

Milton Martins de Oliveira Junior
Orientador: Prof. Dr. Antônio Oseas de Carvalho Filho

**Método Automático para Classificação de
Imagens Histopatológicas da Mama Utilizando
Transfer Learning e Análise de Textura**

Picos - PI
Dezembro de 2019

Milton Martins de Oliveira Junior
Orientador: Prof. Dr. Antônio Oseas de Carvalho Filho

**Método Automático para Classificação de Imagens
Histopatológicas da Mama Utilizando *Transfer Learning*
e Análise de Textura**

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação.

Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
Dezembro de 2019

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

- O48** Oliveira Junior, Milton Martins de.
Método automático para classificação de imagens histopatológicas da mama utilizando Transfer Learning e análise de textura. / Milton Martins de Oliveiras Junior. -- Picos,PI, 2019.
44 f.
CD-ROM: 4 ¾ pol.
- Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação). – Universidade Federal do Piauí, Picos, 2020.
“Orientador(A): Prof. Dr. Antônio Oseas de Carvalho Filho.”
1. Câncer de Mama – Tecnologia. 2. Imagens Histopatológicas. 3. Transfer Learning. I. Título.
- CDD 005.1**

Elaborada por Rafael Gomes de Sousa CRB 3/1163

MÉTODO AUTOMÁTICO PARA CLASSIFICAÇÃO DE IMAGENS HISTOPATOLÓGICAS
DA MAMA UTILIZANDO TRANSFER LEARNING E ANÁLISE DE TEXTURA

MILTON MARTINS DE OLIVEIRA JUNIOR

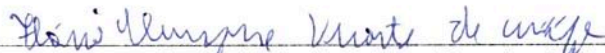
Monografia Aprovada como exigência parcial para obtenção do
grau de Bacharel em Sistemas de Informação.

Data de Aprovação

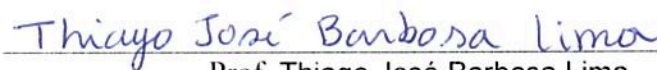
Picos - PI, 02 de dezembro de 2019



Prof. Antônio Oseas de Carvalho Filho



Prof. Flávio Henrique Duarte de Araújo



Prof. Thiago José Barbosa Lima

Agradecimentos

Agradeço primeiramente à Deus, pela vitória conquistada.

Aos meus pais, Milton Martins de Oliveira e Rosa Maria de Sousa Oliveira pelo incentivo e apoio durante toda a graduação.

A minha avó, Benedita Maria de Sousa e meu avô Manoelino Martins de Oliveira (*in memoriam*), que foram exemplos de caráter e dignidade.

Ao meu avô Inácio de Paula Sousa, por todo o apoio e por que me ensinar valores importantes para toda a vida.

A todos os meus tios(as) e primos(as) pelo amparo e assistência familiar.

Agradeço à minha namorada Taynan de Lima Silva que esteve ao meu lado durante o percurso acadêmico.

Agradeço também a todos os meus amigos e colegas, que foram extremamente importantes e ofereceram apoio durante o curso.

Ao meu orientador Prof. Dr. Antonio Oseas de Carvalho Filho por se mostrar sempre disponível para retirar minhas dúvidas e me auxiliar no desenvolvimento deste trabalho, proporcionando-me oportunidades de crescimento acadêmico.

Por fim, a todos que de forma direta ou indireta contribuíram para a realização deste sonho.

Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito.

Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.

Marthin Luther King

Resumo

O câncer de mama é uma doença causada pela multiplicação desordenada de células mamárias anormais, formando massas. Essa doença é uma das principais causas de mortalidade entre as mulheres. A detecção precoce é essencial para que o tratamento comece o mais cedo possível, proporcionando melhores taxas de sobrevivência entre pacientes. A triagem de doenças através de imagens histopatológicas pode ser um meio promissor para a discriminação do tecido examinado. Nos últimos anos, vêm crescendo o número de sistemas de Detecção Auxiliada por Computador (CAD). Estes sistemas sofisticados são cruciais para complementar as análises feitas pelos especialistas. Neste trabalho, apresentaremos uma abordagem computacional para classificar tecidos mamários normais ou tumorais por meio de imagens histopatológicas. Foi utilizado o *dataset* fornecido pela Breast Cancer Histology ICIAR 2018, que contém 400 imagens de biópsias da mama. Houve a extração de características utilizando abordagens de *Transfer Learning* (ResNet50, VGG16 e VGG19) e índices de diversidade atuando como descritores de textura, como também a concatenação dos melhores conjuntos. Os atributos extraídos foram utilizados para classificação nos algoritmos *Random Forest* e *Support Vector Machine*. Os resultados alcançaram uma acurácia de 91,75%, mostrando ser uma metodologia promissora para a classificação de tumores e câncer de mama.

Palavras-chaves: Câncer de Mama. Imagens Histopatológicas. Índices de Diversidade. *Transfer Learning*.

Abstract

Breast cancer is a disease caused by the disordered multiplication of abnormal breast cells, forming masses. This disease is a major cause of mortality among women. Early detection is essential for treatment to begin as early as possible, providing improved survival rates among patients. Screening for disease through histopathological imaging may be a promising means for discriminating the tissue examined. In recent years, the number of Computer Aided Detection (CAD) systems has been growing. These sophisticated systems are crucial in complementing expert analysis. In this paper, we will present a computational approach to classify normal or tumoral breast tissues through histopathological images. We used the dataset provided by Breast Cancer Histology ICIAR 2018, which contains 400 breast biopsy images. Characteristics were extracted using Transfer Learning approaches (ResNet50, VGG16 and VGG19) and diversity indices acting as texture descriptors, as well as the concatenation of the best sets. The extracted attributes were used for classification in the Random Forest and Support Vector Machine algorithms. The results reached an accuracy of 91.75%, showing to be a promising methodology for the classification of tumors and breast cancer.

Lista de ilustrações

Figura 1 – Ao lado esquerdo, o processo de realização de duas tarefas de <i>Machine Learning</i> . Na direita, o <i>Transfer Learning</i> sendo utilizado para passar o conhecimento da tarefa de origem, para a de destino.	22
Figura 2 – Representação resumida da metodologia proposta.	31
Figura 3 – Exemplos da Base de imagens BACH ICIAR 2018.	32

Lista de tabelas

Tabela 1 – Interpretação dos elementos da ecologia dos índices de diversidade no Processamento de Imagens	24
Tabela 2 – Resumo dos trabalhos relacionados.	29
Tabela 3 – Matriz de confusão	35
Tabela 4 – Interpretação de concordância para os valores do índice Kappa.	36
Tabela 5 – Tempo aproximado de extração de característica de uma imagem em segundos.	37
Tabela 6 – Resultados com <i>Random Forest</i>	37
Tabela 7 – Resultados com <i>Support Vector Machine</i> com $C=1$	38
Tabela 8 – Resultados com <i>Support Vector Machine</i> com $C=10$	38
Tabela 9 – Resultados dos classificadores com as concatenação dos atributos da ResNet50 e dos índices de diversidade.	39
Tabela 10 – Comparação do método proposto com alguns trabalhos relacionados.	40

Lista de abreviaturas e siglas

A	Acurácia
CAD	Detecção Auxiliada por Computador
CADx	Diagnóstico Auxiliada por Computador
Colab	<i>Google Colab</i>
CNN	Rede Neural Convolutiva
E	Especificidade
FN	Falso Negativo
FP	Falso Positivo
k	Índice de Concordância Kappa
PDI	Processamento Digital de Imagens
RBF	Função de Base Radial
ROC	Área sobre a Curva
S	Sensibilidade
SVM	<i>Support Vector Machine</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

Lista de símbolos

\arctan	Arco da tangente
i	Espécie
D^1	Índice de Simpson
D^2	Índice de Margalef
D^3	Índice de McIntosh
E^1	Índice de Camargo
E^2	Índice Smith & Wilson B
E^3	Índice de Smith & Wilson 1-D
E^4	Índice de Smith & Wilson $-\ln D$
H'	Índice Shannon
\ln	Logaritmo
N	Total de indivíduos
n_i	Número de indivíduos
π	Pi
p_i	Proporção de indivíduos da espécies i
p_j	Proporção de indivíduos da espécies j
Sp	Total de espécies da espécies i
S_{min}	Índice de Shannon <i>minimum</i>
Σ	Somatório
U	Distância Euclidiana

Sumário

1	Introdução	14
1.1	Objetivos	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
1.2	Organização do Trabalho	16
2	Referencial Teórico	17
2.1	Câncer de Mama	17
2.2	Processamento Digital de Imagens	17
2.2.1	Aquisição das Imagens	18
2.2.2	Pré-Processamento	18
2.2.3	Segmentação	18
2.2.4	Extração de Características	19
2.2.5	Classificação/Reconhecimento	19
2.3	<i>Deep Learning</i>	20
2.3.1	<i>Redes Neurais Convolucionais</i>	20
2.3.2	<i>Transfer Learning</i>	21
2.3.2.1	VGGNets	22
2.3.2.2	ResNet50	23
2.4	Índices de Diversidade	23
2.4.0.1	Índice de Shannon	24
2.4.0.2	Índice de Simpson	24
2.4.0.3	Índice de Margalef	25
2.4.0.4	Índice de McIntosh	25
2.4.0.5	Índice de Camargo	25
2.4.0.6	Índice de Smith & Wilson B	26
2.4.0.7	Índice de Smith & Wilson 1-D	26
2.4.0.8	Índice de Smith & Wilson $-lnD$	26
2.4.0.9	Índice de Shannon <i>maximum</i>	26
2.4.0.10	Índice de Shannon <i>minimum</i>	27
3	Trabalhos Relacionados	28
4	Metodologia	31
4.1	Aquisição de Imagens	32
4.2	Extração de Características	33

4.2.1	<i>Transfer Learning</i>	33
4.2.2	Índices de Diversidade Ecológicos	33
4.3	Classificação	34
4.3.1	<i>Random Forest</i>	34
4.3.2	<i>Support Vector Machine</i>	34
4.4	Métricas de Validação	35
4.4.1	Acurácia	35
4.4.2	Sensibilidade	36
4.4.3	Especificidade	36
4.4.4	Kappa	36
4.4.5	Área Sobre a Curva	36
5	Resultados e Discussões	37
6	Conclusão	41
	Referências	42

1 Introdução

O Câncer de mama trata-se de uma doença causada pela multiplicação desordenada das células mamárias. Isto acaba acarretando na geração de células anormais, que quando se multiplicam, transformam-se em um tumor.

