

William de Oliveira Torres  
Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho

**Análise de Textura de Nódulos Pulmonares  
para Sistemas CAD Utilizando Índice de  
Diversidade Funcional**

Picos - PI  
Dezembro de 2019

William de Oliveira Torres  
Orientador: Prof. Dr. Antonio Oseas de Carvalho Filho

## **Análise de Textura de Nódulos Pulmonares para Sistemas CAD Utilizando Índice de Diversidade Funcional**

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**

**T693a** Torres, William de Oliveira.  
Análise de textura de nódulos pulmonares para Sistemas CAD utilizando Índice de Diversidade Funcional. / William de Oliveira Torres. -- Picos,PI, 2019.  
36 f.  
CD-ROM: 4 ¾ pol.

Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação. – Universidade Federal do Piauí, Picos, 2019.  
“Orientador(A): Prof. Dr. Antonio Oseas de Carvalho Filho.”

1. Imagens Médicas. 2. Análise de Texturas – Computação Real. 3. Câncer de Pulmão – Saúde - Tecnologia. I. Título.

**CDD 005.73**

*Elaborada por Rafael Gomes de Sousa CRB 3/1163*

ANÁLISE DE TEXTURA DE NÓDULOS PULMONARES PARA SISTEMAS CAD  
UTILIZANDO ÍNDICE DE DIVERSIDADE FUNCIONAL

WILLIAM DE OLIVEIRA TORRES

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

José Anatiel Gonçalves Santos Landim  
\_\_\_\_\_  
Prof. José Anatiel Gonçalves Santos Landim

Jonnison Lima Ferreira  
\_\_\_\_\_  
Prof. Jonnison Lima Ferreira

# Agradecimentos

Em primeiro lugar agradeço a Deus, pela vitória conquistada.

Aos meus pais, Djanira de Oliveira Torres e Everaldo Florentino Torres pelo incentivo e apoio dado no decorrer de toda a graduação.

Aos meus tios(as), primos(as), padrinho e madrinha pelas mensagens de incentivo e amparo familiar.

Agradeço a todos meus amigos e colegas, que foram de uma importância incalculável e que ofereceram apoio durante todo o curso.

Ao meu orientador Prof. Dr. Antonio Oseas de Carvalho Filho por se mostrar disponível para retirar dúvidas e auxiliar no desenvolvimento dessa atividade, proporcionando grandes oportunidades e crescimento acadêmico.

Por fim, gostaria de agradecer a todos que de forma direta ou indireta contribuíram para a realização desse objetivo, desse sonho.

*Conheça todas as teorias, domine todas as técnicas, mas ao tocar uma alma humana,  
seja apenas outra alma humana.*

*Carl Jung*

# Resumo

O câncer de pulmão é uma das principais causas de mortes por câncer no mundo. As chances de sobrevivência aumentam ao detectar a doença em seus estágios iniciais. Uma das melhores ferramentas usadas para diagnosticar é a tomografia computadorizada. O problema desse método é que, seu resultado depende diretamente do especialista que está analisando a imagem; esse processo é cansativo e pode levar ao erro. Sistemas de Detecção Auxiliado por Computação (*Computed-Aided Detection Systems* - CAD) vem se mostrando uma boa forma de auxílio aos especialistas. Esse trabalho apresenta o uso de descritores baseados no Índice de Diversidade Funcional (IDF) na etapa de redução de falsos positivos (FP) em tomografia computadorizadas. Para validar a eficiência da metodologia, foram usadas as métricas: acurácia, sensibilidade, especificidade, ROC e Kappa, alcançado 97.73%, 98.40%, 95.87%, 0.941, e 0.996, respectivamente. A metodologia proposta apresentou resultados satisfatórios e superiores ao estado da arte, com isso pode-se concluir que, o IDF é robusto e capaz de participar de um Sistema de Detecção Auxiliado por Computação real.

**Palavras-chaves:** Câncer de Pulmão, Imagens Médicas, Análise de Texturas, Diversidade Funcional.

# Abstract

The lung cancer is one of the leading causes of cancer deaths in the world. The chances of survival of this diagnose is higher when it is detected in the initial stages. One of the best tools used to perform the diagnosis is computerized tomography. The problem of this method is that it depends directly on the specialist who is analyzing the image; this process is tiring and can lead to error. Computer-Aided Detection Systems has emerged as a way to help the specialists. This work presents the use of descriptors based on functional diversity index (FDI) in the step of reducing false positives in computerized tomography. To validate the efficiency of the methodology, the following metrics were used: accuracy, sensitivity, specificity, ROC, and Kappa, and they achieved values of 97.73%, 98.40%, 95.87%, 0.941, and 0.996, respectively. The proposed methodology presented satisfactory results and superior to the state of the art, it can be concluded that, the use of the FDI proved to be a robust method and being able to participate in a real Computer-Aided Detection System.



# Lista de ilustrações

Figura 1 – Fluxograma do CAD usado como base do trabalho focado em nosso método de redução de FPs (caixa amarela). . . . .	21
Figura 2 – A Matriz Funcional exhibe a relação entre espécies $E$ e a importância delas em relação a uma Característica Funcional $CF$ . Após a Matriz Funcional ser gerada, é criada então a Matriz de Distâncias onde, utilizando o cálculo de Distância Euclidiana, é obtido a distância entre uma espécie e outra. . . . .	23
Figura 3 – Baseado na Matriz de Distâncias, um Dendrograma é gerado onde espécies são agrupadas de acordo com sua proximidade a outra. Por fim, é gerado o índice de diversidade funcional baseado na soma de todos os galhos do dendrograma. . . . .	24
Figura 4 – Uma visão geral do método proposto para extrair recursos de um VOI segmentado. IDF1, IDF2, IDF3 e IDF4 são as quatro características calculadas. . . . .	25
Figura 5 – Resultados de todas as distâncias juntas em cada quantização com o classificador <i>Random Forest</i> . Q5, Q7 e Q9 são a quantização com 5,7 e 9 bits, respectivamente. . . . .	29

# Lista de tabelas

Tabela 1 – Resumo dos trabalhos relacionados . . . . .	15
Tabela 2 – Níveis de precisão da classificação, segundo o índice Kappa. . . . .	20
Tabela 3 – Correlação entre os termos utilizados pela biologia e os termos adotados no método proposto. . . . .	24
Tabela 4 – Resultados dos testes das imagens com a quantização igual a 5 Bits. . .	27
Tabela 5 – Resultados dos testes das imagens com a quantização igual a 7 Bits. . .	28
Tabela 6 – Resultados dos testes das imagens com a quantização igual a 9 Bits. . .	28
Tabela 7 – Comparação do método proposto com trabalhos relacionados. . . . .	30

# Lista de abreviaturas e siglas

CAD	- <i>Computed-Aided Detection Systems</i>
IDF	- Índice Diversidade Funcional
TC	- Tomografia Computadorizada
VP	- Verdadeiro Positivo
VN	- Verdadeiro Negativo
FP	- Falso Positivo
FN	- Falso Negativo
VOI	- <i>Volumes of Interest</i>
GLCM	- <i>Gray level Co-occurrence Matrix</i>
CF	- Característica Funcional
E	- Espécie

# Sumário

<b>1</b>	<b>Introdução</b>	<b>12</b>
<b>2</b>	<b>Trabalhos Relacionados</b>	<b>14</b>
<b>3</b>	<b>Referencial Teórico</b>	<b>16</b>
3.1	Câncer de Pulmão	16
3.2	Tomografia Computadorizada	16
3.3	Processamento Digital de Imagens	17
3.4	Extração de Características	18
3.5	Classificação de Imagens	18
3.6	Validação dos Resultados	19
<b>4</b>	<b>Metodologia</b>	<b>21</b>
4.1	Aquisição das Imagens	21
4.2	Índice de Diversidade Funcional	22
4.3	Classificadores	24
<b>5</b>	<b>Resultados</b>	<b>27</b>
<b>6</b>	<b>Conclusão</b>	<b>31</b>
<b>7</b>	<b>Publicações</b>	<b>32</b>
	<b>Referências</b>	<b>33</b>

# 1 Introdução

No ano de 2018, o câncer de pulmão teve um total de 2.093,876 novos casos em todo o mundo, levando a 1.761,007 mortes, representando 11,6% e 18,4% dos casos e mortes de câncer, respectivamente. O câncer de pulmão, quando comparado com outros 32 tipos, é o mais comumente diagnosticado e também o mais letal (BRAY et al., 2018).

