

**UNIVERSIDADE FEDERAL DO PIAUÍ
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**REDUÇÃO DE FALSOS POSITIVOS BASEADO NAS RELAÇÕES
FILOGENÉTICAS ENTRE ESPÉCIES**

THAYANE DE OLIVEIRA SIMÕES

PICOS – PI

2017

THAYANE DE OLIVEIRA SIMÕES

**REDUÇÃO DE FALSOS POSITIVOS BASEADO NAS RELAÇÕES
FILOGENÉTICAS ENTRE ESPÉCIES**

Trabalho de Conclusão de Curso submetido à Coordenação do Curso de Bacharelado em Sistemas de Informação da Universidade Federal do Piauí, Campus Senador Helvídio Nunes de Barros, no período 2016.2 como requisito para obtenção do título de Bacharel em Sistemas de Informação.

Orientadora: Profª. Ma. Patrícia Medyna Lauritzen de Lucena Drumond

PICOS – PI

2017

FICHA CATALOGRÁFICA**Serviço de Processamento Técnico da Universidade Federal do Piauí****Biblioteca José Albano de Macêdo****S589r** Simões, Thayane de Oliveira.

Redução de falsos positivos baseado nas relações filogenéticas entre espécies / Thayane de Oliveira Simões.– 2016.

CD-ROM : il.; 4 ¾ pol. (f.)

Trabalho de Conclusão de Curso (Curso Bacharelado em Sistemas de Informação) – Universidade Federal do Piauí, Picos, 2016.

Orientador(A): Prof.^a Ma. Patrícia Medyna Lauritzen de Lucena Drumond.

1. Câncer-Falso Positivo. 2. Câncer de Pulmão-Índice de Diversidade Filogenética. 3. Redução de Falso Positivo-Câncer. I. Título.

CDD 005

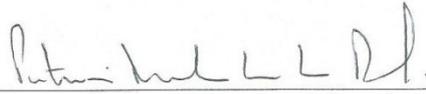
REDUÇÃO DE FALSOS POSITIVOS BASEADO NAS RELAÇÕES FILOGENÉTICAS
ENTRE ESPÉCIES

THAYANE DE OLIVEIRA SIMÕES

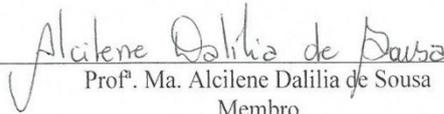
Monografia aprovada como exigência parcial para obtenção do grau de
Bacharel em Sistemas de Informação.

Data de Aprovação

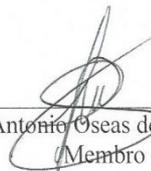
Picos - PI, 18 de Janeiro de 2017



Profª. Ma. Patricia Medyna Lauritzen de Lucena Drumond
Orientador



Profª. Ma. Alcilene Dalilia de Sousa
Membro



Prof. Dr. Antonio Oseas de Carvalho Filho
Membro

RESUMO: O câncer de pulmão tornou-se a patologia com maior incidência de mortalidade dessa categoria. Assim, ferramentas de cunho computacional que utilizam técnicas de processamento de imagens digitais e reconhecimento de padrões, têm sido amplamente explorados. Essa demanda tem como finalidade auxiliar na detecção precoce de nódulos pulmonares, visto que o câncer de pulmão é o mais comum de todos os tumores malignos. Este trabalho apresenta uma técnica de redução de falsos positivos que permite a extração de informações de textura baseado nos índices de diversidade filogenética em imagens de tomografia computadorizada, esses índices são utilizados como características para os classificadores: *MultiLayer Perceptron*, *J48*, *Random Forest* e *Random Tree* identificando a presença de nódulo e não nódulo. Para validação foram testadas 24.127 imagens da base LIDC-IDRI de um total de 833 exames, obtendo um índice kappa de 0,9909, 0,008 de taxa de falsos positivos, curva ROC igual a 100, uma taxa de acerto de 99,6%, sensibilidade de 99,7%, e especificidade de 99,6%. Os resultados obtidos com a utilização dos índices de diversidade filogenética para redução de falsos positivos mostraram-se bastante promissores.

PALAVRAS-CHAVE: Redução de falsos positivos. Câncer de pulmão. Índices de Diversidade Filogenética.

ABSTRACT: Lung cancer became the pathology with a higher incidence of mortality from this category. So computer oriented tools that use techniques of digital image processing and pattern recognition, have been widely exploited. This demand is intended to assist in early detection of lung nodules, since lung cancer is the most common of malignant tumors. This paper presents a technique to reduce false positives which allows the texture information extraction based on phylogenetic diversity indexes on computed tomography images, these indexes are used as features for the classifiers: *MultiLayer Perceptron*, *J48*, *Random Forest* and *Random Tree* by identifying the presence of lump and nodule not. For validation were tested 24.127 bases images LIDC-IDRI of 833 total exams, obtaining a kappa index of 0.9909, 0.008 of false positive rate, ROC curve equal to 100, a hit rate of 99.6%, 99.7% sensitivity and specificity of 99.6%. the results obtained with the use of phylogenetic diversity indexes to reduce false positives were very promising.

KEYWORDS: False Positive Reduction. Lung Cancer. Phylogenetic Diversity Indexes.