Segundo o Instituto Nacional do Câncer (INCA), o câncer de mama é o mais comum entre as mulheres no mundo e no Brasil, depois do de pele não melanoma, correspondendo a cerca de 25% dos casos novos a cada ano (INCA, 2019).

A detecção do câncer de mama pode ocorrer nas fases iniciais e na maioria dos casos, isso possibilita um tratamento menos agressivo ao paciente. Os sintomas da doença devem ser investigados para que ocorra a confirmação do diagnóstico. A identificação do câncer de mama geralmente consiste em exames por palpação, mamografia e ultra-sonografia. Após o exame, é realizada a biópsia (quando necessário), do tecido mamário para determinar se o diagnóstico indica um crescimento maligno de tecido. Por meio das biópsias tornam-se viável avaliar a histologia microscópica e os elementos do tecido. Assim, pode distinguir o tecido normal de lesões benignas e malignas.

O tecido coletado durante a biópsia geralmente é corado com hematoxilina e eosina (H&E) antes da análise visual realizada pelos especialistas. Segundo Elmore et al. (2015), o processo de diagnóstico usando biópsias coradas com H&E não é uma atividade simples e a concordância diagnóstica média entre especialistas é de aproximadamente 75%.

Nos dias atuais, é crescente o interesse no desenvolvimento e aplicação de técnicas de processamento de imagens, onde os principais objetivos são aumentar a precisão e rapidez do diagnóstico e fornecer uma segunda opinião aos médicos. Dentro das técnicas de Processamento Digital de Imagens (PDI), as soluções de aprendizado profundo tem se mostrado muito eficazes para resolução de problemas semelhantes.

As técnicas de PDI tem mostrado robustez para análise e interpretação de imagens médicas (LEDERLIN et al., 2013). Este auxílio ao especialista vem se tornando crucial para que se obtenha um diagnóstico preciso o mais rápido possível. O desenvolvimento dos métodos automáticos tem sido cada vez mais explorados, pois possibilitam detectar áreas suspeitas de anomalias.

O aprendizado profundo se mostra robusto e eficiente em descobrir estruturas complexas em dados de alta complexidade e, portanto, pode ser aplicado em vários contextos distintos. Os métodos de *deep learning* são capazes de transformar representações brutas em representações mais abstratas (LECUN; BENGIO; HINTON, 2015).

Nos últimos anos, houve um crescente interesse no desenvolvimento e uso de técnicas de processamento de imagens, onde os principais objetivos são aumentar a precisão do diagnóstico e fornecer uma segunda opinião aos médicos. Como alternativa para auxiliar no diagnóstico por imagem, entram sistemas como Detecção Auxiliada por computa-

dor (CAD) e Diagnóstico Auxiliado por Computador (CADx) (OLIVEIRA et al., 2015), responsáveis por realizar detecção e diagnóstico de doenças de forma automática, respectivamente.

No presente trabalho, é proposta uma abordagem computacional para realizar a classificação de imagens de biópsia identificando se as mesmas possuem ou não tecidos cancerígenos. O intuito deste método é fornecer uma segunda opinião aos especialistas buscando tornar o diagnóstico mais preciso, precoce e de baixo custo.

Para viabilizar o desenvolvimento da metodologia proposta, aplicaremos a técnica de *Transfer Learning*, que consiste em utilizar redes neurais convolucionais pré-treinadas. Desta maneira, é possível utilizar o conhecimento já adquirido por elas para a extração de características. Além disso, será utilizada a análise de textura, que será feita por meio de índices de diversidade ecológicos. Por fim, utilizamos os classificadores *Random Forest* e *Support Vector Machine* (SVM) para realizar a classificação.

Como contribuição apresentamos um método totalmente automático para classificação de imagens histopatológicas em cancerígenas e saudáveis. No âmbito da computação apresentamos novas técnicas que são capazes de mensurar propriedades de textura das imagens, que quando utilizadas em conjunto com a técnica de *Transfer Learning* possam fornecer melhores resultados para o especialista ou atuar na triagem de pacientes.

1.1 Objetivos

1.1.1 Objetivo Geral

A partir do contexto apresentado, o presente trabalho possui como objetivo principal desenvolver um método automático para a classificação de imagens histopatológicas da mama em doentes e saudáveis.

1.1.2 Objetivos Específicos

1. Avaliar o desempenho da extração de características utilizando *Transfer Learning*;
2. Avaliar a viabilidade do uso de índices de diversidade como forma de extração de características;
3. Realizar a concatenação de características para alcançar melhores resultados;
4. Testar o desempenho dos classificadores SVM e *Random Forest* nas características extraídas; e
5. Construir uma metodologia que forneça uma segunda opinião ao especialista no diagnóstico do câncer de mama.

1.2 Organização do Trabalho

Além da Introdução, este trabalho está dividido na seguinte ordem: no Capítulo 2, é apresentado o referencial teórico; no Capítulo 3, apresentam-se os trabalhos relacionados; no Capítulo 4, descreve-se a metodologia do trabalho; no Capítulo 5, são descritos os resultados e discussões da execução do projeto; e por fim, são apresentadas as conclusões e sugestões para trabalhos futuro no Capítulo 6.

2 Referencial Teórico

Com o propósito de facilitar o entendimento do trabalho aqui proposto, serão abordados neste Capítulo conceitos, conhecimentos e informações pertinentes ao trabalho.

2.1 Câncer de Mama

O Câncer de mama é a patologia maligna de maior mortalidade na população feminina (BUITRAGO; UEMURA; SENA, 2011). É uma doença que pode ter seu quadro agravado quando o diagnóstico ocorre numa fase tardia, o que diminui as possibilidades de cura do paciente.

De acordo com a Organização Mundial da Saúde (OMS), há duas estratégias para a detecção precoce do câncer: diagnóstico precoce, ou abordagem ágil e oportuna de pessoas com sinais e sintomas de câncer; e rastreamento, aplicação regular de um teste em pessoas aparentemente saudáveis, pertencentes a uma faixa etária de maior risco para a doença, com o objetivo de identificá-las em fase pré-clínica e reduzir a mortalidade por essa causa (TOMAZELLI et al., 2016).

Alguns dos sinais e sintomas deste câncer podem ser notados em fases iniciais, por exemplo: nódulos fixos e indolores na mama, axila e pescoço; pele da mama avermelhada; alterações no mamilo; líquido anormal sendo expelido pelos mamilos (INCA, 2019).

Muitos avanços vêm ocorrendo no tratamento do câncer de mama nas últimas décadas. Nos dias atuais, o conhecimento sobre as várias maneiras da doença se apresentar e as diversas terapêuticas estão disponíveis.

Segundo o (INCA, 2019), o tratamento do câncer de mama depende da fase em que a doença se encontra (estadiamento) e do tipo do tumor. Pode incluir cirurgia, radioterapia, quimioterapia, hormonioterapia e terapia biológica.

2.2 Processamento Digital de Imagens

O PDI é uma forma de processamento de dados, em que sua entrada e saída são imagens. O objetivo de seu uso é melhorar o aspecto visual de certas feições estruturais para o analista humano e fornecer outros subsídios para a sua interpretação, inclusive gerando produtos que possam ser posteriormente submetidos a outros processamentos (CÂMARA et al., 1996).

Graças aos avanços tecnológicos obtidos nos últimos anos, nota-se que vem crescendo o leque de aplicações na área de PDI (FILHO et al., 2016). Esses avanços fazem com que o número de novas pesquisas cresça e as já existentes sejam aprimoradas e expandidas.

O uso das técnicas de PDI se expandiram, e são usadas para vários tipos de tarefas em diversas áreas, como: computação gráfica, processamento de dados, visão computacional, inteligência artificial, reconhecimento de padrões. Essas técnicas vêm sendo aproveitadas em diferentes áreas, como na medicina, na biologia, na astronomia, na área industrial, no auxílio a lei, entre outras aplicações.

No geral, o processo do PDI é subdividido em algumas etapas. Essas etapas podem vir a mudar de um trabalho para o outro, mas geralmente essas etapas são: aquisição da base de imagens, pré-processamento, segmentação, extração de características e classificação ou reconhecimento.

2.2.1 Aquisição das Imagens

A Aquisição das imagens é a primeira etapa a ser realizada. Para a realização desta etapa são necessários: um instrumento formador de imagem um dispositivo físico sensível a uma faixa do espectro eletromagnético, capaz de produzir um sinal elétrico proporcional ao nível de energia percebida; e um digitalizador, que converterá a saída elétrica do dispositivo de sensoramento físico para a forma digital (GONZALEZ; WOODS; EDDINS, 2002).

2.2.2 Pré-Processamento

No pré-processamento, é feita uma melhora na qualidade da imagem. Esta melhora envolve duas categorias principais: métodos que operam no domínio espacial e métodos que operam no domínio de frequência.