A detecção de nódulos pulmonares é crucial pois há uma grande chance de evoluir para câncer (Huang et al., 2013). A detecção de nódulos por tomografia computadorizada (TC) é uma tarefa complexa, tendo em vista que eles podem ter um contraste semelhante a outras estruturas de baixa densidade e tamanho pequeno em uma área complexa (por exemplo, conectada a vasos sanguíneos ou nas bordas do pulmão) (LEE; KOUZANI; HU, 2010). Outro fator que dificulta a detecção é que os especialistas têm um grande número de TCs para analisar. Esse processo exaustivo e repetitivo pode levar a distrações e resultar em um mal-entendido da análise, especialmente quando as imagens exibem simultaneamente outras anomalias. Portanto, esse tipo de análise é frequentemente vulnerável a erros (OST; FEIN; FEINSILVER, 2003).

Devido a complexidade da tarefa de detecção de nódulos pulmonares, Sistemas de Detecção Auxiliadas por Computação (*Computer Aided Detection Systems* - CAD) começaram a ser desenvolvidos. CADs usam técnicas computacionais com a finalidade auxiliar os especialistas no momento da análise dos exames. O CAD usado como base por esse trabalho possui 5 etapas em seu processo, são elas: 1) Aquisição de imagens: etapa onde as imagens médicas são obtidas; 2) Segmentação: nessa etapa obtemos as regiões de interesse na imagem, no caso de TC do pulmão são obtidos os nódulos pulmonares; 3) Redução de Falsos Positivos (FP): analisa todas as áreas de interesse obtidas na etapa anterior e então válida se o que foi segmentado realmente é de interesse (Contribuição deste trabalho); 4) Extração de características: responsável pela obtenção de atributos para caracterização da imagem; 5) Classificação: trata-se da última etapa do sistema, nela as características extraídas são utilizadas como entrada para um algoritmo de classificação, onde a saída é, usualmente, o diagnóstico da imagem ou da área de interesse.

O presente trabalho trata a utilização de atributos baseados em índice de diversidade funcional na etapa de redução de falso positivos em imagens de TC do pulmão. O objetivo é reduzir FPs na detecção dos nódulos pulmonares, ou seja, após a segmentação dos nódulos muitas regiões são falso positivos (não-nódulos) e isso acaba prejudicando as etapas posteriores de um sistema CAD.

Para a tarefa de redução de FPs entre nódulos e não-nódulos em imagens de TCs é proposto um descritor que trabalha a caracterização apenas em texturas contidas em regiões segmentadas em exames de TC. Para essa tarefa, os conceitos de índice de diversidade funcional, provindos da biologia, foram empregados para extração de atributos de

texturas. Optou-se por analisar exclusivamente a textura das imagens em decorrência de certas situações, como por exemplo: a) mesmo para os mais experientes especialistas, há regiões em que a diferenciação de contraste e textura entre tipos de tecidos pode passar despercebida; b) devido a natureza das imagens do exame, uma parte expressiva dos sistemas CAD fornece regiões arredondadas, abrindo assim uma maior chance de erros na classificação quando se utiliza à análise de forma para a descrição.

A principal contribuição desse trabalho é uma nova forma para se descrever texturas baseados em um índice ecológico, o índice de diversidade funcional para a redução de FPs em imagens TCs pulmonares. O uso desse descritor provê a possibilidade estimar o comportamento das texturas sobre cada imagem.

## 2 Trabalhos Relacionados

A segmentação pode gerar candidatos a nódulos incorretos, conhecidos como FP. O desafio enfrentado é reduzir o número de FPs sem perder os nódulos reais. Há trabalhos reconhecidos na literatura que abordam o mesmo problema enfrentado pelo método proposto. Para efeito de comparação, iremos listar alguns deles logo abaixo.

Em (Boroczky; Zhao; Lee, 2005) propõem um método de seleção de subconjuntos de características baseado em Algoritmos Genéticos para realiza a redução de FPs de nódulos pulmonares, ela utiliza 189 imagens, que a partir de cada imagem são extraídas 23 características, em seguida são selecionados os melhores atributos. O melhor resultado obtido foi utilizando 9 características, tendo o resultado de 92,4% de sensibilidade e 85,4% de especificidade usando a validação cruzada *leave-one-out*.

A proposta de (Setio et al., 2016) é utilizar a técnica *multi-view convolutional networks* (ConvNets) para a redução de FPs. Foram utilizadas um total de 888 imagens oriundas da base LIDC-IDRI. Atingindo uma sensibilidade de 87.9% da tarefa de redução de FPs.

No trabalho de (El-Regaily et al., 2017) os autores propõem o uso de um conjunto de redes neurais convolucionais (*Ensemble of Convolutional Neural Networks* E-CNNs) para a redução de FPs. São utilizadas 3 CNNs, cada uma diferenciando na quantidade de camadas e tamanho das imagens de entrada. Foram utilizadas um total de 400 imagens de CTs. Esse método obteve uma sensibilidade de 77,77%, especificidade de 69,5% e acurácia de 70,53%.

A proposta de (FILHO et al., 2017) realiza a classificação de nódulos pulmonares, para isso, ele usa medidas de forma, analisando sua forma usando diagramas de forma, medidas de proporção e uma análise baseada em cilindro. Além disso, usa o classificador de máquina de vetores de suporte (Support Vector Machine - SVM). O método obteve uma acurácia de 95.33%.

Para a redução de FPs em TCs pulmonares, (NAQI; SHARIF; JAFFAR, 2018) utilizam *deep learning* baseada em *autoencoder* (LIOU et al., 2014) e *softmax*. Um total de 1018 imagens originadas da base LIDC-IDRI. Na etapa de redução de FP, foi obtida uma sensibilidade de 95.6%.

Por fim, o trabalho de (Qin et al., 2018) aplica um modelo de CNN 3D baseado na *DenseNet* para redução de FPs. São utilizadas 888 imagens vindas da base do desafio LUNA16. Esse método obteve uma sensibilidade de 98.30% na tarefa de redução de FPs.

Para facilitar a visualização dos trabalhos relacionados, a Tabela 1 mostra de forma mais direta o trabalho, a base de imagens utilizada, a quantidade de exames na amostra e a técnica utilizada para a tarefa.

Alguns trabalhos apresentados anteriormente tratam exclusivamente da redução de FPs, já outros, a redução é apenas uma das etapas presentes no trabalho. Apesar de

Tabela 1: Resumo dos trabalhos relacionados

Trabalho	Base	Amostra	Técnica
<a href="#">Boroczky, Zhao e Lee (2005)</a>	Privada	189	Seleção de subconjuntos de características baseado em Algoritmos Genéticos.
<a href="#">Setio et al. (2016)</a>	LIDIC-IDRI	888	<i>Multi-view Convolutional Networks</i> (ConvNets).
<a href="#">El-Regaily et al. (2017)</a>	LIDIC-IDRI	400	Conjunto de redes neurais convolucionais ( <i>Ensemble of Convolutional Neural Networks</i> E-CNNs).
<a href="#">Filho et al. (2017)</a>	LIDIC-IDRI	833	Análise de forma usando diagramas de forma, medidas de proporção e uma análise baseada em cilindro.
<a href="#">Naqi, Sharif e Jaffar (2018)</a>	LIDIC-IDRI	1018	Usam <i>deep learning</i> baseada em <i>autoencoder</i> e softmax.
<a href="#">Qin et al. (2018)</a>	LUNA16	888	CNN 3D baseado na <i>DenseNet</i> .

obterem bons resultados, devemos levar em consideração que os trabalhos possuem algum tipo de problema como: 1) quantidade reduzida de imagens, 2) devido ao comportamento da doença, são geradas imagens com formas semelhantes entre nódulos e não-nódulos após a segmentação; 3) alto custo de processamento quando são utilizados algoritmos de *deep learning*. Dessa forma, utilizamos a base de imagens LIDIC-IDRI([III et al., 2011](#)) para contornar o problema da quantidade de imagens, e atributos de textura para que o método seja menos sensível à segmentação e, ao mesmo tempo, necessite de um poder de processamento menor que um algoritmo de *deep learning*.



## 3 Referencial Teórico

Com o propósito de facilitar a compreensão do trabalho aqui proposto, serão abordados neste capítulo conceitos e informações necessárias para a compreensão da proposta. O restante do capítulo está dividido nas seguintes seções: [3.1 Câncer de Pulmão](#), [3.2 Tomografia computadorizada](#), [3.3 Processamento Digital de Imagens](#), [3.4 Extração de características](#), [3.5 Classificação de Imagens](#) e por fim [3.6 Validação dos Resultados](#).

