et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. **Radiology**, Radiological Society of North America, v. 286, n. 1, p. 286–295, 2017.

INCA. **Estimativa — 2020 Incidência de Câncer no Brasil**. Rio de Janeiro, Brazil, 2019.

INSTITUTE, National Cancer. **Comprehensive Cancer Information - National Cancer Institute**. [S.l.: s.n.], 2020. <<https://www.cancer.gov/>>. (Accessed on 04/17/2020).

INZA, Iñaki et al. Filter versus wrapper gene selection approaches in DNA microarray domains. **Artificial intelligence in medicine**, Elsevier, v. 31, n. 2, p. 91–103, 2004.

IOPE, R; LEMKE, Ney; WINCKLER, G von. GridUNESP: a multi-campus Grid infrastructure for scientific computing. In: PROCEEDINGS of the 3rd Latin American Conference on High Performance Computing (CLCAR 2010); Gramado: 25-28 August. [S.l.: s.n.], 2010. p. 76–84.

IVANOVICI, M; RICHARD, N; DECEAN, H. Fractal dimension and lacunarity of psoriatic lesions-A colour approach. **medicine**, v. 6, n. 4, p. 7, 2009.

JAFFAR, M Arfan; SIDDIQUI, Abdul Basit; MUSHTAQ, Mubashar. Ensemble classification of pulmonary nodules using gradient intensity feature descriptor and differential evolution. **Cluster Computing**, Springer, v. 21, n. 1, p. 393–407, 2018.

JALALIAN, Afsaneh et al. Computer-assisted diagnosis system for breast cancer in computed tomography laser mammography (ctlm). **Journal of digital imaging**, Springer, v. 30, n. 6, p. 796–811, 2017.

JANSSEN, Ronald J; MOURÃO-MIRANDA, Janaina; SCHNACK, Hugo G. Making individual prognoses in psychiatry using neuroimaging and machine learning. **Biological Psychiatry: Cognitive Neuroscience and Neuroimaging**, Elsevier, 2018.

JAYACHANDRAN, Srinath; GHOSH, Ashlin. Deep Transfer Learning for Texture Classification in Colorectal Cancer Histology. **arXiv preprint arXiv:2004.01614**, 2020.

JIANG, Shancheng et al. Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. **Expert Systems with Applications**, Elsevier, v. 82, p. 216–230, 2017.

JOHN, George H; LANGLEY, Pat. Estimating continuous distributions in Bayesian classifiers. In: MORGAN KAUFMANN PUBLISHERS INC. PROCEEDINGS of the Eleventh conference on Uncertainty in artificial intelligence. [S.l.: s.n.], 1995. p. 338–345.

JØRGENSEN, Alex Skovsbo et al. Using cell nuclei features to detect colon cancer tissue in hematoxylin and eosin stained slides. **Cytometry Part A**, Wiley Online Library, v. 91, n. 8, p. 785–793, 2017.

KALKAN, Habil et al. Automated classification of local patches in colon histopathology. In: IEEE. PATTERN Recognition (ICPR), 2012 21st International Conference on. [S.l.: s.n.], 2012. p. 61–64.

KARNAN, M; LOGHESHWARI, T. Improved implementation of brain MRI image segmentation using ant colony system. In: IEEE. 2010 IEEE International Conference on Computational Intelligence and Computing Research. [S.l.: s.n.], 2010. p. 1–4.

KATHER, Jakob Nikolas et al. Multi-class texture analysis in colorectal cancer histology. **Scientific reports**, Nature Publishing Group, v. 6, p. 27988, 2016.

KEČO, Dino; SUBASI, Abdulhamit; KEVRIC, Jasmin. Cloud computing-based parallel genetic algorithm for gene selection in cancer classification. **Neural Computing and Applications**, Springer, v. 30, n. 5, p. 1601–1610, 2018.

KIM, Ji-Hyun. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. **Computational statistics & data analysis**, Elsevier, v. 53, n. 11, p. 3735–3745, 2009.

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

KOIZUMI, Mitsuru et al. Diagnostic performance of a computer-assisted diagnosis system for bone scintigraphy of newly developed skeletal metastasis in prostate cancer patients: search for low-sensitivity subgroups. **Annals of nuclear medicine**, Springer, v. 31, n. 7, p. 521–528, 2017.