Técnicas de processamento no domínio espacial baseiam-se em filtros que manipulam o plano da imagem, enquanto que as técnicas de processamento no domínio da frequência se baseiam em filtros que agem sobre o espectro da imagem (ESQUEF; ALBUQUERQUE; ALBUQUERQUE, 2003). É comum utilizar ambas as técnicas combinadas, para realçar determinadas características de uma imagem.

2.2.3 Segmentação

A segmentação é a etapa onde é realizada a separação da imagem como um todo em partes que constituem a mesma, mas que se diferenciem entre si. Os *pixels* de interesse são geralmente denominados “objetos”, e é chamado de “fundo” os *pixels* que podem ser desprezados ou que não tem utilidade.

Dentre todas as etapas do processamento de imagens, a segmentação é considerada uma das mais críticas no tratamento da informação, pois é aqui onde são definidas as regiões de interesse para análises e processamentos posteriores. Assim, distorções ou erros nesta etapa refletem nas demais, o que pode vir a causar resultados indesejados causando contribuições negativas na eficiência do processo.

A segmentação é um processo adaptativo, ou seja, ela busca sempre se adequar as características particulares de cada imagem e objetivo que se pretende alcançar. Ainda que exista uma vasta diversidade de técnicas de segmentação de imagens, ainda assim há um grande interesse no estudo, desenvolvimento e adaptação de novas técnicas.

A segmentação possui duas abordagens principais, a similaridade entre os *pixels* e a descontinuidade entre eles. Na similaridade a técnica mais usada é a Binarização (*Image Thresholding*), pois é uma técnica eficiente e simples do ponto de vista computacional. A binarização é usada quando as amplitudes dos níveis de cinza são suficientes para caracterizar os “objetos” presentes na imagem (ESQUEF; ALBUQUERQUE; ALBUQUERQUE, 2003).

Já as técnicas baseadas em descontinuidade, procuram determinar variações bruscas entre *pixels* próximos, o que permite identificar os grupos de *pixels* que formam o limite das bordas dos objetos da imagem. Essa é uma técnica de segmentação muito eficiente para a detecção de bordas

2.2.4 Extração de Características

A etapa de extração de características é uma das etapas finais de um sistema de processamento de imagens. É nesta etapa que são retiradas as informações úteis da imagem que foi segmentada ou pré-processada. Quando o objetivo do processamento é a obtenção de informações em números, é realizada uma extração de atributos.

De forma geral, a extração de características pode ser subdividida em algumas categorias de análise, sendo elas por: textura e forma. O objetivo geral de uma análise por textura, é descrever aspectos da imagem que dizem respeito a suavidade, rugosidade e regularidade. Já na análise baseada na forma da imagem, o intuito é extrair informações que mensuram sobre propriedades estruturais da imagem (FILHO et al., 2016).

Nessa etapa pode-se utilizar aprendizado profundo, como por exemplo uma Rede Neural Convolutiva, porém, extração ocorre de maneira diferente. Essas redes profundas possuem em sua arquitetura muitas camadas que são responsáveis pela extração de características. Uma característica interessante é a capacidade da rede treinar as camadas de extração subsequentes, ao invés de usar valores predeterminados.

2.2.5 Classificação/Reconhecimento

Na classificação, o objetivo é realizar, de forma automática, a identificação dos objetos segmentados na imagem. Isto é realizado por um algoritmo que mede a similaridade dos objetos e classes, define uma função discriminante que separe os grupos e classifica os objetos de acordo com a sua função programada.

Podemos subdividir os classificadores em dois tipos, supervisionados e não supervisionados. Em ambos os casos são necessárias duas fases, que são o treinamento e a

classificação de fato.

A classificação supervisionada se baseia na identificação de diferentes classes com comportamentos espectrais diferenciados. Para isso, algoritmos de classificação são adotados com o intuito de extrair as feições de interesse em um espaço multidimensional (CARVALHO; JÚNIOR, 2007).

Na classificação não-supervisionada o usuário utiliza algoritmos pra reconhecer as classes da imagem, e ao ter as áreas definidas, o usuário não se preocupa com a homogeneidade das classes. Esta classificação ocorre pelo submetimento dos *pixels* de uma área a um algoritmo de agrupamento que determina o agrupamento do dado.

2.3 *Deep Learning*

Deep Learning, trata-se de um conjunto de métodos de *machine learning*, com foco em problemas com o nível de abstração elevado (LECUN et al., 1998), e é vista como uma das sub-áreas da inteligência computacional. As técnicas de aprendizado profundo também são muito utilizadas para realizar classificação, eliminando variações irrelevantes das informações por meio de uma ampliação dos aspectos de entrada.

Dentre suas características, estão a presença de estruturas de grafos profundos separados por níveis de processamento. Esses algoritmos se assemelham à complexa rede de neurônios humanos, onde as ligações são representadas pelos nós dos grafos. Por sua robustez e abstração, os métodos de aprendizado profundo podem ser usados para muitos objetivos e aplicações, como por exemplo: reconhecimento facial, segmentação e classificação de imagens.

2.3.1 *Redes Neurais Convolucionais*

Uma Rede Neural Convolutiva (CNN) é uma variação das redes de perceptrons de múltiplas camadas, tendo sido inspirada no processo biológico de processamento de dados visuais (VARGAS; PAES; VASCONCELOS, 2016). Este tipo de rede vem sendo amplamente utilizada em aplicações de classificação, extração de características, detecção e reconhecimento em imagens e vídeos.

Uma CNN é capaz de aplicar filtros de dados visuais mantendo a relação dos *pixels* vizinhos durante o processamento da rede. É uma arquitetura de aprendizado profundo composta por uma ou várias camadas convolucionais e subsequentes, seguidas opcionalmente de camadas totalmente conectadas, como em uma rede neural multicamada padrão (LECUN et al., 1998).

A utilização de CNN tem se tornado bem popular em tarefas de visão computacional. Elas possuem uma facilidade em serem treinadas quando a quantidade de amostras rotuladas que representam as diferentes classes alvos é grande. O treinamento delas é

facilitado pelo fato de cada unidade de camada não estar conectada a todas as unidades da camada seguinte.

De forma simplificada, pode ser considerada uma CNN toda rede neural que utiliza a convolução no lugar da multiplicação de matrizes em pelo menos uma de suas camadas (GOODFELLOW et al., 2016). Na atualidade, as redes convolucionais existentes trazem decisões de projeto que são importantes, decisões essas que foram adotadas e aprimoradas ao longo de seu desenvolvimento.

As CNNs são arquiteturas compostas por diversas camadas, entre elas temos a camada convolucional. Esta é responsável pela parte mais pesada do processamento computacional. Ela é constituída por filtros que recebem um arranjo de *pixels*, que pode ser chamado de volume. Os conjuntos de filtros dessa camada são capazes de aprender de acordo com seu treinamento.

Esses filtros realizam convoluções com os dados de entradas para obter um mapa de características, que indicam regiões onde características específicas em relação ao filtro, são encontradas na entrada. Assim, os valores reais dos filtros são alterados ao longo do treinamento, tornando possível que a rede aprenda a identificar regiões significantes para extrair características dos dados.

É comum que em seguida a uma camada convolucional, exista uma camada de *pooling*. Essa camada tem como objetivo reduzir gradualmente a dimensão espacial do volume de entrada, para que seja reduzido também o custo computacional da rede e evitar *overfitting* (ARAÚJO et al., 2017a).

Na operação de *pooling*, ocorre uma substituição dos valores pertencentes a uma região do mapa de atributos que foi gerado pela camada convolucional, por alguma das métricas dessa região. É comum a utilização de uma técnica em que os valores são substituídos pelo valor máximo de uma região (GOODFELLOW et al., 2016), isso é chamado de *max pooling*.

2.3.2 *Transfer Learning*

Transfer Learning é a reutilização de um modelo pré-treinado de CNN em um novo problema. Basicamente, é o processo de transferir o conhecimento de uma tarefa de origem para uma tarefa de destino permitindo progresso rápido ou duradouro ao modelar a segunda tarefa (PAN; YANG, 2009).

A ideia geral sugere a utilização de uma rede neural que já foi treinada em um novo conjunto de dados, com o objetivo de solucionar um problema diferente. Em alguns estudos, é difícil obter um conjunto de dados grande o suficiente, o que descarta a possibilidade de treinar a rede do zero. Outro fator é o custo computacional para treinar uma CNN. Quanto mais profunda a arquitetura, mais caro e custoso o treinamento da mesma.

A *Transfer Learning* pode ser bem eficaz quando comparada com o método normal de *machine learning*, se saindo melhor em alguns aspectos como velocidade, precisão e a

necessidade de um conjunto de treino menor. A Figura 1 resume como os dois processos ocorrem.

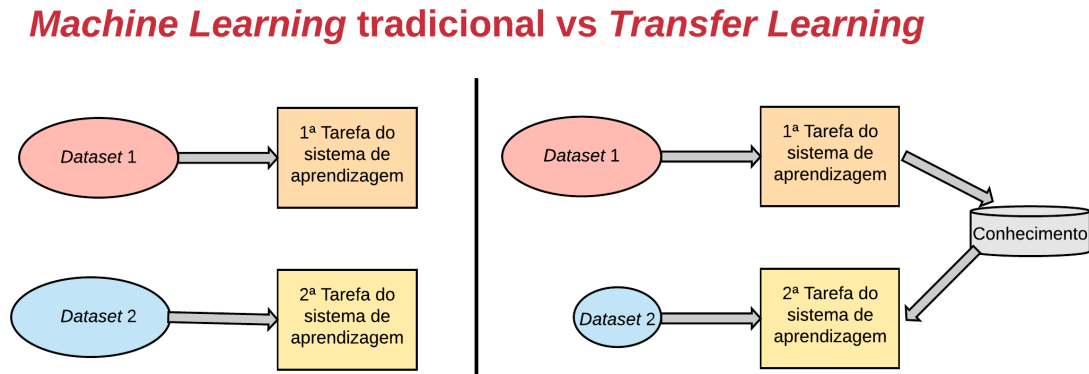


Figura 1: Ao lado esquerdo, o processo de realização de duas tarefas de *Machine Learning*. Na direita, o *Transfer Learning* sendo utilizado para passar o conhecimento da tarefa de origem, para a de destino.