### 3.1 Câncer de Pulmão

O câncer de pulmão é o crescimento descontrolado das células no pulmão, levando então à formação de nódulos pulmonares. Segundo dados publicados em [Bray et al. \(2018\)](#), dentre os tipos que podem atingir ambos os sexos, o câncer de pulmão é o mais comumente diagnosticado representando aproximadamente 1 a cada 10 dos casos de diagnósticos e a principal causa de óbito, representando aproximadamente 1 a cada 5 mortes por câncer no mundo.

De acordo com [INCA \(2019\)](#), o tabagismo ou a exposição passiva ao tabaco são fatores importantes de risco que contribuem com o desenvolvimento do câncer de pulmão. Cerca de 85% dos casos diagnosticados feitos, o câncer está ligado ao consumo do tabaco ou de seus derivados.

O diagnóstico da doença pode ser feito pelos seguintes exames: Raio-X do tórax, complementado por TC e/ou broncoscopia. Uma vez obtida a confirmação da doença, é feito o estadiamento, que avalia o estágio de evolução, ou seja, verifica se a doença está restrita ao pulmão ou disseminada para outros órgãos. O estadiamento é feito por meio de vários exames, como biópsia pulmonar guiada por tomografia, biópsia por broncoscopia, tomografia de tórax, ressonância nuclear, PET-CT, cintilografia óssea, mediastinoscopia, ecobroncoscopia, entre outros ([INCA, 2019](#)).

Para um tratamento adequado, é necessário o diagnóstico histológico e o estadiamento para definir se a doença está localizada no pulmão ou se existem focos em outros órgãos. Para os pacientes com doença localizada, e, particularmente, sem linfonodo (gânglio) aumentado (íngua) no mediastino (região entre os dois pulmões), o tratamento é cirúrgico, seguido ou não de quimioterapia e/ou radioterapia ([INCA, 2019](#)).

### 3.2 Tomografia Computadorizada

Desde a década de 70, a TC vem sendo aclamada por ser uma ferramenta revolucionária na área de diagnósticos por imagens médicas. Muito disso se deve a forma com que a

imagem é capturada na TC. A principal diferença da TC para a radiografia convencional, é possibilitar a visão transversal do corpo do paciente ao especialista (SEERAM, 2010).

Como explicada por Seeram (2010), a TC usa um computador para processar e reconstruir dados de transmissão de raios-X coletados do paciente. Para isso, é utilizado conhecimento multidisciplinar baseado em física, matemática, engenharia e ciência da computação. Sua evolução técnica continuada é marcada por diversos avanços recentes que fazem da TC uma das ferramentas mais importantes do século XXI.

Por sua natureza, a adoção dessa técnica, permite que a análise seja feita em qualquer parte do corpo do paciente, além de que, esse tipo de exame trata-se de uma exame não invasivo.

### 3.3 Processamento Digital de Imagens

Com a evolução tecnológica ocorrida nos últimos anos, o desenvolvimento de novos algoritmos para lidar com sinais bidimensionais permitem cada vez mais aplicações na área de processamento digital de imagens, com isso, despertasse o interesse para estudos nessa área (ROCHA, 2014).

No processamento digital de imagens, as entradas e saídas são imagens, cujos procedimentos englobam a extração de atributos e melhoram o aspecto visual de certas características, fornecendo informações que possam ser suficientes para a percepção e interpretação humana e seus dados possam ser interpretados de forma automática através de máquinas (OLIVEIRA, 2013).

Uma imagem pode ser definida como uma função bidimensional, onde  $x$  e  $y$  são coordenadas espaciais e a amplitude de  $f$  no par de coordenadas  $f(x, y)$  é denominado intensidade ou nível de cinza. Quando  $x$  e  $y$  e a intensidade dos valores de  $f$  são finitos e discretos, tem-se então uma imagem digital clássica. Esses elementos possuem uma localização e um valor particular, chamados de *pixels* (SOUSA, 2011) e (SILVA, 2016).

As interpretações das informações podem ser feitas através de algoritmos computacionais, tomando como entrada uma imagem, que passa por técnicas de processamento digital de imagem para extração de características, após esta etapa os dados extraídos são usados para o reconhecimento ou classificação, considerando um catálogo de padrões ou um banco de padrões (CONCI; AZEVEDO; LETA, 2008).

Para análise de imagens digitais é fundamental que as regiões de interesse sejam separadas das demais, com isso evita-se que haja perda de processamento por parte da máquina com dados irrelevante, para o sistema em desenvolvimento (SOUSA, 2011).

## 3.4 Extração de Características

Na análise de imagens, visão computacional e reconhecimento de padrões, os sistemas necessitam-se reconhecer elementos ou objetos de interesse nas imagens, para isso, são estabelecidos parâmetros quantificáveis, que dependem de cor, posição, orientação, dimensões, textura, para poder ser realizado computacionalmente (CONCI; AZEVEDO; LETA, 2008).

A extração de características é uma forma utilizada para extrair dados em uma imagem que possa assim caracterizar a mesma, onde, a partir de um conjunto grande de características, seja extraída aquelas que são mais importantes para o problema, formando um conjunto mais reduzido de características por imagem digital (JR, 2014).

A extração de características tem como objetivo mensurar e quantificar padrões perceptíveis ou não visualmente, com a finalidade de categorizar objetos nas imagens, de acordo com aspectos morfológicos de interesse (JR, 2014).

Características de textura são determinadas como um conjunto de alguns padrões existentes em superfícies físicas, perceptíveis ao olho humano, trazendo grande quantidade de informação sobre a natureza da superfície, como suavidade, rugosidade e regularidade. A análise feita através de texturas é útil em aplicações, por se aproximar da avaliação feita pelo sistema visual humana (JR, 2014).

Após obter o conjunto de atributos que caracterizam a imagem, é a vez de verificar a eficácia dos mesmos. Para essa tarefa aplicamos os dados em um algoritmo de classificação e, baseado em sua saída, é possível verificar se as características utilizadas são capazes de descrever a imagem de forma eficaz.

## 3.5 Classificação de Imagens

O processo de classificação pode ser realizado considerando classes previamente definidas ou não definidas. Nas classes previamente definidas cada parâmetro que caracteriza cada classe é obtido através de treinamento, executados antes da aplicação do algoritmo de classificação. Com as classes não definidas, não se dispõe de parâmetros ou informações previamente coletadas a aplicação do algoritmo de classificação (JR, 2012).

A classificação consiste em reconhecer novos objetos, a partir de descritores gerados dos mesmos, tomando decisões a partir da extração de características do mundo real em imagens, isso é feito por meio de indagações simples a respeito dos parâmetros extraídos, ou por meio de algoritmos de inteligência artificial (CONCI; AZEVEDO; LETA, 2008).

No processamento de imagens digitais, classificar uma imagem consiste em pegar um conjunto de características extraídas de um grupo de imagens do objeto de interesse e treinar um sistema para que possa reconhecer novos objetos a partir dessas características.

O reconhecimento de padrões ou características, usa alguns métodos ou técnicas, desde

as mais simples, baseadas em vizinhos mais próximos, funções discriminantes, árvore de decisão, até as mais complexas, baseadas em redes *baysianas*, lógica *fuzzy* e redes neurais, para a classificação de imagens de acordo com os descritores extraídos (CONCI; AZEVEDO; LETA, 2008).

Para o desenvolvimento deste trabalho, a classificação será feita através da aplicação de algoritmos de *machine learning*, que implementa técnicas de reconhecimento de padrões redução de falsos positivos de nódulos pulmonares em imagens de tomografia computadorizadas pulmonares.

### 3.6 Validação dos Resultados

Para validar o resultado obtido da etapa de classificação, foram utilizadas estatísticas/métricas que nos permitem descobrir a eficiência do método para classificar corretamente os VOIs que possuem lesões sobre os VOIs que não possuem a lesão. As métricas foram, acurácia (Ac), sensibilidade (Sens), especificidade (Esp) (DUDA; HART, 1973), área sob a curva ROC (ROC) (MATOS, 2017), e índice Kappa (COHEN, 1968).