KOMINAMI, Yoko et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. **Gastrointestinal endoscopy**, Elsevier, v. 83, n. 3, p. 643–649, 2016.

- KOOI, Thijs et al. Large scale deep learning for computer aided detection of mammographic lesions. **Medical image analysis**, Elsevier, v. 35, p. 303–312, 2017.
- KOTSIANTIS, Sotiris B; ZAHARAKIS, I; PINTELAS, P. Supervised machine learning: A review of classification techniques. **Emerging artificial intelligence applications in computer engineering**, v. 160, p. 3–24, 2007.
- KOTTHOFF, Lars et al. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. **The Journal of Machine Learning Research**, JMLR. org, v. 18, n. 1, p. 826–830, 2017.
- KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: ADVANCES in neural information processing systems. [S.l.: s.n.], 2012. p. 1097–1105.
- KUMAR, G Ravi; RAMACHANDRA, GA; NAGAMANI, K. An efficient feature selection system to integrating svm with genetic algorithm for large medical datasets. **International Journal**, v. 4, n. 2, 2014.
- KUNCHEVA, Ludmila I; JAIN, Lakhmi C. Designing classifier fusion systems by genetic algorithms. **IEEE Transactions on Evolutionary computation**, IEEE, v. 4, n. 4, p. 327–336, 2000.
- KUO, Bor-Chen et al. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, IEEE, v. 7, n. 1, p. 317–326, 2014.
- LEE, Jung Hwan et al. Spine Computed Tomography to Magnetic Resonance Image Synthesis Using Generative Adversarial Networks: A Preliminary Study. **Journal of Korean Neurosurgical Society**, Korean Neurosurgical Society, 2020.
- LEITE, Mariana et al. Etiology-based classification of brain white matter hyperintensity on magnetic resonance imaging. **Journal of Medical Imaging**, International Society for Optics e Photonics, v. 2, n. 1, p. 014002, 2015.

LI, Wenyuan et al. Path R-CNN for prostate cancer diagnosis and gleason grading of histological images. **IEEE transactions on medical imaging**, IEEE, v. 38, n. 4, p. 945–954, 2018.

LICENSE, GNU General Public. GNU General Public License. **Retrieved December**, v. 25, p. 2014, 1989.

LINDEN, Ricardo. **Algoritmos genéticos: uma importante ferramenta da inteligência computacional**. [S.l.]: Brasport, 2006.

LIU, Liping et al. Modified Cuckoo Search Algorithm with Variational Parameters and Logistic Map. **Algorithms**, Multidisciplinary Digital Publishing Institute, v. 11, n. 3, p. 30, 2018.

LOPES, Uilian Kenedi. Redes neurais convolucionais aplicadas ao diagnóstico de tuberculose por meio de imagens radiológicas. Universidade do Vale do Rio dos Sinos, 2017.

LU, Chunhong; ZHU, Zhaomin; GU, Xiaofeng. An intelligent system for lung cancer diagnosis using a new genetic algorithm based feature selection method. **Journal of medical systems**, Springer, v. 38, n. 9, p. 97, 2014.

LUCAS, Diogo C. Algoritmos genéticos: uma introdução. **Universidade Federal do Rio Grande do Sul**, v. 24, 2002.

MARTINEZ, Edson Zangiacomi; LOUZADA-NETO, Francisco;

PEREIRA, Basilio de Bragança. A curva ROC para testes diagnósticos. **Cad. saúde colet.,(Rio J.)**, v. 11, n. 1, p. 7–31, 2003.

MELLENDEZ, Jaime et al. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. **IEEE Trans. Med. Imaging**, v. 34, n. 1, p. 179–192, 2015.

MIRANDA, Márcio Nunes de. **Algoritmos Genéticos: Fundamentos e Aplicações**. [S.l.: s.n.], 2007.

MISAWA, Masashi et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. **Gastroenterology**, Elsevier, v. 150, n. 7, p. 1531–1532, 2016.