Para este trabalho foram selecionadas 3 arquiteturas de CNN previamente treinadas, com pesos pré-definidos e estas foram testadas e tiveram suas performance avaliadas quanto ao *Transfer Learning*. Os modelos utilizados foram as VGGNets (VGG16 e VGG19) (SIMONYAN; ZISSERMAN, 2014) e a Residual Network (ResNet50) (HE et al., 2016).

2.3.2.1 VGGNets

A VGG é uma rede neural convolucional profunda proposta e desenvolvida por um grupo de Oxford (SIMONYAN; ZISSERMAN, 2014). A mesma garantiu a primeira e segunda posições nas tarefas de localização e classificação, respectivamente, no desafio *ImageNet* 2014. Esta rede geralmente possui as profundidades de 16 e 19 camadas.

Esta rede traz consigo alguns padrões: seus filtros são de tamanho 3x3, *pools* máximos são alocados após cada duas convoluções e o número de filtros deveria ser dobrado após cada *max-pooling*.

A VGG é uma rede amplamente utilizada pela comunidade, inclusive para *Transfer Learning*. Ela possui arquitetura homogênea e robusta para extração de características.

Neste trabalho, realizamos a extração de características por meio da VGG com 16 e com 19 camadas. A VGG19 é um modelo originado pela VGG16, com a principal diferença em seu número de camadas (A VGG19 possui apenas 3 camadas de convolução a mais).

Inicialmente, o conceito da VGG19 era diferente da VGG16, sendo que a VGG19 foi utilizada com o objetivo de detecção de objetos (SIMONYAN; ZISSERMAN, 2014) quando foi desenvolvida.

2.3.2.2 ResNet50

A ResNet50 é uma arquitetura que tem como principal característica suas conexões residuais. Esta arquitetura foi a vencedora do desafio *ImageNet* (DENG et al., 2009) em 2015, onde realizam a classificação de imagens em 1000 classes diferentes. Esta arquitetura, trabalha de maneira a aprender resíduos, em vez de aprender características (HE et al., 2016).

A parte residual pode ser entendida como a subtração de características aprendidas da entrada dessa camada e a ResNet faz isso usando conexões de atalho, que são conexões em que o gradiente é diretamente retro-propagado para camadas anteriores.

A ResNet50 é capaz de solucionar o problema do desaparecimento de gradiente, que é ocasionado por meio da extrema profundidade da rede. Devido a essa profundidade da rede, os gradientes calculados onde é a função de perda acabam diminuindo para zero após as diversas aplicações da regra da cadeia. Dessa maneira, os pesos nunca tem seus valores atualizados e a rede torna-se incapaz de aprender. Já na arquitetura residual, os gradientes podem proceder através das conexões de salto de maneira direta.

2.4 Índices de Diversidade

A diversidade é um termo muito utilizado na área da ecologia. O seu objetivo é informar variedade de espécies presentes em uma comunidade ou área. Para realizar a medição da diversidade são utilizados dois componentes: A riqueza da espécie que consiste no número das espécies em uma determinada região e a abundância relativa, que é o número de indivíduos de uma determinada espécie numa determinada área.

O resultado obtido por qualquer índice de diversidade é representado por um único valor. A forma mais básica de utilizar índices de diversidade em imagens, é quando a comunidade representa uma imagem ou região da mesma, as espécies sendo os níveis de cinza e os indivíduos sendo os *pixels*.

Estes índices podem agir como descritores de textura das imagens possibilitando a extração de características das mesmas. Desta maneira, é possível reconhecer a heterogeneidade e homogeneidade das imagens, o que auxilia a classificar as mesmas como normais ou doentes.

A Tabela 1 mostra como são interpretados cada um dos elementos destes índices quando aplicados ao processamento de imagens.

Tabela 1: Interpretação dos elementos da ecologia dos índices de diversidade no Processamento de Imagens

Metodologia Proposta	Ecologia
Região de Interesse da Imagem	Comunidade
Níveis de Cinza da Imagem	Espécies
Pixeis da Imagem	Indivíduos

Foi realizada a extração das características baseada em textura utilizando 10 índices: Índice de Shannon, Índice de Simpson, Índice de Margalef, Índice de McIntosh, Índice de Camargo, Índice de Smith & Wilson B, Índice de Smith & Wilson 1-D, Índice de Smith & Wilson $-\ln D$, Índice de Shannon máximo e Índice de Shannon mínimo.

2.4.0.1 Índice de Shannon

O Índice de Shannon assume que os indivíduos são randomicamente amostrados a partir de uma comunidade infinitamente grande, e que todas as espécies estão representadas na amostra (SHANNON; WEAVER, 1949). O cálculo deste índice ocorre conforme a Equação 2.1

$$H' = - \sum_{i=1}^S p_i \ln p_i, \quad (2.1)$$

onde S é total de espécies, p_i é a proporção de indivíduos pertencentes a espécie i na amostra, calculado com a Equação 2.2

$$p_i = \frac{n_i}{N}, \quad (2.2)$$

onde n_i é o número de indivíduos na espécie i na amostra e N é o número total de indivíduos na comunidade.

2.4.0.2 Índice de Simpson

O Índice de Simpson (D) é a medida de probabilidade de dois indivíduos selecionados aleatoriamente de uma comunidade infinitamente grande, pertencerem à mesma espécie (SIMPSON, 1949). O Índice é calculado da pela Equação 2.3

$$D^1 = 1 - \sum_{i=1}^S p_i^2, \quad (2.3)$$

onde p_i é o número de indivíduos na espécie i na comunidade, n_i é o número de indivíduos na espécie i e N é o número total de indivíduos na comunidade. S é o total de espécies predominante na comunidade.

2.4.0.3 Índice de Margalef

Índice de Margalef (D) é usado na ecologia para estimar a biodiversidade de uma comunidade com base na distribuição numérica dos indivíduos das diferentes espécies em função do número total de indivíduos existentes na amostra analisada (MARGALEF, 1969). O Índice é calculado como mostra a Equação 2.4

$$D^2 = \frac{(S - 1)}{\ln(N)}, \quad (2.4)$$

onde S é o numero de espécies e $\ln(N)$ é o logaritmo do número de indivíduos pertencente a comunidade.

2.4.0.4 Índice de McIntosh

No índice proposto por McIntosh (1967), a comunidade pode ser vista como um ponto em um hiper-volume S-dimensional e a distância euclidiana da comunidade para a origem pode ser utilizada como uma medida de diversidade. E uma medida de probabilidade de dois indivíduos selecionados aleatoriamente de uma comunidade infinitamente grande pertencerem a mesma espécie (ARAÚJO et al., 2017b). O Índice é calculado a partir da Equação 2.5,

$$D^3 = \frac{(N - U)}{N - \sqrt{N}}, \quad (2.5)$$

onde N é o número total de indivíduos na amostra e U é dado pela Equação 2.6

$$U = \sqrt{\sum n_i^2}, \quad (2.6)$$

onde n_i é o número de indivíduos na espécie i e o somatório é realizado sobre todas as espécies. U é a distancia euclidiana da comunidade desde a origem quando plotada em um hiper volume S-dimensional.

2.4.0.5 Índice de Camargo

O índice de Camargo (1993) é a relação entre pares de espécies i e j . O Índice é definido pela Equação 2.7

$$E^1 = 1 - \left[\sum_{i=1}^S \sum_{j=i+1}^S \frac{P_i - P_j}{S} \right], \quad (2.7)$$

onde p_i é a proporção de espécies i na amostra, p_j é a proporção de espécies j na amostra e S é o número total de espécies.

2.4.0.6 Índice de Smith & Wilson B

O índice de B de [Smith e Wilson \(1996\)](#) é baseado na variação de abundância. Essa variância é calculada usando abundâncias de logaritmo, o que significa que o índice examina as diferenças proporcionais entre as espécies. Com a obtenção da variância, a mesma é multiplicada pelo fator $-2 / \pi \arctan$ para dar um valor entre o intervalo de 0 a 1. A Equação 2.8 define o índice

$$E^2 = 1 - \left[\frac{2}{\pi \arctan \left\{ \frac{\sum_{i=1}^S (\ln n_i - \sum_{j=1}^S \frac{\ln n_j}{S})^2}{S} \right\}} \right], \quad (2.8)$$

sendo n_i o número de indivíduos nas espécies i , n_j é o número de indivíduos nas espécies j e S é o número total de espécies.

2.4.0.7 Índice de Smith & Wilson 1-D

O índice de uniformidade de ([SMITH; WILSON, 1996](#)) 1-D utiliza o cálculo da dominância de ([SIMPSON, 1949](#)) (D) de várias maneiras. A mais comum é usar o complemento de D como índice de diversidade e para dados contínuos ou com números grandes de cortes discretos. O cálculo é definido conforme a Equação 2.9

$$E^3 = \frac{1 - D}{1 - \frac{1}{S}}, \quad (2.9)$$

onde D é o índice de diversidade de Simpson e S é o número total de espécies.