Os cálculos de acurácia, sensibilidade e especificidade são feitos em cima de quatro valores que são fornecidos ao final da avaliação, são eles verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN). Nesse trabalho, o caso positivo é atribuído para os nódulos e os não-nódulos são dados como caso negativo, ou seja, o VP é o nódulo (caso positivo) que foi classificado como positivo, o FP é um não-nódulo (caso negativo) que foi classificado como positivo, o VN é um não-nódulo (caso negativo) que foi classificado negativo, por fim, o FN é quando um nódulo (caso positivo) é classificado como negativo.

Sensibilidade é uma métrica usada para medir a eficácia do método em identificar corretamente os casos VPs, ou seja, o qual eficaz é o método para identificar os nódulos. Para obter o valor da sensibilidade, é utilizada a Equação 3.1, onde a sensibilidade é igual ao verdadeiro positivo dividido pela soma dos verdadeiros positivos mais o falsos negativos.

$$Sens = \frac{VP}{VP + FN} \quad (3.1)$$

A especificidade é a métrica adotada para caracterizar o quão eficaz o método é na identificação correta de VN, ou seja, o quão eficaz é o método para identificação de não-nódulos. Para obter seu valor é usada a Equação 3.2, onde a especificidade é dada pela divisão dos verdadeiros negativos pela soma dos verdadeiros negativos mais os falsos positivos.

$$Esp = \frac{VN}{VN + FP} \quad (3.2)$$

A acurácia calcula a taxa de sucesso, ou seja, a porcentagem de VPs e VNs no teste, pode ser calculada de acordo com a Equação 3.3, onde a acurácia é dada pela soma dos verdadeiros positivos mais os verdadeiros negativos dividido pela soma dos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

$$Ac = \frac{VP + VN}{VP + VN + FN + FP} \quad (3.3)$$

A área sobre a curva ROC é uma das melhores formas assimilar a associação muitas vezes contraditória entre especificidade e sensibilidade. Ela traduz a probabilidade de que, diante de um caso positivo e um outro negativo, a regra do classificador tenda ao caso positivo. (MATOS, 2017)

Por fim, usamos também o índice Kappa (K) para avaliar os resultados (COHEN, 1968). Quanto mais próximo de 1 for o valor Kappa obtido, maior será a concordância. De acordo com Landis e Koch (1977), há uma categorização dos níveis de desempenho de classificação pelo índice Kappa que pode ser vista na Tabela 2.

Tabela 2: Níveis de precisão da classificação, segundo o índice Kappa.

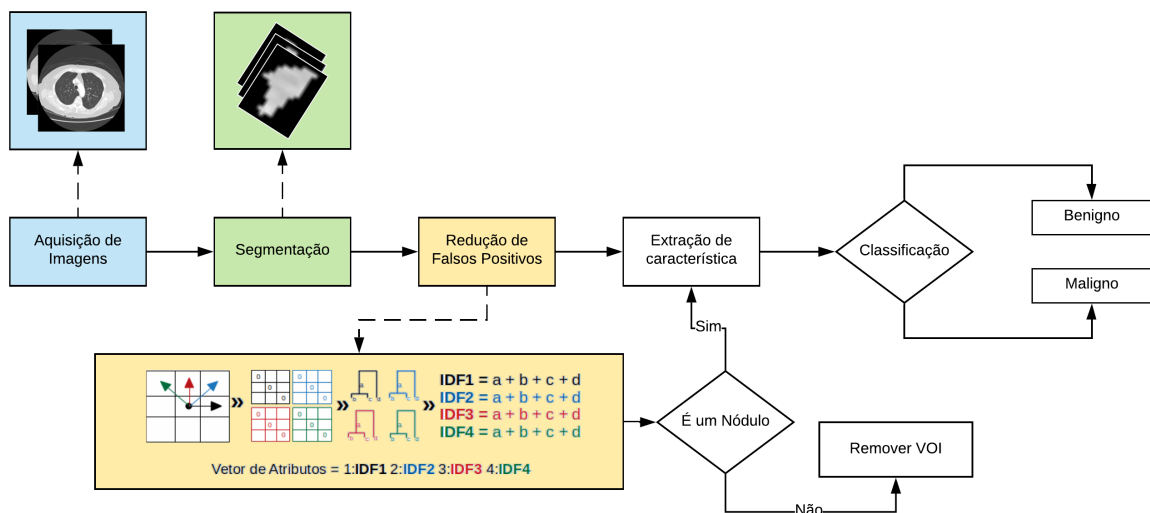
Índice Kappa (k)	Qualidade
$K < 0.2$	Pobre
$0.2 \leq K < 0.4$	Razoável
$0.4 \leq K < 0.6$	Bom
$0.6 \leq K < 0.8$	Muito Bom
$K \geq 0.8$	Excelente

Fonte: Landis e Koch (1977)

## 4 Metodologia

Neste capítulo, será exibido a metodologia proposta para classificação de regiões suspeitas em nódulo e não nódulo, isto é, para redução de falsos positivos. A primeira etapa foi a aquisição de imagens, elas foram adquiridas a partir da base LIDC-IDRI (III et al., 2011) usando o método de (FILHO et al., 2014) para segmentação dos *Volumes of Interest* (VOI). Após esta etapa, aplicou-se a extração de características utilizando os índices de diversidade funcional (Torres et al., 2018). Para finalizar, é realizada a classificação e a validação dos dados extraídos. A Figura 1 apresenta uma visão geral de um CAD completo, com ênfase na etapa de redução de falsos positivos que é foco desse trabalho.

Figura 1: Fluxograma do CAD usado como base do trabalho focado em nosso método de redução de FPs (caixa amarela).



Fonte: Próprio autor.

### 4.1 Aquisição das Imagens

A base de imagens usada para esse trabalho foi a LIDC-IDRI (III et al., 2011), trata-se de um recurso de acesso gratuito e que foi projetado para treinar e avaliar com precisão os CADs para detecção de nódulos e diagnósticos de câncer de pulmão. Além disso, diversos trabalhos na literatura também usam esse conjunto de dados, como podemos ver em (FILHO et al., 2017), (FILHO et al., 2017), e (Dey; Lu; Hong, 2018), permitindo uma melhor comparação dos resultados.

As TCs contidas nessa base têm origem em tomógrafos diferentes, isso proporciona uma dificuldade adicional na tarefa de classificação de nódulos pulmonares, pois cada

equipamento produz as imagens de maneira diferente. Devido a isso, usá-los torna a generalização uma tarefa mais complexa para algoritmos. Com isso, se uma metodologia conseguir uma alta taxa de acertos, provavelmente terá bons resultados em protocolos diferentes.

As amostras utilizadas neste trabalho apresentam dois tipos de VOIs: nódulo e não-nódulo. Para extrair os VOIs dos nódulos, utilizou-se as marcações fornecidas pelos especialistas da *Lung Image Database Consortium* (LIDC). Para a segmentação dos não-nódulos foi utilizada a metodologia de (FILHO et al., 2014). Assim, foi possível usar candidatos a não-nódulos que foram encontrados em um cenário real e nódulos fornecidas pelos especialistas.

O método de Filho et al. (2014) é composta por três etapas. A primeira etapa, é realizada a segmentação e reconstrução do parênquima pulmonar, seguida por um realce para destacar suas estruturas. Na segunda etapa, os candidatos a nódulos são segmentados utilizando clusterização. Finalmente, no terceiro estágio, as características desproporção esférica, densidade esférica, esfericidade, distância radial ponderada, alongamento e índice de forma radial *Boyce-Clark* são extraídos, selecionados e classificados usando um *Support Vector Machine*.

Quando juntos, os VOIs não-nódulos vindos do método (FILHO et al., 2014) e as VOIs nódulos vindas das marcações dos especialistas, foram obtidas 17741 VOIs não-nódulos e 6407 VOIs de nódulos formando assim as 24148 VOIs que foram utilizada nos testes presente trabalho.

## 4.2 Índice de Diversidade Funcional

Na ecologia, é comum o uso de índices de riqueza e abundância das espécies como uma forma de caracterização da biodiversidade. No entanto, estudos mais recentes focados em diversidade vêm empregando o conceito de diversidade funcional. A diversidade funcional é considerada uma das melhores formas de prever a produtividade e a vulnerabilidade de um ecossistema do que outros índices (SCHLEUTER et al., 2010).

Conforme Tilman (2001), a diversidade funcional no funcionamento das comunidades expressa a influência que as espécies e suas características possuem. Sendo assim, ao se mensurar a diversidade funcional, se consegue ter ideia da pluralidade de características funcionais que compõem os fenótipos dos organismos que agem nos processos da comunidade.