MITCHELL, Tom M et al. Machine learning. 1997. **Burr Ridge, IL: McGraw Hill**, v. 45, n. 37, p. 870–877, 1997.

MOHAMED, Ali W; SABRY, Hegazy Z; KHORSHID, Motaz. An alternative differential evolution algorithm for global optimization. **Journal of advanced research**, Elsevier, v. 3, n. 2, p. 149–165, 2012.

MUNI, Durga Prasad; PAL, Nikhil R; DAS, Jyotirmay. Genetic programming for simultaneous feature selection and classifier design, 2006.

NAGARAJAN, G et al. Hybrid genetic algorithm for medical image feature extraction and selection. **Procedia Computer Science**, Elsevier, v. 85, p. 455–462, 2016.

NAIYAR, Madeeha; ASIM, Yousra; SHAHID, Aqsa. Automated colon cancer detection using structural and morphological features. In: IEEE. FRONTIERS of Information Technology (FIT), 2015 13th International Conference on. [S.l.: s.n.], 2015. p. 240–245.

NIKOLAIDIS, NS; NIKOLAIDIS, IN; TSOUROUROS, CC. A Variation of the Box-Counting Algorithm Applied to Colour Images. **arXiv preprint arXiv:1107.2336**, 2011.

ORLOV, Nikita V et al. Automatic classification of lymphoma images with transform-based global features. **IEEE Transactions on Information Technology in Biomedicine**, IEEE, v. 14, n. 4, p. 1003–1013, 2010.

ÖZÇİFT, Akin; GÜLTEN, Arif. Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases. **Digital Signal Processing**, Elsevier, v. 23, n. 1, p. 230–237, 2013.

PARIKH, Jay; TICKMAN, Ronald. Image-Guided Tissue Sampling: Where Radiology Meets Pathology. **The breast journal**, Wiley Online Library, v. 11, n. 6, p. 403–409, 2005.

- PATIDAR, Shivnarayan; PACHORI, Ram Bilas; ACHARYA, U Rajendra. Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals. **Knowledge-Based Systems**, Elsevier, v. 82, p. 1–10, 2015.
- PAUL, Desbordes et al. Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. **Computerized Medical Imaging and Graphics**, Elsevier, v. 60, p. 42–49, 2017.
- PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **Journal of machine learning research**, v. 12, Oct, p. 2825–2830, 2011.
- PICOLI, Ivan Luiz. Uma abordagem de classificação não supervisionada de cargas de trabalho de sistemas analíticos em Apache Hadoop através de análise de log, 2013.
- PRAVEEN, GB; AGRAWAL, Anita. Hybrid approach for brain tumor detection and classification in magnetic resonance images. In: IEEE. COMMUNICATION, Control and Intelligent Systems (CCIS), 2015. [S.l.: s.n.], 2015. p. 162–166.
- AL-RAJAB, Murad; LU, Joan; XU, Qiang. Examining applying high performance genetic data feature selection and classification algorithms for colon cancer diagnosis. **Computer methods and programs in biomedicine**, Elsevier, v. 146, p. 11–24, 2017.
- RAJPURKAR, Pranav et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. **arXiv preprint arXiv:1711.05225**, 2017.
- RATHORE, Saima et al. Classification of colon biopsy images based on novel structural features. In: IEEE. 2013 IEEE 9th International Conference on Emerging Technologies (ICET). [S.l.: s.n.], 2013. p. 1–6.
- EL-REGAILY, Salsabil A et al. Survey of Computer Aided Detection Systems for Lung Cancer in Computed Tomography. **Current Medical Imaging Reviews**, Bentham Science Publishers, v. 14, n. 1, p. 3–18, 2018.
- REMAMANY, Krishna Priya et al. Brain tumor segmentation in MRI images using integrated modified PSO-fuzzy approach. **Int. Arab J. Inf. Technol.**, v. 12, 6A, p. 797–805, 2015.

RIBEIRO, Matheus Gonçalves et al. Analysis of the Influence of Color Normalization in the Classification of Non-Hodgkin Lymphoma Images. In: IEEE. 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). [S.l.: s.n.], 2018. p. 369–376.