2.4.0.8 Índice de Smith & Wilson $-\ln D$

O índice de uniformidade de [Smith e Wilson \(1996\)](#) 1/D converte a dominância de [Simpson \(1949\)](#) (D) em uma medida de diversidade. O índice é definido pela Equação 2.10

$$E^4 = \frac{-\ln D}{\ln S}, \quad (2.10)$$

onde D é o índice de diversidade de Simpson e S é o número total de espécies.

2.4.0.9 Índice de Shannon *maximum*

Este índice é simplesmente o valor máximo que o índice de Shannon pode produzir para um conjunto de dados fornecido. O calculo é dado por $\ln(S)$, onde S é o número total de espécies.

2.4.0.10 Índice de Shannon *minimum*

Este índice é simplesmente o valor mínimo que o índice de Shannon pode produzir para um conjunto de dados fornecido. O mesmo é dado pela Equação 2.11

$$S_{min} = -\ln(N) - \left[\frac{(N - S + 1) \ln(N - S + 1)}{N} \right], \quad (2.11)$$

onde N é o número total de indivíduos na amostra e S é o número total de espécies.

3 Trabalhos Relacionados

Atualmente, vem se difundindo cada vez mais o uso de técnicas de *machine learning* para atividades em áreas diversas, e a área médica é uma delas. Nessa seção serão abordados brevemente alguns trabalhos correlatos diretamente ao método aqui proposto.

Em [Araújo et al. \(2017c\)](#) é proposto um método para a classificação de imagens de biópsia mamárias, utilizando CNN. As imagens são classificadas em quatro classes, tecido normal, lesão benigna, carcinoma in situ e carcinoma invasivo, e em duas classes, carcinoma e não carcinoma. As características extraídas pela CNN também são usados no classificador SVM. A acurácia alcançada foi de 77,8% para quatro classes e 83,3% para carcinoma / não carcinoma.

[Nawaz et al. \(2018\)](#) apresenta o ajuste fino de uma CNN profunda (ALEXNET) alterando e inserindo camadas convolucionais da camada de entrada e a camada totalmente conectada. Este método apresenta resultados experimentais que alcançam a acurácia de 81,25%.

[Golatkhar, Anand e Sethi \(2018\)](#) propõe um método baseado em aprendizado profundo, onde é realizado o ajuste fino de uma CNN Inception-v3 para classificação de imagens de tecido mamário coradas com H&E liberadas para o desafio BACH ([ARESTA et al., 2019](#)). Foi obtida uma acurácia média de 85% nas quatro classes e 93% para não-câncer (isto é, normal ou benigno) vs. tumor maligno (carcinoma invasivo ou in situ).

[Rakhlin et al. \(2018\)](#) desenvolve uma abordagem computacional baseada em redes neurais convolucionais profundas para classificação de imagens histológicas de câncer de mama. Neste trabalho, a classificação de 4 classes alcançou acurácia de 87,2%. Para classificação de 2 classes para detectar carcinomas, relatamos 93,8% de acurácia.

[Ferreira et al. \(2018\)](#) propõe uma Inception ResNet v2 que por meio de *Transfer Learning*, ajuste fino e aumento de dados, realiza a classificação de imagens histológicas de câncer de mama. A rede é capaz de aprender quais características são relevantes para a classificação. Este trabalho utiliza a base BACH ICIAR 2018 e atingiu acurácia de 76% para a classificação de 4 classes no conjunto de teste.

Em [Vesal et al. \(2018\)](#) é proposta uma abordagem baseada em *Transfer Learning* para a classificação de imagens histopatológicas da mama em 4 classes: normal, benigno, carcinoma in situ e carcinoma invasivo. As imagens são da base BACH ICIAR 2018. Primeiramente as imagens são normalizadas para corrigir variações de cores induzidas durante a preparação das lâminas. Após isto, são extraídos patches das imagens e estes são usados para ajustar as CNNs Inception-V3 e ResNet50, ambas pré-treinadas. Foi utilizado *cross-validation* com $k=3$ para avaliação. A Inception-V3 alcançou uma acurácia de teste média de 97,08% para quatro classes, superando marginalmente a rede ResNet50, que alcançou uma média de 96,66%.

Dentro da literatura, há poucos trabalhos semelhantes onde são utilizados os índices de diversidade ecológicos para caracterizar textura, voltado ao problema do câncer de mama.

O trabalho de [Carvalho, Filho e Dalilia \(2017\)](#), onde é apresentado um método de classificação de tecidos da mama. São usados descritores de textura baseados em índices filogenéticos para extração de características, em seguida feita uma classificação usando Bagging, BayesNet e *Random Forest*. A base utilizada é a DDSM. Os resultados alcançaram uma acurácia de 90%, sensibilidade de 88,5% e especificidade de 91,7%.

Como na literatura há trabalhos que utilizam este método de análise de textura para a tarefa de texturização de imagens suspeitas de câncer de mama, foi selecionado o trabalho de [Reis et al. \(2012\)](#). Em seu trabalho é investigado o potencial grau de caracterização de textura dos índices de diversidade aplicado em imagens digitais de faces humanas. São utilizados os índices como extratores de características e o SVM como classificador. Dentre os 18 testes realizados foram obtidas acurácias superiores a 95%.

Abaixo, a Tabela 2 resume os trabalhos relacionados, as ferramentas e tecnologias utilizadas e a base de imagens.

Tabela 2: Resumo dos trabalhos relacionados.

Trabalho	Método	Base de Imagens
(REIS et al., 2012)	Índices de Diversidade	CBCL, ORL, Libor Spacek e CMU AMP Face Expression
(ARAÚJO et al., 2017c)	CNN	<i>Bioimaging Challenge 2015 Breast Histology</i>
(CARVALHO; FILHO; DALILIA, 2017)	Índices Filogenéticos	DDSM
(GOLATKAR; ANAND; SETHI, 2018)	Ajuste Fino da Inception-v3	BACH ICIAR 2018
(RAKHLIN et al., 2018)	<i>Deep CNN</i>	BACH ICIAR 2018
(NAWAZ et al., 2018)	CNN AlexNet	BACH ICIAR 2018
(FERREIRA et al., 2018)	CNN Inception CNN ResNet v2	BACH ICIAR 2018
(VESAL et al., 2018)	ResNet50 e Inception-v3	BACH ICIAR 2018
Método Proposto	<i>Transfer Learning</i> e Índices de Diversidade + <i>Random Forest</i> e SVM	BACH ICIAR 2018

De acordo com a literatura pesquisada, as técnicas de PDI e *machine learning* vem tornando-se imprescindíveis no auxílio aos especialistas em diagnósticos envolvendo imagens médicas. Uma desvantagem do aprendizado profundo é seu alto custo computacional, pois estes algoritmos necessitam um grande período de treino e grande quantidade de imagens para atingir bons resultados. Este empecilho pode ser resolvido utilizando *Transfer Learning*, que trata-se da utilização de redes neurais convolucionais pré treinadas. Esta abordagem é efetiva e muito útil principalmente quando o conjunto de dados é pequeno.

Na abordagem aqui proposta, serão utilizados alguns modelos de CNN pré-treinados (ResNet50, VGG16 e VGG19) juntamente com os índices de diversidade para a extração de características. É feita a comparação da eficiência dos dois métodos testando os atributos

extraídos nos classificadores *Random Forest* e SVM. Há também a concatenação entre as características dos índices de diversidade e do modelo de *Transfer Learning* que obteve o melhor resultado, e são refeitos os testes com os dois classificadores.

4 Metodologia

A metodologia que será seguida para a realização deste trabalho consiste em: aquisição do *dataset* de imagens; extração das características aplicando a técnica de *Transfer Learning* e análise de textura utilizando índices de diversidade ecológicos; classificação com *Random Forest* e *Support Vector Machine*; e validação dos resultados. A Figura 2 ilustra a metodologia.

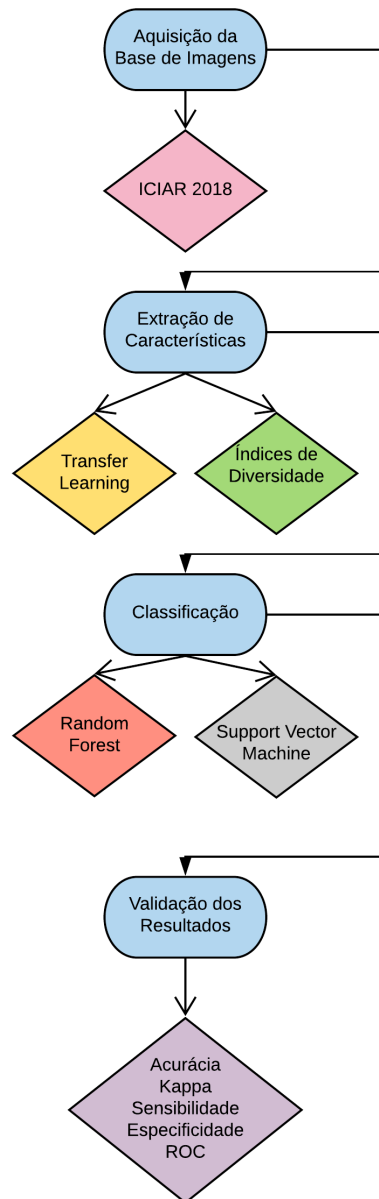


Figura 2: Representação resumida da metodologia proposta.

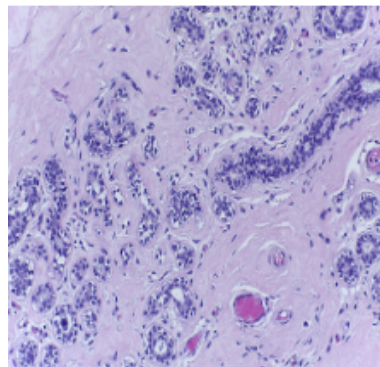
4.1 Aquisição de Imagens

Para este trabalho, a base de imagens utilizada foi a base de imagens BACH ICIAR 2018 (ARESTA et al., 2019). A mesma é utilizada em um desafio que contém como um de seus objetivos a classificação do tipo de câncer de mama. A base de dados utilizada no trabalho consiste em 400 imagens divididas em 4 classes e cada uma das classes possui 100 imagens.