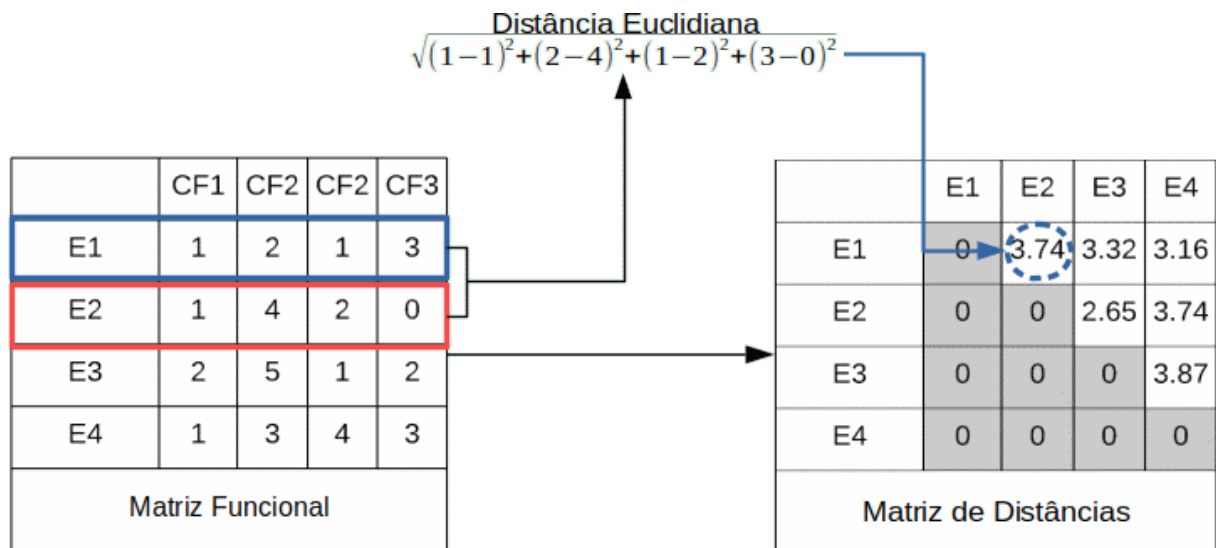
Para exemplificar, imaginemos duas comunidades ambas com a mesma quantidade de espécies, vamos chamar elas de *A* e *B*. Se *A* for composta apenas por aves e *B* for composta por lagartos, mamíferos e aves. *B* será mais diversificado pois há mais espécies funcionalmente distintas do que *A* (CIANCIARUSO; SILVA; BATALHA, 2009).

Calcular a diversidade funcional é calcular a importância que cada espécie (E) desempenha em cada característica funcional (CF) na comunidade em que está presente. Diferentemente das características fenotípicas, a diversidade funcional é calculada independentemente da quantidade de membros da espécie na comunidade. Isso quer dizer que, considerando uma imagem como sendo uma comunidade, ao invés de, considerarmos unicamente as relações entre as espécies, consideramos as relações que uma espécie têm com seus vizinhos. Isso nos permite caracterizar e explorar texturas sobre certa região.

A Diversidade Funcional (IDF), consiste na soma dos comprimentos dos braços de um dendrograma funcional, ou seja, um dendrograma gerado a partir de uma matriz de espécies x características funcionais (PETCHEY; GASTON, 2002).

Para o cálculo do IDF em uma imagem, primeiramente gera-se uma matriz de coocorrência de níveis de cinza (*Gray level Co-occurrence Matrix* - GLCM) (V.; UNNIKRISSH-NAN; BALAKRISHNAN, 2012), chamada matriz funcional em nosso método, como mostramos na Figura 2. Depois disso, criamos a matriz de distância (Figura 2) e logo após o dendrograma (Figura 3) é gerado pelo agrupamento da menor distância entre as espécies da matriz de distância e, finalmente, a soma de seus galhos está resultando em uma característica única.

Figura 2: A Matriz Funcional exibe a relação entre espécies E e a importância delas em relação a uma Característica Funcional CF. Após a Matriz Funcional ser gerada, é criada então a Matriz de Distâncias onde, utilizando o cálculo de Distância Euclidiana, é obtido a distância entre uma espécie e outra.



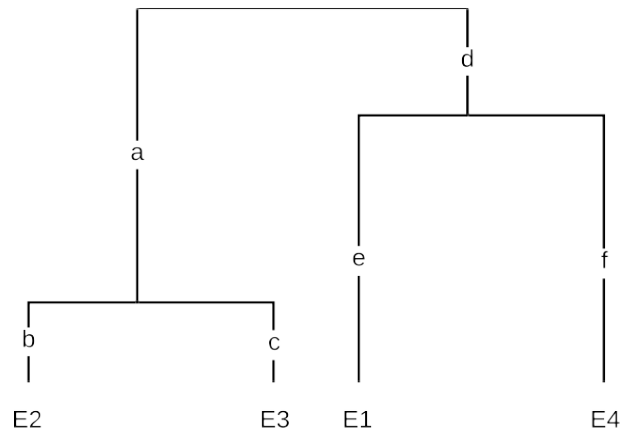
Fonte: Próprio autor.

Para uma melhor compreensão, a Tabela 3 exibe a correlação entre os termos utilizados pela biologia e os termos adotados no método proposto.

Para a extração de características de textura utilizando o IDF, reduziu-se a imagem que originalmente possuem 12 bits para 9, 7, e 5 bits utilizando a técnica de *quantization* (GERSHO, 1978), para reduzir o tamanho da matriz de coocorrência e reduzir o tempo de processamento. Foram calculadas as matrizes de coocorrência predefinidas para



Figura 3: Baseado na Matriz de Distâncias, um Dendrograma é gerado onde espécies são agrupadas de acordo com sua proximidade a outra. Por fim, é gerado o índice de diversidade funcional baseado na soma de todos os galhos do dendrograma.



$$\begin{aligned} \text{IDF}(E1 \text{ e } E2) &= a + b + d + e \\ \text{IDF}(E3 \text{ e } E4) &= a + c + d + f \\ \text{IDF}^* &= a + b + c + d + e + f \end{aligned}$$

Fonte: Próprio autor.

Tabela 3: Correlação entre os termos utilizados pela biologia e os termos adotados no método proposto.

Termos de Biologia	Termos do método
Comunidade	VOI
Espécies	Valor de <i>pixel</i> da comunidade
Indivíduos	Número de <i>pixels</i> de uma espécie
Características fenotípicas das espécies	Relação de textura entre espécies

quatro angulações ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ ) distância entre *pixels*  $D$  variando entre 1 a 3. Os índices gerados nas etapas subsequentes são armazenados em um vetor de características, para serem classificados posteriormente. Para facilitar a compreensão, a Figura 4 ilustra como funciona o algoritmo responsável pela etapa de extração de características.

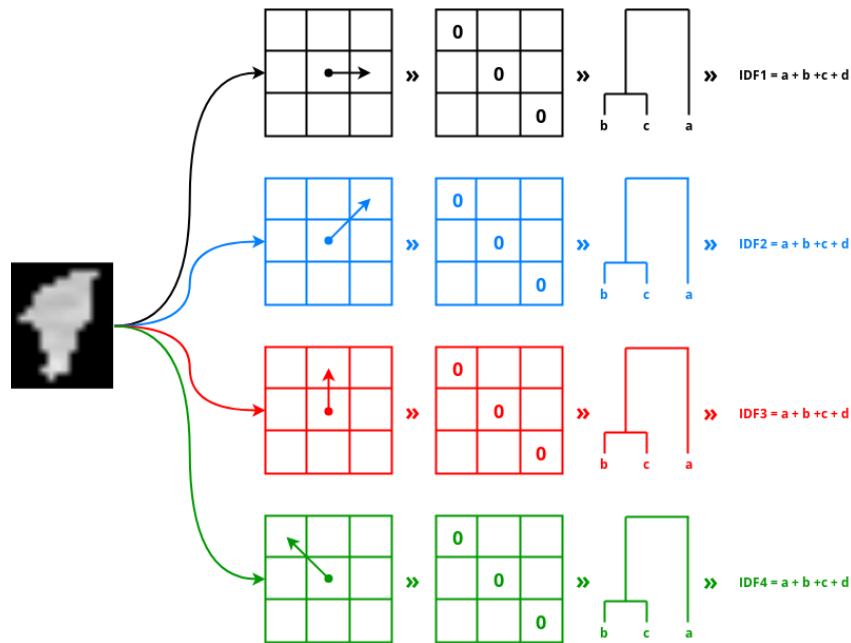
### 4.3 Classificadores

A classificação das características obtidas foram feitas pela suíte de algoritmos de aprendizagem de máquinas e mineração de dados WEKA (WITTEN et al., 2016), que contém ferramentas para pré-processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização.