RIBEIRO, Matheus Gonçalves et al. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. **Expert Systems with Applications**, v. 120, p. 262–278, 2019. ISSN 0957-4174. DOI:

<https://doi.org/10.1016/j.eswa.2018.11.034>. Disponível em:

<http://www.sciencedirect.com/science/article/pii/S0957417418307541>.

ROBERTO, Guilherme F et al. Features based on the percolation theory for quantification of non-Hodgkin lymphomas. **Computers in biology and medicine**, Elsevier, v. 91, p. 135–147, 2017.

RODRIGUES, Diego Alves. Deep Learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento automotivo. Universidade Federal da Paraíba, 2018.

ROSENBLATT, Frank. Principles of neurodynamics. Spartan Book, 1962.

SAEYS, Yvan; INZA, Iñaki; LARRAÑAGA, Pedro. A review of feature selection techniques in bioinformatics. **bioinformatics**, Oxford University Press, v. 23, n. 19, p. 2507–2517, 2007.

SAFAVIAN, S Rasoul; LANDGREBE, David. A survey of decision tree classifier methodology. **IEEE transactions on systems, man, and cybernetics**, IEEE, v. 21, n. 3, p. 660–674, 1991.

SAHINER, Berkman; CHAN, Heang-Ping; HADJIISKI, Lubomir. Classifier performance prediction for computer-aided diagnosis using a limited dataset. **Medical Physics**, Wiley Online Library, v. 35, n. 4, p. 1559–1570, 2008.

SHAMIR, Lior et al. IICBU 2008: a proposed benchmark suite for biological image analysis. **Medical & biological engineering & computing**, Springer, v. 46, n. 9, p. 943–947, 2008.

- SHARMA, Puneet et al. **Method and system for non-invasive assessment of coronary artery disease**. [S.l.]: Google Patents, set. 2015. US Patent 9,119,540.
- SIEGEL, Rebecca L; MILLER, Kimberly D; JEMAL, Ahmedin. Cancer statistics, 2020. **CA: A Cancer Journal for Clinicians**, Wiley Online Library, v. 70, n. 1, p. 7–30, 2020.
- SILVA, Giovanni da et al. Classification of malignancy of lung nodules in CT images using Convolutional Neural Network. In: SBC. ANAIS do XVI Workshop de Informática Médica. [S.l.: s.n.], 2020. p. 21–29.
- SINGH, Siddharth et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: a systematic review and meta-analysis of individual participant data. **Clinical Gastroenterology and Hepatology**, Elsevier, v. 13, n. 3, p. 440–451, 2015.
- SIRINUKUNWATTANA, Korsuk et al. Gland segmentation in colon histology images: The glas challenge contest. **Medical image analysis**, Elsevier, v. 35, p. 489–502, 2017.
- SMITH, Matthew G; BULL, Larry. Genetic programming with a genetic algorithm for feature construction and selection. **Genetic Programming and Evolvable Machines**, Springer, v. 6, n. 3, p. 265–281, 2005.
- SONG, Yang et al. Low dimensional representation of fisher vectors for microscopy image classification. **IEEE transactions on medical imaging**, IEEE, v. 36, n. 8, p. 1636–1649, 2017.
- SWERDLOW, Steven H; COOK, James R. As the world turns, evolving lymphoma classifications—past, present and future. **Human Pathology**, Elsevier, v. 95, p. 55–77, 2020.
- TAGUCHI, Hiroshi et al. Lung cancer detection based on helical CT images using curved-surface morphology analysis. In: INTERNATIONAL SOCIETY FOR OPTICS e PHOTONICS. MEDICAL Imaging 1999: Image Processing. [S.l.: s.n.], 1999. v. 3661, p. 1307–1315.
- TAPE, Thomas G. The area under an ROC curve. **Interpreting Diagnostic Tests [cited 2016 Nov 20]**. Available from: <http://gim.unmc.edu/dxtests/roc3.htm>, 2006.

TCHRAKIAN, Nairi et al. Quantitative image analysis of the MIB-1 proliferation index in Non-Hodgkin Lymphoma. **Journal of Histology & Histopathology**, Herbert Publications, v. 5, n. 1, p. 3, 2018.