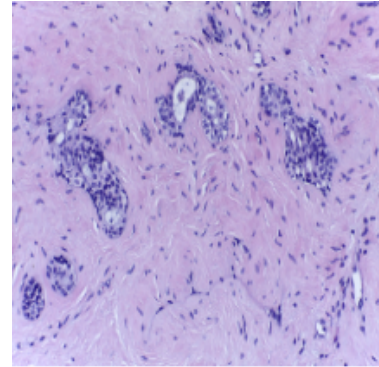
Cada imagem é rotulada com uma das quatro classes balanceadas: carcinoma normal (*normal*), benigno (*benign*), in situ (*InSitu*) e carcinoma invasivo (*invasive*), onde a classe é definida como um tipo de câncer predominante na imagem. A anotação da imagem foi realizada por dois médicos especialistas.

Na abordagem aqui proposta, a base foi agrupada em duas classes de imagens. A classe normal permanece sem alterações, enquanto as outras 3 classes (benigna, in situ e invasivo) se tornam uma classe que denominamos doente. Desta maneira, teremos 100 imagens que não possuem carcinomas, enquanto outras 300 possuem.

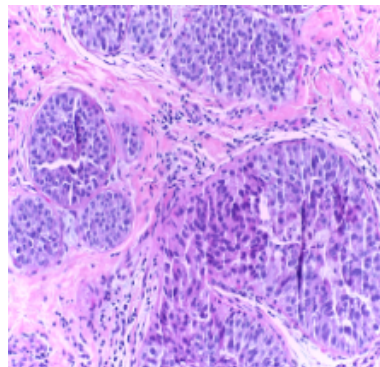
A seguir, são mostrados exemplos de imagens contidas no *dataset* utilizado durante este trabalho.



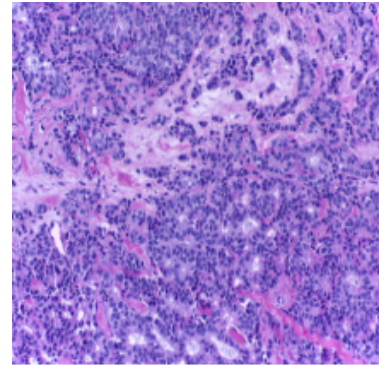
(a) Biópsia classificada como normal.



(b) Biópsia classificada como benigno.



(c) Biópsia classificada como *In-Situ*.



(d) Biópsia classificada como invasivo.

Figura 3: Exemplos da Base de imagens BACH ICIAR 2018.

4.2 Extração de Características

A etapa de extração de características foi concebida utilizando duas técnicas diferentes: *Transfer Learning* e análise de textura. Foram utilizadas 3 arquiteturas de CNN unidas ao *Transfer Learning* e índices de diversidade ecológica atuando como descritores de textura para realizar a etapa em questão.

4.2.1 *Transfer Learning*

Para mensurar a efetividade das diferentes formas de implementar uma CNN, foram utilizadas diferentes arquiteturas para a extração dos atributos das imagens. Visando melhores resultados e diminuir o tempo e custo computacional das CNNs, utilizamos *Transfer Learning*.

Para este trabalho, foram selecionados os três dos principais modelos para serem estudados e terem sua performance avaliados para o problema em questão. Os modelos utilizados neste estudo foram a ResNet50, a VGG16 e a VGG19. Estas arquiteturas foram escolhidas pois apresentam bons resultados ao solucionar problemas semelhantes, como foi visto em alguns trabalhos relacionados.

Todos estes modelos foram pré-treinados na mesma base de imagem, a ImageNet (DENG et al., 2009). Os modelos foram treinados para classificar imagens entre 1000 classes diferentes. Para utilizar estes modelos como extrator de características, foi retirada a última camada, e foram salvos os valores que estavam contidos nesta camada.

4.2.2 Índices de Diversidade Ecológicos

O objetivo geral de uma análise por textura, é descrever aspectos da imagem que dizem respeito a suavidade, rugosidade e regularidade. Desta maneira, é possível reconhecer a heterogeneidade e homogeneidade das imagens, o que auxilia a classificar as mesmas como normais ou doentes. A Tabela 1 mostra como são interpretados cada um dos elementos destes índices quando aplicados ao processamento de imagens.

Os índices de diversidade de espécies no contexto da ecologia indicam variedade de espécies, possibilitando abordar informações de como estão distribuídas as espécies em uma comunidade. Com isso podemos abordar riqueza e igualdade de espécies. A riqueza refere-se ao numero de espécies, quanto maior a riqueza, mais espécies há em uma comunidade. A igualdade refere-se a abundância relativa, ou seja, quanto mais próximas as abundâncias das espécies dentro de uma comunidade, maior é a igualdade.

Neste trabalho, os índices de diversidade são utilizados para medir a diversidade na distribuição dos níveis de cinza presente nas regiões de interesse de uma imagem. A finalidade é utilizar o valor do índice para caracterizar textura nas imagens fornecendo os dados necessários para o reconhecimento de um padrão que possibilite ao classificador

distinguir as regiões que correspondem o tecido tumoral e àquelas que correspondem a tecidos normais, não-massas.

4.3 Classificação

A etapa da classificação tem como objetivo mensurar a similaridade dos objetos e classes contidos na imagem, definir uma função discriminante que separe estes grupos e classificar os objetos de acordo com sua função programada. Em nossa abordagem, foram utilizados os algoritmos *Random Forest* e *Support Vector Machine*, pois os mesmos apresentam bons resultados em diversas tarefas semelhantes que envolvem classificação.

Foi utilizado o método de *cross validation k-folds* com o $k=10$. Este processo realiza a divisão do conjunto de dados em 10 grupos, separando 9 grupos de treino, e um para teste, realizando 10 cruzamentos alterando o grupo de teste. Desta maneira todos os grupos são testados, e é gerada uma média que atribui-se como resultado.

4.3.1 *Random Forest*

O *Random Forest* é um algoritmo de classificação que combina múltiplas árvores de previsão, onde cada árvore depende dos valores de um vetor independente amostrado aleatoriamente e com a distribuição de amostras para todas as árvores da floresta (BREIMAN, 2001).

Neste trabalho, utilizamos os valores de parâmetros setados por padrão do *Random Forest*. Os mesmo vêm setados das seguintes maneiras:

- I: Número de interações = 100;
- K: Número de atributos que serão utilizados para investigação = 0;
- M: Número mínimo de instancias por folha = 1;
- V: Proporção mínima da variação da classe numérica = 0.001;
- S: *Seed* para o gerador de números aleatórios = 1.

4.3.2 *Support Vector Machine*

O *Support Vector Machine* (SVM) é um algoritmo de classificação supervisionada de *machine learning* proposto por (DRUCKER et al., 1997). Neste trabalho, realizamos testes com os diversos parâmetros do SVM para obter melhores resultados.

Os parâmetros utilizados no SVM neste trabalho são apresentados à seguir:

- S: Tipo de SVM utilizado = C-SVC;

- K: Tipo de função do *kernel* = linear, polinomial, RBF (Função de Base Radial) e sigmoide;
- C: Custo da função = 1 e 10;
- D: Degrau na função de *kernel* (aplica-se somente no *kernel* polinomial) = 6;
- G: Número do *gamma* na função do kernel = 0;
- R: Número do tetha na função do kernel = 0;
- E: Tolerância do critério de finalização = 0,001;
- P: Valor do epsilon na função de *loss* = 0.1.

4.4 Métricas de Validação

Para que os resultados obtidos com o presente método possam ser validados, foram utilizadas as métricas: acurácia, especificidade, sensibilidade, coeficiente Kappa e Área Sobre a Curva (ROC).

Para que algumas dessas métricas sejam calculados é necessária uma matriz de confusão. A mesma esta ilustrada segundo a Tabela 3.

Tabela 3: Matriz de confusão

Resultado	Doença	
	Normal	Doente
Positivo	Verdadeiro	Falso
	Positivo (VP)	Positivo (FP)
Negativo	Falso	Verdadeiro
	Negativo (FN)	Negativo (VN)

4.4.1 Acurácia

Acurácia (A) é a proporção de predições corretas, sem levar em consideração o que é positivo e o que é negativo. Esta medida é altamente suscetível a desbalanceamentos do conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema. Representado pela Equação 4.1

$$A = \frac{VP + VN}{VP + FP + VN + FN}. \quad (4.1)$$

4.4.2 Sensibilidade

Sensibilidade (S) é a proporção de verdadeiros positivos, ou seja, a capacidade do sistema em prever corretamente a condição para casos que realmente a têm. Representado pela Equação 4.2

$$S = \frac{VP}{VP + FN}. \quad (4.2)$$

4.4.3 Especificidade

Especificidade (E) é a proporção de verdadeiros negativos, isto é, a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a têm. Representado pela Equação 4.3

$$E = \frac{VN}{VN + FP}. \quad (4.3)$$

4.4.4 Kappa

O coeficiente de concordância Kappa (k) é uma estatística que mede a concordância entre dois ou mais julgadores, quando realizam uma avaliação nominal ou original de um mesmo conjunto ou amostra (COHEN, 1960). O Kappa pode ser interpretado como é exibido na Tabela 4.

Tabela 4: Interpretação de concordância para os valores do índice Kappa.