Os experimentos foram realizadas com a validação cruzada ( $k = 10$ ). Este método dividiu os dados em  $k$  grupos; usa  $k-1$  grupos para treinar o modelo e um para o teste. Repetimos esse processo  $k$  vezes, sempre mudando o grupo de teste. No final, todos os grupos são aplicados como um conjunto de testes e geramos uma pontuação média.

Para classificação, foi utilizado os classificadores: *Reduced Error Pruning Tree* (REP-Tree) (KALMEGH, 2015), *Instance-Bases learning with parameter  $\mathbf{K}$*  (IBK) (AHA; KIBLER, 1991), *Random Forest* (BREIMAN, 2001) e SVM (CORTES; VAPNIK, 1995).

Figura 4: Uma visão geral do método proposto para extrair recursos de um VOI segmentado. IDF1, IDF2, IDF3 e IDF4 são as quatro características calculadas.



Fonte: Próprio autor.

Esses algoritmos foram escolhidos a partir de um conjunto inicial de algoritmos candidatos devido à sua aplicação em trabalhos relacionados. Esses classificadores são robustos à presença de ruído, apresentando bons resultados na categorização de padrões de nódulos e não-nódulos em imagens de TC de pulmão. Os classificadores usados aqui tiveram seus parâmetros aprimorados através de tentativas e erros; no final, os parâmetros utilizados foram:

- **REPTree:** foi usado o número mínimo de instâncias igual a 5, a variação mínima para uma divisão em 0.001, o número de *folds* como 3, o *seed* para o embaralhamento aleatório de dados é 1 e *no pruning*.
- **IBK:** o número de vizinhos é 3, o número máximo de instâncias de treinamento mantidas é 0, foi habilitado a seleção do número de vizinhos mais próximos entre 1 e o valor  $k$  especificado usando a avaliação de espera única nos dados de treinamento, usando o algoritmo *Linear NN-Search* para o algoritmo de busca de vizinhos mais próximo usando a distância euclidiana.
- **Random Forest:** o tamanho de cada bolsa é 100, o número de iterações é 1000, o número de *slots* de execução é 3, o número de atributos a serem investigados aleatoriamente é 0, o número mínimo de instâncias por folha é 1.0, a proporção de variação numérica mínima de classe de a variação do trem para a divisão é 0.001 e a semente do gerador de números aleatórios é 1.

- **SVM:** utilizou-se o tipo de *kernel* SVM C-SVC ( $C = 1.0$ ), o tipo de função do *kernel* é polinomial com grau 3, foi definido o gama na função do *kernel* igual a 0,0, definimos o parâmetro *nu* a 0,5, o *epsilon-SVR* é 0.001, a tolerância do critério de terminação é 0,001, o *epsilon* na função de perda do *epsilon-SVR* é 0,1.

## 5 Resultados

Para validar o método, foi usada a base de imagens LIDC-IDRI. A extração das características da textura foi realizada a partir dos descritores desenvolvidos utilizando o IDF e classificando as imagens em nódulos e não-nódulos, utilizando os classificadores com validação cruzada ( $k = 10$ ).

Foram implementadas duas formas de testes. Na primeira forma, as imagens foram classificadas usando todas as distâncias em cada quantização separadamente (Tabelas 4, 5 e 6). Na segunda forma, os vetores de características de todas as distâncias de cada quantização foram combinados e depois classificados (Figura 5). As estatísticas apresentadas aqui servem para aumentar a compreensão dos resultados; no entanto, para escolher os melhores e os piores resultados, usamos o índice *Kappa* como métrica.

As tabelas onde estão os resultados dos testes, as siglas adotadas *D*, *Ac*, *Sens*, *Esp* e *ROC* representam Distância, Acurácia, Sensibilidade, Especificidade e área sob a curva ROC, respectivamente.

Tabela 4: Resultados dos testes das imagens com a quantização igual a 5 Bits.

D	Classificador	Ac(%)	Sens(%)	Esp(%)	ROC	Kappa
1	<b>Random Forest</b>	<b>90,70</b>	<b>91,82</b>	<b>87,02</b>	<b>0,9450</b>	<b>0,7518</b>
	REPTree	89,11	90,51	84,34	0,9060	0,7074
	SVM	74,49	74,27	92,12	0,5200	0,0585
	IBK	82,88	83,29	80,61	0,7550	0,4927
2	Random Forest	88,01	91,16	76,85	0,9250	0,6562
	REPTree	86,57	91,02	74,45	0,8990	0,6405
	SVM	74,93	76,57	58,22	0,5720	0,1839
	IBK	80,51	83,79	69,48	0,7750	0,4830
3	Random Forest	52,39	92,46	23,11	0,9250	0,6562
	REPTree	84,71	92,55	67,77	0,8990	0,6305
	SVM	76,06	80,49	56,92	0,6460	0,322
	IBK	79,88	86,24	62,14	0,7750	0,4830

-Negrito é o melhor resultado.

Fonte: Próprio autor.

A Tabela 4 apresenta os resultados dos testes realizados com as imagens em quantização em 5 bits e cada distância separada. A melhor combinação foi o classificador *Random Forest* com a distância 1, tendo alcançado a acurácia, sensibilidade, especificidade, área ROC e Kappa igual a 90.70%, 91.82%, 87.02%, 0.945 e 0.7518 respectivamente. Já a pior foi com classificador SVM com a distância 1, tendo atingido a acurácia, sensibilidade, especificidade, área ROC e *Kappa* igual a 74.49%, 74.27%, 92.12%, 0.5200 e 0.0585 respectivamente.

Tabela 5: Resultados dos testes das imagens com a quantização igual a 7 Bits.

D	Classificador	Ac(%)	Sens(%)	Esp(%)	ROC	Kappa
1	<b>Random Forest</b>	<b>96,15</b>	<b>97,17</b>	<b>93,27</b>	<b>0,9910</b>	<b>0,9008</b>
	REPTree	95,57	96,47	92,98	0,8852	0,8730
	SVM	74,42	74,22	92,62	0,5190	0,0548
	IBK	88,87	91,39	81,23	0,8950	0,7081
2	Random Forest	93,25	97,21	83,66	0,9810	0,9321
	REPTree	92,65	96,88	82,53	0,9690	0,8179
	SVM	74,88	76,50	58,05	0,5710	0,1804
	IBK	87,98	86,05	99,08	0,7750	0,6408
3	Random Forest	93,23	97,21	83,66	0,9810	0,8321
	REPTree	89,66	97,44	74,20	0,9540	0,7550
	SVM	75,82	80,22	56,36	0,6410	0,3129
	IBK	86,38	93,75	70,43	0,9000	0,6707

-Negrito é o melhor resultado.

Fonte: Próprio autor.

A Tabela 5 apresenta os resultados dos testes realizados com as imagens em quantização em 7 bits e cada distância separada. A melhor combinação foi o classificador *Random Forest* com a distância 1, tendo alcançado a acurácia, sensibilidade, especificidade, área ROC e *Kappa* igual a 96.15%, 97.17%, 93.27%, 0.991 e 0.9008 respectivamente. Já a pior foi com classificador SVM com a distância 1, tendo o atingido a acurácia, sensibilidade, especificidade, área ROC e *Kappa* igual a 74.42%, 74.22%, 92.62%, 0.5190 e 0.0548 respectivamente.

Tabela 6: Resultados dos testes das imagens com a quantização igual a 9 Bits.

D	Classificador	Ac(%)	Sens(%)	Esp(%)	ROC	Kappa
1	Random Forest	88,05	91,08	82,60	0,9420	0,7099
	REPTree	87,09	89,61	78,92	0,8960	0,6565
	SVM	74,44	74,23	93,33	0,5190	0,0552
	IBK	80,98	85,28	66,43	0,7890	0,4893
2	<b>Random Forest</b>	<b>89,88</b>	<b>94,13</b>	<b>79,08</b>	<b>0,9550</b>	<b>0,7457</b>
	REPTree	88,49	93,08	76,74	0,9280	0,7103
	SVM	74,88	76,49	58,04	0,5710	0,1800
	IBK	80,79	88,37	65,72	0,7990	0,4874
3	Random Forest	88,22	96,30	72,09	0,9510	0,7208
	REPTree	87,42	95,65	70,94	0,9320	0,7017
	SVM	75,82	80,22	56,36	0,6410	0,3129
	IBK	81,80	88,55	64,71	0,9320	0,5429

-Negrito é o melhor resultado.