TIWARI, Arti; SRIVASTAVA, Shilpa; PANT, Millie. Brain tumor segmentation and classification from magnetic resonance images: review of selected methods from 2014 to 2019. **Pattern Recognition Letters**, Elsevier, v. 131, p. 244–260, 2020.

TIWARI, Puneet; SACHDEVA, Jainy et al. Computer aided diagnosis system-a decision support system for clinical diagnosis of brain tumours. **International Journal of Computational Intelligence Systems**, Directory of Open Access Journals, v. 10, n. 1, p. 104, 2017.

VAN GINNEKEN, Bram; ROMENY, BM Ter Haar; VIERGEVER, Max A. Computer-aided diagnosis in chest radiography: a survey. **IEEE Transactions on medical imaging**, IEEE, v. 20, n. 12, p. 1228–1241, 2001.

VAN ROSSUM, Guido; DRAKE, Fred L. **Python language reference manual**. [S.l.]: Network Theory United Kingdom, 2003.

VAPNIK, Vladimir. **The nature of statistical learning theory**. [S.l.]: Springer science & business media, 2013.

VAPNIK, Vladimir N. An overview of statistical learning theory. **IEEE transactions on neural networks**, IEEE, v. 10, n. 5, p. 988–999, 1999.

VELU, CM; KASHWAN, KR. Visual data mining techniques for classification of diabetic patients. In: IEEE. ADVANCE Computing Conference (IACC), 2013 IEEE 3rd International. [S.l.: s.n.], 2013. p. 1070–1075.

VRJI, KS Angel; JAYAKUMARI, J. Automatic detection of brain tumor based on magnetic resonance image using CAD System with watershed segmentation. In: IEEE. SIGNAL Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on. [S.l.: s.n.], 2011. p. 145–150.

- WANG, Pin et al. Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. **Biomedical Signal Processing and Control**, Elsevier, v. 57, p. 101789, 2020.
- WANG, Xiaosong et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE. COMPUTER Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. [S.l.: s.n.], 2017. p. 3462–3471.
- WEISS, Sholom M; KAPOULEAS, Ioannis. An empirical comparison of pattern recognition, neural nets and machine learning classification methods. **Readings in machine learning**, p. 177–183, 1990.
- WELIKALA, RA et al. Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. **Computerized Medical Imaging and Graphics**, Elsevier, v. 43, p. 64–77, 2015.
- WHITLEY, Darrell. A genetic algorithm tutorial. **Statistics and computing**, Springer, v. 4, n. 2, p. 65–85, 1994.
- WILCOXON, Frank. Individual comparisons by ranking methods. In: BREAKTHROUGHS in statistics. [S.l.]: Springer, 1992. p. 196–202.
- XERRI, Luc et al. The heterogeneity of follicular lymphomas: from early development to transformation. **Virchows Archiv**, Springer, v. 468, n. 2, p. 127–139, 2016.
- YANG, Jihoon; HONAVAR, Vasant. Feature subset selection using a genetic algorithm. In: FEATURE extraction, construction and selection. [S.l.]: Springer, 1998. p. 117–136.
- YANG, Xin-She; DEB, Suash. Cuckoo search via Lévy flights. In: IEEE. 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC). [S.l.: s.n.], 2009. p. 210–214.
- YU, Songyang; GUAN, Ling. A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. **IEEE transactions on medical imaging**, IEEE, v. 19, n. 2, p. 115–126, 2000.

ZEZOS, Petros et al. 439-Toward Computer-Based Automated Mayo Score Classification in Ulcerative Colitis through Classical and Deep Machine Learning. **Gastroenterology**, Elsevier, v. 154, n. 6, s–99, 2018.

ZHANG, Xin et al. Design, synthesis and evaluation of genistein-polyamine conjugates as multi-functional anti-Alzheimer agents. **Acta pharmaceutica sinica B**, Elsevier, v. 5, n. 1, p. 67–73, 2015.

ZHI, Hui; LIU, Sanyang. Face recognition based on genetic algorithm. **Journal of Visual Communication and Image Representation**, Elsevier, v. 58, p. 495–502, 2019.

ZHUO, Li'an et al. Calibrated Stochastic Gradient Descent for Convolutional Neural Networks, 2019.