Kappa (k)	Classificação
$k < 0.2$	Fraca
$0.2 \leq k < 0.4$	Razoável
$0.4 \leq k < 0.6$	Moderada
$0.6 \leq k < 0.8$	Forte
$k \geq 0.8$	Perfeita

4.4.5 Área Sobre a Curva

A Área Sobre a Curva possibilita quantificar a exatidão de um teste diagnóstico (proporcional à área sob a curva), além de possibilitar a comparação de testes de diagnósticos. A Área sobre a Curva ROC constitui um dos índices mais usados para sumarizar a qualidade da curva.

As curvas ROC estão sendo usadas para avaliar a capacidade de discriminação de vários métodos estatísticos que combinam várias pistas, resultados de testes e entre outros propósitos preditivos (PARADELA; LOURENÇO; VERAS, 2005).

5 Resultados e Discussões

Para viabilizar que uma abordagem computacional possa auxiliar um especialista na classificação de biópsias mamárias, é crucial que haja um bom resultado nos testes realizados e que as métricas de avaliação do método sejam bem equilibradas. É preferível também que estes testes não sejam tão demorados.

Para a realização destes experimentos, foi utilizada a plataforma Google Colaboratory (CARNEIRO et al., 2018). A mesma trata-se de um serviço baseado em *cloud* que utiliza Jupyter Notebooks para disseminar a pesquisa em *machine learning*. No Colab, como também é conhecido, é fornecido um ambiente totalmente configurado e pronto para ser utilizado, com acesso gratuito a uma GPU robusta.

Algo a ser levado em conta é o tempo. a extração de características utilizando *Transfer Learning* é mais demorada do que a feita pelos índices de diversidade. A Tabela 5 ilustra o tempo médio aproximado que os extratores de características levam para atuar em cada imagem.

Tabela 5: Tempo aproximado de extração de característica de uma imagem em segundos.

Extrator	Tempo médio
ResNet50	≈ 10 segundos
VGG16	≈ 12 segundos
VGG19	≈ 14 segundos
Índices de Diversidade	≈ 3 segundos

Nos testes realizados com o *Random Forest*, as características extraídas por meio de *Transfer Learning* levaram a resultados que se aproximam um do outro. No entanto, os índices de diversidade tiveram um desempenho um pouco abaixo dos demais. A Tabela 6 mostra os resultados referentes ao classificador *Random Forest*.

Tabela 6: Resultados com *Random Forest*

Extrator	A (%)	k	S (%)	E (%)	ROC (%)
ResNet50	82,75	0,41	81,64	94,28	91,1
VGG16	84,5	0,5	89,58	83,8	93,2
VGG19	82,5	0,41	91,8	81,7	92,1
Índices de Diversidade	71,5	0,09	36,53	76,7	65,4

Já no classificador SVM, realizamos testes variando o parâmetro do *kernel*, alternando mesmo entre: função de base radial (RBF), sigmoide, polinomial e linear. Foram também realizados testes com o parâmetro de custo variando entre 1 ou 10. A Tabela 7 exibe o desempenho do classificador em questão com C=1 e em todos os *kernels*.

Tabela 7: Resultados com *Support Vector Machine* com C=1

Método de Extração	Kernel	A (%)	k	S (%)	E (%)	ROC (%)
ResNet50	RBF	87,5	0,61	86,54	92,1	76,3
	Sigmoide	85,25	0,51	95,55	83,94	71,2
	Polinomial	83,5	0,45	92,5	82,5	68
	Linear	89,75	0,72	79,79	93,02	86,2
VGG16	RBF	88,75	0,61	88,73	88,75	80,2
	Sigmoide	82,25	0,39	93,93	81,19	65,2
	Polinomial	84,25	0,54	71,41	86,91	75,5
	Linear	86,25	0,62	73,1	90,42	81,2
VGG19	RBF	86	0,56	86,66	85,88	74,7
	Sigmoide	84,25	0,48	91,11	83,38	69,8
	Polinomial	83,25	0,47	82,45	83,57	69,8
	Linear	88	0,67	77,65	91,17	83
Índices de Diversidade	RBF	75	0	-	75	50
	Sigmoide	75	0	-	75	50
	Polinomial	74	0,11	44,44	76,92	54,7
	Linear	75	0	-	75	50

Com o parâmetro de custo = 10, a ResNet50 seguiu superior nos resultados, como é mostrado na Tabela 8.

Tabela 8: Resultados com *Support Vector Machine* com C=10

Método de Extração	Kernel	A (%)	k	S (%)	E (%)	ROC (%)
ResNet50	RBF	91,75	0,77	86	93,5	87,8
	Sigmoide	90,5	0,72	89,7	90,7	83,7
	Polinomial	88	0,66	79,5	90,4	82,0
	Linear	89,75	0,72	79,8	93,0	86,2
VGG16	RBF	89	0,7	79,8	91,8	84,3
	Sigmoide	78,25	0,39	57,3	84,2	69,2
	Polinomial	85,5	0,6	72,3	89,5	79,7
	Linear	86,25	0,62	73,2	90,4	81,2
VGG19	RBF	87	0,64	76,7	90	81
	Sigmoide	84,25	0,55	72,8	87,1	75,8
	Polinomial	85,5	0,58	76,9	87,6	77,0
	Linear	88	0,67	77,7	91,2	83,0

Diante dos resultados expostos nas Tabelas 6, 7 e 7 pode-se observar que utilizando a *Transfer Learning* para a extração de características obtivemos resultados superiores aos índices de diversidade. As redes pré-treinadas atingiram bons números, em ambos os classificadores, já os índices atingiram apenas uma média razoável durante os experimentos.

Dentre as redes pré-treinadas, a ResNet50 obteve resultados ligeiramente superiores aos outros extratores. Uma razão para isto ocorrer é sua concepção robusta para a extra-

ção de características devido a sua homogeneidade. Também observamos que diante dos resultados apresentados, o classificador SVM obteve resultados melhores quando comparado ao *Random Forest* para o problema abordado.

Um evento interessante ocorreu nos testes do SVM. Como exibe a Tabela 7 alguns valores de kappa e Sensibilidade foram zerados. Isto ocorreu devido o classificador reconhecer todas as instancias como pertencentes a classe doente quando os *kernels* RBF, sigmoide e linear eram utilizados. Quando testado com o *kernel* polinomial o SVM foi capaz de classificar as imagens nas duas classes, apesar dos resultados não serem tão satisfatórios.

Foi realizado um terceiro experimento, onde houve a concatenação das características extraídas com a ResNet50 e os índices de diversidade. Os resultados do teste citado acima aparecem nas Tabela 9. Foram testados os 2 classificadores novamente, com o *Random Forest* em sua parametrização padrão e variando os parâmetros de *kernel* e custo (C) do SVM entre os valores mencionados anteriormente nesta seção.

Tabela 9: Resultados dos classificadores com as concatenação dos atributos da ResNet50 e dos índices de diversidade.

Classificador	Kernel	A (%)	k	S (%)	E (%)	ROC (%)
SVM (C=1)	RBF	89	0,67	88,8	89	80,7
	Sigmoide	82,25	0,39	91,3	81,1	65,2
	Polinomial	84,5	0,56	73,1	87,4	81,2
	Linear	86,25	0,62	73,2	83	92,6
SVM (C=10)	RBF	91,5	0,76	85,1	93,5	87,7
	Sigmoid	89,5	0,69	88,2	89,8	82,0
	Polynomial	90,25	0,73	81,4	93,1	86,5
	Linear	89,75	0,72	79,8	93	86,2
<i>Random Forest</i>	-	81,75	0,37	90,9	80,9	91,9

Após a realização deste último teste, foi notada uma leve melhora em alguns resultados. O classificador SVM continua a superar o *Random Forest* na tarefa de classificação aqui abordada. É importante mencionar que houve também a leve baixa em alguns resultados, como é exibido na Tabela 9.

Após estes testes, pudemos constatar que o melhor resultado foi utilizando as características extraídas pela ResNet50 no classificador SVM, com o *kernel* RBF e C=10.

A comparação com outros trabalhos dentro da literatura é uma tarefa um pouco complicada, já que há uma algumas diferenças nas maneiras de abordagem do problema e também nas ferramentas e técnicas utilizadas para solucioná-los. A Tabela 10 exibe alguns números comparativos.

No trabalho de [Carvalho, Filho e Dalilia \(2017\)](#), utilizou-se os índices como descritores de textura, porém, em um *dataset* para classificação de mamografias em benigno e maligno. No mesmo apresentaram bons resultados, mas nas imagens histopatológicas os mesmos não tiveram tal sucesso.

Tabela 10: Comparação do método proposto com alguns trabalhos relacionados.

Trabalho	Método	Base de Imagens	A (%)	k	S (%)	E (%)	ROC (%)
(GOLATKAR; ANAND; SETHI, 2018)	Ajuste Fino da Inception-v3	BACH ICIAR 2018	93	-	-	-	-
(CARVALHO; FILHO; DALILIA, 2017)	Índices Filogenéticos	DDSM	90	-	88,5	91,7	-
(ARAÚJO et al., 2017c)	CNN	<i>Bioimaging Challenge 2015 Breast Histology</i>	83,3	-	-	-	-
(NAWAZ et al., 2018)	AlexNet	BACH ICIAR 2018	81,25	-	-	-	-
Método Proposto	<i>Transfer Learning e Índices de Diversidade + Random Forest e SVM</i>	BACH ICIAR 2018	91,75	0,77	86	93,5	87,8

Já nos trabalhos de Araújo et al. (2017c) e Nawaz et al. (2018) são utilizados métodos baseados em redes neurais convolucionais. É uma abordagem eficaz, no entanto, na maioria dos casos requer um grande poder de processamento, tempo de treinamento da rede e vasta quantidade de dados necessária para que se obtenha bons resultados. O trabalho de Golatkar, Anand e Sethi (2018) realiza o ajuste fino da arquitetura Inception-v3, onde a última camada é re-treinada para realizar a tarefa de classificação.