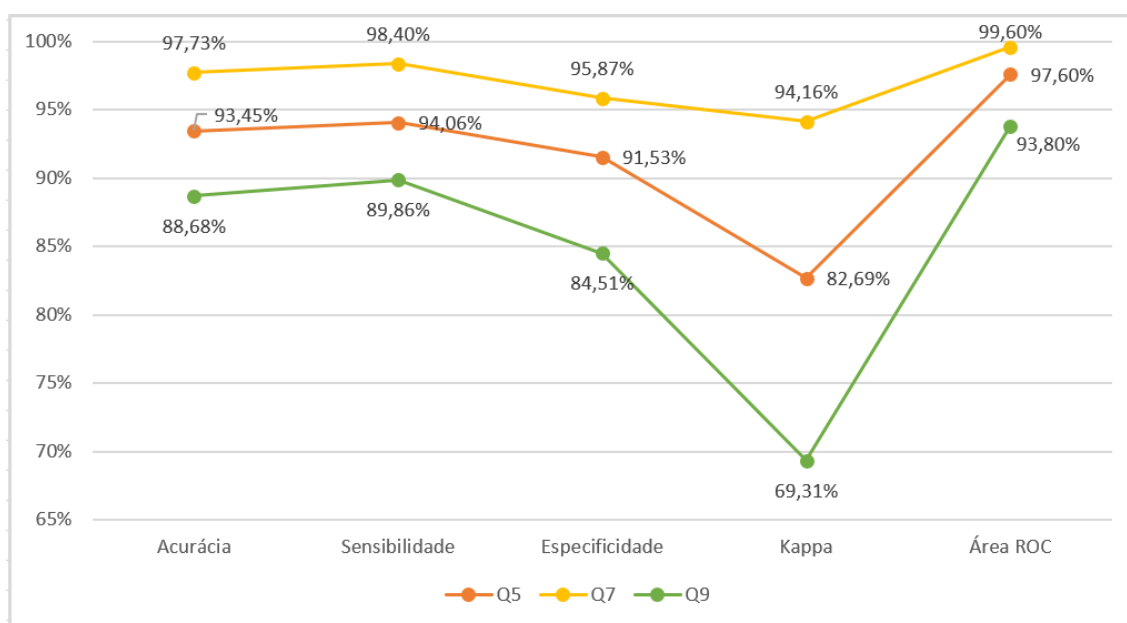
Fonte: Próprio autor.

A Tabela 6 apresenta os resultados com as imagens em quantização em 9 bits e cada distância. A melhor combinação foi o classificador *Random Forest* com a distância 2, tendo

alcançado a acurácia, sensibilidade, especificidade, área ROC e Kappa igual a 89.88%, 94.13%, 79.08%, 0.955 e 0.7457, respectivamente. Já o pior foi com classificador SVM com a distância 1, tendo o atingido a acurácia, sensibilidade, especificidade, área ROC e *Kappa* igual a 74.44%, 74.23%, 93.33%, 0.519 e 0.0552, respectivamente.

Por fim, foi concatenado os recursos em todas as distâncias para cada experimento de quantização. O classificador *Random Forest* também obteve o melhor resultado neste experimento. Na figura 5 mostramos um gráfico onde é possível ver a comparação entre cada quantização no melhor classificador.

Figura 5: Resultados de todas as distâncias juntas em cada quantização com o classificador *Random Forest*. Q5, Q7 e Q9 são a quantização com 5, 7 e 9 bits, respectivamente.



Fonte: Próprio autor.

Como mostra o gráfico da Figura 5, o melhor resultado entre as quantizações com todas as distâncias juntas ficou com a quantização de 7 bits, onde a acurácia, sensibilidade, especificidade, *Kappa* e área ROC igual a 97,73%, 98,40%, 95,87%, 0.9416 e 0.996 respectivamente. Já o pior foi com a quantização em 9 bits, tendo atingido a acurácia, sensibilidade, especificidade, *Kappa* e área ROC igual a 88,68%, 89,86% , 84,51%, 0.6931 e 0.938 respectivamente.

A tarefa de comparação dos trabalhos relacionados não é uma tarefa trivial, pois, muitos dos métodos propostos utilizam bases distintas, técnicas com abordagens diferentes. Todos estes fatores devem ser levados em consideração no momento da comparação, assim, não é possível fazer uma comparação fidedigna. Dessa forma, o intuito com a Tabela 7 é apenas de apresentar um levantamento quantitativo das métricas usadas pelos autores na avaliação dos resultados em cada trabalho.

Como mostrado na Tabela 7, os resultados obtidos pelo método proposto são promissores e podem ser comparáveis com os demais trabalhos da literatura. Além disso,

Tabela 7: Comparação do método proposto com trabalhos relacionados.

Trabalho	Base	Amostra	Ac(%)	Sens(%)	Esp(%)	ROC	Kappa
Boroczky, Zhao e Lee (2005)	Privada	189	-	92,40	85,40	-	-
Setio et al. (2016)	LIDC-IDRI	888	-	89,70	-	-	-
El-Regaily et al. (2017)	LIDC-IDRI	400	70,53	77,77	69,50	-	-
Filho et al. (2017)	LIDC-IDRI	833	95,33	-	-	-	-
Naqi, Sharif e Jaffar (2018)	LIDC-IDRI	1018	-	95,60	-	-	-
Qin et al. (2018)	LUNA16	888	-	98,30	-	-	-
<b>IDF</b>	<b>LIDC-IDRI</b>	<b>833</b>	<b>97,73</b>	<b>98,40</b>	<b>95,87</b>	<b>0,941</b>	<b>0,996</b>

-Negrito é o melhor resultado.

Fonte: Próprio autor.

superamos o estado da arte para redução de falso positivos em imagens de TC do pulmão para detecção de nódulos.

## 6 Conclusão

O emprego do índices de diversidade funcional na tarefa de redução de falsos positivos nas imagens de TC, especificamente, na caracterização das propriedades de textura das VOIs candidatas a nódulo pulmonar. O método proposto obteve-se resultados promissores, uma acurácia de 97,73%, área sob curva ROC de 94,10% e índice *Kappa* de 99,60%. Com base nestes resultados o método proposto foi superior ao estado da arte.

A metodologia apresentada neste trabalho poderá integrar-se a uma ferramenta CAD, sendo aplicada em casos reais e atuais na detecção de nódulos pulmonares, beneficiando assim o especialista que poderá utilizar a ferramenta para ter auxílio no diagnóstico, tendo assim, uma segunda opinião.

Como trabalhos futuros, pretende-se: usar a metodologia para diagnóstico dos nódulos em benigno e maligno; implantação de mais índices funcionais para caracterização de texturas em imagens; aplicação do método juntamente aos outros índices funcionais a diferentes tipos de bases de imagens



## 7 Publicações

William de Oliveira Torres, Antonio Oseas de Carvalho Filho, Alcilene Dalilia de Sousa, Francisco Airton Pereira da Silva: *Functional Diversity applied to the false positive reduction in breast tissues based on digital mammography*. **ISCC - IEEE Symposium on Computers and Communications**, Natal. Rio Grande do Norte. 2018.