A abordagem aqui proposta utiliza redes pré-treinadas, por isso o tempo de execução é diminuído. Além disto, os testes foram realizados no Colab, que provê *hardware* robusto para realização de testes.

É importante ressaltar que alguns dos trabalhos relacionados realizam a classificação entre carcinoma / não carcinoma, diferente de nossa abordagem, que pretende identificar e separar biópsias normais de biópsias que possuem tumores. Há também trabalhos que realizam a classificação de 4 classes (normal, benigno, *inSitu* e invasivo).

6 Conclusão

No presente trabalho, testamos e avaliamos a eficiência de algoritmos na tarefa de classificação automática de imagens histopatológicas da mama em normal e doentes. Utilizando índices de diversidade e técnicas de *deep learning* como *Transfer Learning* para extração das características os resultados obtidos foram satisfatórios e foram alcançados com certa rapidez na classificação com o *Random Forest* e SVM.

A abordagem de aprendizado profundo VGG16 mostrou-se mais eficiente dentre as demais na tarefa de extração usadas no trabalho em questão. Já para a classificação, o SVM mostrou-se superior ao *Random Forest*, baseando-se nas métricas de avaliação que utilizamos para este trabalho.

O uso de *Transfer Learning* combinado com o SVM pode ser usado de forma eficiente na análise automática de biópsias mamárias, mesmo com um número pequeno de imagens. Acreditamos que a abordagem proposta poderá ser integrada a uma ferramenta de processamento digital de imagens sendo aplicada a casos reais, auxiliando especialistas fornecendo-lhes uma segunda opinião no processo de diagnóstico da biópsia. Com isso, espera-se garantir ao paciente um prognóstico mais favorável quanto à sua saúde e uma melhor qualidade de vida.

Como trabalhos futuros, destacamos:

- Realizar novos testes com uma quantidade maior de imagens / outros *datasets*;
- Realizar o ajuste fino para obter melhores resultados;
- Adaptar a abordagem para a classificação de 4 classes; e
- Utilizar outros classificadores.

Referências

- ARAÚJO, F. H. et al. Redes neurais convolucionais com tensorflow: Teoria e prática. *SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. III Escola Regional de Informática do Piauí. Livro Anais-Artigos e Minicursos*, Sociedade Brasileira de Computação, v. 1, p. 382–406, 2017. Citado na página [21](#).
- ARAÚJO, J. D. L. et al. Diagnóstico de glaucoma em imagens de fundo de olho utilizando os índices de diversidade de shannon e mcintosh. In: SBC. *17º Workshop de Informática Médica (WIM 2017)*. [S.l.], 2017. v. 17, n. 1/2017. Citado na página [25](#).
- ARAÚJO, T. et al. Classification of breast cancer histology images using convolutional neural networks. *PloS one*, Public Library of Science, v. 12, n. 6, p. e0177544, 2017. Citado 3 vezes nas páginas [28](#), [29](#) e [40](#).
- ARESTA, G. et al. Bach: Grand challenge on breast cancer histology images. *Medical Image Analysis*, Elsevier, 2019. Citado 2 vezes nas páginas [28](#) e [32](#).
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página [34](#).
- BUITRAGO, F.; UEMURA, G.; SENA, M. C. F. Fatores prognósticos em câncer de mama. *Comun. ciênc. saúde*, v. 22, n. Sup. Espec. 1, p. 69–81, 2011. Citado na página [17](#).
- CÂMARA, G. et al. Spring: Integrating remote sensing and gis by object-oriented data modelling. *Computers & graphics*, Elsevier, v. 20, n. 3, p. 395–403, 1996. Citado na página [17](#).
- CAMARGO, J. Must dominance increase with the number of subordinate species in competitive interactions? *J Theor Biol*, v. 161, p. 537–542, 1993. Citado na página [25](#).
- CARNEIRO, T. et al. Performance analysis of google colab as a tool for accelerating deep learning applications. *IEEE Access*, IEEE, v. 6, p. 61677–61685, 2018. Citado na página [37](#).
- CARVALHO, E. D.; FILHO, A. O. de C.; DALILIA, A. Classificação de tecidos da mama em maligno e benigno baseado em mamografias digitais usando descritores de textura. 2017. Citado 3 vezes nas páginas [29](#), [39](#) e [40](#).
- CARVALHO, L. M. T. de; JÚNIOR, F. W. A. Classificação digital do uso do solo comparando os métodos “pixel a pixel” e orientada ao objeto em imagem quickbird henriqueta veloso ferreira bernardi maíra dzedzej. 2007. Citado na página [20](#).
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. Citado na página [36](#).
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255. Citado 2 vezes nas páginas [23](#) e [33](#).

- DRUCKER, H. et al. Support vector regression machines. In: *Advances in neural information processing systems*. [S.l.: s.n.], 1997. p. 155–161. Citado na página 34.
- ELMORE, J. G. et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*, American Medical Association, v. 313, n. 11, p. 1122–1132, 2015. Citado na página 14.
- ESQUEF, I. A.; ALBUQUERQUE, M. P. d.; ALBUQUERQUE, M. P. d. Processamento digital de imagens. *CENTRO BRASILEIRO DE PESQUISAS FÍSICAS–CBPF*, 2003. Citado 2 vezes nas páginas 18 e 19.
- FERREIRA, C. A. et al. Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 763–770. Citado 2 vezes nas páginas 28 e 29.
- FILHO, A. O. d. C. et al. Métodos para sistemas cad e cadx de nódulo pulmonar baseada em tomografia computadorizada usando análise de forma e textura. Universidade Federal do Maranhão, 2016. Citado 2 vezes nas páginas 17 e 19.
- GOLATKAR, A.; ANAND, D.; SETHI, A. Classification of breast cancer histology using deep learning. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 837–844. Citado 3 vezes nas páginas 28, 29 e 40.
- GONZALEZ, R.; WOODS, R.; EDDINS, S. Fundamentals of digital image processing. *prentice Hall, upper saddle river, NJ*, 2002. Citado na página 18.
- GOODFELLOW, I. et al. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. Citado na página 21.
- HE, K. 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. Citado 2 vezes nas páginas 22 e 23.
- INCA, I. N. do C. *Câncer de mama*. 2019. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Acessado em: 23 de jun. 2019. Citado 2 vezes nas páginas 14 e 17.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015. Citado na página 14.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 20.
- LEDERLIN, M. et al. Management strategy of pulmonary nodule in 2013. *Diagnostic and interventional imaging*, Elsevier, v. 94, n. 11, p. 1081–1094, 2013. Citado na página 14.
- MARGALEF, R. Diversity and stability: a practical proposal and a model of interdependence. 1969. Citado na página 25.
- MCINTOSH, R. P. An index of diversity and the relation of certain concepts to diversity. *Ecology*, Wiley Online Library, v. 48, n. 3, p. 392–404, 1967. Citado na página 25.

- NAWAZ, W. et al. Classification of breast cancer histology images using alexnet. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 869–876. Citado 3 vezes nas páginas 28, 29 e 40.
- OLIVEIRA, F. S. S. de et al. Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and svm. *Computers in biology and medicine*, Elsevier, v. 57, p. 42–53, 2015. Citado na página 15.
- PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009. Citado na página 21.
- PARADELA, E. M. P.; LOURENÇO, R. A.; VERAS, R. P. Validação da escala de depressão geriátrica em um ambulatório geral. *Revista de saúde pública*, SciELO Public Health, v. 39, p. 918–923, 2005. Citado na página 36.
- RAKHLIN, A. et al. Deep convolutional neural networks for breast cancer histology image analysis. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 737–744. Citado 2 vezes nas páginas 28 e 29.
- REIS, A. B. S. et al. Estudo de índices de diversidade como descritores de textura para classificação de imagens de faces. Universidade Federal do Maranhão, 2012. Citado na página 29.
- SHANNON, C. E.; WEAVER, W. The mathematical theory of information. University of Illinois press, 1949. Citado na página 24.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 22.
- SIMPSON, E. H. Measurement of diversity. *nature*, Nature Publishing Group, v. 163, n. 4148, p. 688, 1949. Citado 2 vezes nas páginas 24 e 26.
- SMITH, B.; WILSON, J. B. A consumer’s guide to evenness indices. *Oikos*, JSTOR, p. 70–82, 1996. Citado na página 26.
- TOMAZELLI, J. G. et al. Avaliação das ações de detecção precoce do câncer de mama no brasil por meio de indicadores de processo: estudo descritivo com dados do sismama, 2010-2011. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 26, p. 61–70, 2016. Citado na página 17.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: *Proceedings of the XXIX Conference on Graphics, Patterns and Images*. [S.l.: s.n.], 2016. p. 1–4. Citado na página 20.
- VESAL, S. et al. Classification of breast cancer histology images using transfer learning. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2018. p. 812–819. Citado 2 vezes nas páginas 28 e 29.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA
“JOSÉ ALBANO DE MACEDO”**

Identificação do Tipo de Documento

- () Tese
() Dissertação
() Monografia
() Artigo

Eu, Milton Martins da Oliveira Júnior,
autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de
02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar,
gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação
Método automático para classificação de imagens histopatológicas
do mamão utilizando Transfer Learning e análise de textura
de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título
de divulgação da produção científica gerada pela Universidade.

Picos-PI 27 de Janeiro de 2020.

Milton Martins da Oliveira Júnior
Assinatura

Milton Martins da Oliveira Júnior
Assinatura