## Referências

- AHA, D.; KIBLER, D. Instance-based learning algorithms. *Machine Learning*, v. 6, p. 37–66, 1991. Citado na página 24.
- Boroczky, L.; Zhao, L.; Lee, K. P. Feature subset selection for improving the performance of false positive reduction in lung nodule cad. In: *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*. [S.l.: s.n.], 2005. p. 85–90. ISSN 1063-7125. Citado 3 vezes nas páginas 14, 15 e 30.
- BRAY, F. et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, v. 68, n. 6, p. 394–424, 2018. Citado 2 vezes nas páginas 12 e 16.
- BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. Citado na página 24.
- CIANCIARUSO, M. V.; SILVA, I. A. A.; BATALHA, M. A. A. Diversidades filogenética e funcional: novas abordagens para a Ecologia de comunidades. In: . [S.l.]: scielo, 2009. p. 93 – 103. Citado na página 22.
- COHEN, J. Weighted kappa - nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological bulletin*, v. 70, p. 213–20, 11 1968. Citado 2 vezes nas páginas 19 e 20.
- CONCI, A.; AZEVEDO, E.; LETA, F. *Computação Gráfica - Teoria e Prática*. [S.l.: s.n.], 2008. ISBN 10: 85-352-2329-0. Citado 3 vezes nas páginas 17, 18 e 19.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, Sep 1995. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/BF00994018>>. Citado na página 24.
- Dey, R.; Lu, Z.; Hong, Y. Diagnostic classification of lung nodules using 3d neural networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. [S.l.: s.n.], 2018. p. 774–778. ISSN 1945-8452. Citado na página 21.
- DUDA, R. O.; HART, P. E. *Pattern Classification and Scene Analysis*. New York: Wiley-Interscience Publication, 1973. Citado na página 19.
- El-Regaily, S. A. et al. Lung nodule segmentation and detection in computed tomography. In: *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*. [S.l.: s.n.], 2017. p. 72–78. Citado 3 vezes nas páginas 14, 15 e 30.
- FILHO, A. O. d. C. et al. 3d shape analysis to reduce false positives for lung nodule detection systems. *Medical & Biological Engineering & Computing*, v. 55, n. 8, p. 1199–1213, Aug 2017. Disponível em: <<https://doi.org/10.1007/s11517-016-1582-x>>. Citado 4 vezes nas páginas 14, 15, 21 e 30.
- FILHO, A. O. de C. et al. Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial Intelligence in Medicine*, v. 60, n. 3, p. 165 – 177, 2014. ISSN 0933-3657. Citado 2 vezes nas páginas 21 e 22.

- FILHO, A. O. de C. et al. Lung-nodule classification based on computed tomography using taxonomic diversity indexes and an svm. *Journal of Signal Processing Systems*, v. 87, n. 2, p. 179–196, May 2017. ISSN 1939-8115. Disponível em: <[https://doi.org/10-1007/s11265-016-1134-5](https://doi.org/10.1007/s11265-016-1134-5)>. Citado na página 21.
- GERSHO, A. Principles of quantization. *IEEE Transactions on Circuits and Systems*, v. 25, n. 7, p. 427–436, Jul 1978. ISSN 0098-4094. Citado na página 23.
- Huang, P. et al. A classification system of lung nodules in ct images based on fractional brownian motion model. In: *2013 International Conference on System Science and Engineering (ICSSSE)*. [S.l.: s.n.], 2013. p. 37–40. ISSN 2325-0925. Citado na página 12.
- III, S. G. A. et al. The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, v. 38, n. 2, p. 915–931, 2011. Citado 2 vezes nas páginas 15 e 21.
- INCA. *Câncer de pulmão*. 2019. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-pulmao>>. Citado na página 16.
- JR, G. B. *Classificação de tecidos da mama a partir de imagens mamográficas em massa e não massa usando Índice de Diversidade de McIntosh e máquinas de vetores de suporte*. Tese (Dissertação de Mestrado. Programa de Pós-graduação em Engenharia de Eletricidade) — Universidade Federal do Maranhão, 2012. Citado na página 18.
- JR, G. B. *Detecção de regiões de massas em mamografias usando índices de diversidade, geoestatística e geometria côncava*. Tese (Tese de Doutorado. Programa de Pós-graduação em Engenharia de Eletricidade) — Universidade Federal do Maranhão, 2014. Citado na página 18.
- KALMEGH, S. Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news. In: . [S.l.: s.n.], 2015. Citado na página 24.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, [Wiley, International Biometric Society], v. 33, n. 1, p. 159–174, 1977. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2529310>>. Citado na página 20.
- LEE, S.; KOUZANI, A.; HU, E. Random forest based lung nodule classification aided by clustering. *Computerized Medical Imaging and Graphics*, v. 34, n. 7, p. 535 – 542, 2010. ISSN 0895-6111. Citado na página 12.
- LIOU, C.-Y. et al. Autoencoder for words. *Neurocomputing*, v. 139, p. 84 – 96, 2014. ISSN 0925-2312. Citado na página 14.
- MATOS, C. E. F. *Diagnosis of breast cancer in images mammography through local features and invariants*. Tese (Doutorado), 2017. Citado 2 vezes nas páginas 19 e 20.
- NAQI, S. M.; SHARIF, M.; JAFFAR, A. Lung nodule detection and classification based on geometric fit in parametric form and deep learning. *Neural Computing and Applications*, Oct 2018. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-018-3773-x>>. Citado 3 vezes nas páginas 14, 15 e 30.

OLIVEIRA, F. S. S. *Classificação de tecidos da mama em massa e não- massa usando índice de diversidade taxonômico e máquina de vetores de suporte*. Tese (Dissertação de Mestrado. Curso de Pós-Graduação em Engenharia de Eletricidade) — Universidade Federal do Maranhão, 2013. Citado na página 17.

OST, D.; FEIN, A. M.; FEINSILVER, S. H. The solitary pulmonary nodule. *New England Journal of Medicine*, v. 348, n. 25, p. 2535–2542, 2003. PMID: 12815140. Disponível em: <<https://doi.org/10.1056/NEJMcp012290>>. Citado na página 12.

PETCHEY, O. L.; GASTON, K. J. Functional diversity (fd), species richness and community composition. *Ecology Letters*, Blackwell Science Ltd, v. 5, n. 3, p. 402–411, 2002. ISSN 1461-0248. Disponível em: <<http://dx.doi.org/10.1046/j.1461-0248.2002.00339.x>>. Citado na página 23.

Qin, Y. et al. Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3d cnns. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2018. p. 1005–1009. ISSN 2379-190X. Citado 3 vezes nas páginas 14, 15 e 30.

ROCHA, S. V. *Diferenciação do Padrão de malignidade e benignidade de massas em imagens de mamografias usando padrões locais binários, Geoestatística e Índice de diversidade*. Tese (Tese de Doutorado. Curso de Pós- Graduação em Engenharia de Eletricidade) — Universidade Federal do Maranhão, 2014. Citado na página 17.

SCHLEUTER, D. et al. A user's guide to functional diversity indices. *Ecological Monographs*, v. 80, n. 3, p. 469–484, 2010. Disponível em: <[https://esajournals-onlinelibrary.wiley.com/doi/abs/10.1890/08-2225.1](https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/08-2225.1)>. Citado na página 22.

SEERAM, E. Computed tomography: Physical principles and recent technical advances. *Journal of Medical Imaging and Radiation Sciences*, v. 41, n. 2, p. 87 – 109, 2010. ISSN 1939-8654. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1939865410000317>>. Citado na página 17.

Setio, A. A. A. et al. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1160–1169, May 2016. ISSN 0278-0062. Citado 3 vezes nas páginas 14, 15 e 30.

SILVA, T. F. *Diferenciação do padrão de malignidade e benignidade de massas em mamografias utilizando características geométricas e máquina de vetor de suporte*. Tese (Dissertação de Mestrado. Programa de Pós- graduação em Ciência da Computação) — Universidade Federal do Maranhão, 2016. Citado na página 17.

SOUSA, U. S. *Classificação de massas na mama a partir de imagens mamográficas usando índice de diversidade de shannon-wiener*. Tese (Dissertação de Mestrado. Programa de Pós-graduação em Ciência da Computação) — Universidade Federal do Maranhão, 2011. Citado na página 17.

TILMAN, D. Functional diversity. in encyclopedia of biodiversity. In: . [S.l.]: (S.A. Levin, ed.). Academic Press, San Diego, p. 109-120, 2001. Citado na página 22.

Torres, W. et al. Functional diversity applied to the false positive reduction in breast tissues based on digital mammography. In: *2018 IEEE Symposium on Computers and*

---

*Communications (ISCC)*. [S.l.: s.n.], 2018. p. 01120–01125. ISSN 1530-1346. Citado na página 21.

V., B. S.; UNNIKRISHNAN, A.; BALAKRISHNAN, K. Gray level co-occurrence matrices: Generalisation and some new features. *CoRR*, abs/1205.4831, 2012. Disponível em: <<http://arxiv.org/abs/1205.4831>>. Citado na página 23.

WITTEN, I. H. et al. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. 4th. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016. ISBN 0128042915, 9780128042915. Citado na página 24.



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

**Identificação do Tipo de Documento**

- ( ) Tese  
( ) Dissertação  
 Monografia  
( ) Artigo

Eu, William de Oliveira Torres,  
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 Análise de textura de módulos pulmonares para sistemas CAD utilizando Índice de Diversidade Funcional 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 20 de Janeiro de 2020.

William de Oliveira Torres  
Assinatura

\_\_\_\_\_  
Assinatura