1 INTRODUÇÃO

Câncer é caracterizado pelo crescimento desordenado de células que invadem os tecidos e órgãos, essas células tendem a ser muito invasivas e incontroláveis, podendo surgir tumores malignos (câncer) ou benignos. Normalmente as células saudáveis crescem e se dividem para formar novas células. Quando o câncer se desenvolve, no entanto, este processo é interrompido, as células se tornam cada vez mais anormais ou danificadas (NCI, 2016).

O câncer de pulmão é o mais comum de todos os tumores malignos, apresentando aumento de 2% por ano na sua incidência mundial. Suas causas são as mais diversas possíveis, como histórico familiar, fatores genéticos e o mais comum é o consumo de derivados do tabaco, apresentando 90% dos casos, dentre outros (INCA, 2016).

O diagnóstico é feito através de um Raio-X do tórax ou por Tomografia Computadorizada (TC), sendo o segundo o exame mais indicado para a detecção precoce do câncer de pulmão. No entanto, às vezes, torna-se exaustiva a análise desse exame, dificultando a detecção do nódulo, devido à localização, baixa intensidade e tamanho. Por isso, sistemas assistidos por computador, por exemplo *Assisted Detection Systems Computer* (CAD), vêm

sendo desenvolvidos para auxiliar no diagnóstico dessa doença, fornecendo uma segunda opinião ao especialista.

Processamento Digital de Imagens (PDI), é uma área em crescimento e está sendo bastante utilizada para auxiliar na detecção precoce de inúmeras doenças, inclusive o câncer de pulmão. PDI consiste em um conjunto de operações matemáticas aplicadas sobre a imagem que resulta, na maioria dos casos, em uma nova imagem. Essa manipulação das imagens permite através da análise automática do computador, extrair informações até então imperceptíveis ao olho humano.

Dessa forma, este trabalho apresenta uma metodologia de redução de falsos positivos (número de imagens erradamente classificadas como não-nódulo) utilizando análise de textura, que utiliza técnicas de PDI e reconhecimento de padrões, baseados em índices de diversidade filogenética para extração de características e múltiplos classificadores para classificação de regiões em nódulo e não nódulo.

2 TRABALHOS RELACIONADOS

Na literatura existem várias pesquisas no contexto desse trabalho. Nesta seção enumeram-se alguns deles.

Santos *et. al.* (2014), demonstram uma metodologia para a detecção automática de nódulos pulmonares pequenos (com tamanhos entre 2 e 10 mm) por meio de técnicas de processamento de imagens e reconhecimento de padrões. Utilizam medidas de entropia e Shannon Tsallis como descritores de textura e SVM para a classificação de regiões suspeitas tanto como nódulos ou não-nódulos. Os resultados obtidos com o uso deste conjunto de técnicas, aplicado a uma amostra com 28 exames a partir de uma base de dados pública, mostrou que pequenos nódulos foram detectados com uma sensibilidade de 90,6%, uma especificidade de 85% e uma precisão de 88,4%. A taxa de falsos positivos por exame foi de 1,17.

Eskandarian *et. al.* (2015), apresentam um sistema assistido por computador para detecção de nódulos pulmonares em tomografia computadorizada, utilizando o classificador SVM. Assim, no primeiro passo, por meio de técnicas de mineração de dados, volume de dados são reduzidos. Em seguida, dividida pela área do peito, os nódulos suspeitos são identificados e, eventualmente, os nódulos são detectados. Em comparação com os métodos baseados em limites, utilizaram o SVM pois descreve com mais precisão as áreas dos pulmões.

Os resultados experimentais baseados em dados de 147 pacientes do banco de dados de imagem LIDC-IDRI, o sistema proposto obteve a sensibilidade de 89,9% e de falsos positivos de 3,9 por varredura. Em comparação com sistemas anteriores, o sistema proposto demonstrou um bom desempenho.

Carvalho Filho *et. al.* (2016), desenvolveram uma metodologia baseada em técnicas de processamento de imagem e de reconhecimento de padrões, para a classificação de nódulos pulmonares usando a imagem LIDC-IDRI. Utilizaram os índices de diversidade taxonômica e taxonômica distinção da ecologia para descrever a textura do nódulo e não-nódulo. O cálculo dos índices é baseada em árvores filogenético que são aplicados para a caracterização de candidatos e para classificação foi utilizado o SVM. Para aplicar a metodologia, dividiram o banco de dados completo em dois grupos para treinamento e testes. Na fase de testes, foram utilizados 833 exames, obtendo uma precisão média de 98,11%.

Setio *et al.* (2016), sugerem em seu trabalho um novo sistema CAD para nódulos pulmonares, utilizado multi-viem *Convolutional Networks* (ConvNets), sendo que os recursos característicos são automaticamente aprendidos com os dados de treinamentos. A rede é alimentada com os candidatos a nódulo obtidos através da combinação de três detecção de candidatos, especificamente para sólidos, subsólidos e grandes nódulos. Para cada candidato, foi utilizado um conjunto de *patches* 2-D de plano orientados para realizar a extração. A metodologia proposta compreende múltiplos fluxos de ConvNets 2-D, para o qual as saídas são combinadas usando um método de fusão dedicado para conseguir a classificação final. Em 888 imagens do conjunto de dados LIDC IDRI, esse método atingiu alta sensibilidade de detecção de 85,4% e 90,1% em 1 e 4 falsos positivos por digitalização e taxa de acertos de 93,1%.

Diante do que foi exposto nos trabalhos relacionados, alguns problemas são frequentes quando se refere a ruídos na imagem, tamanho e localização do nódulo, e elevadas taxas de falsos positivos por exame, sendo que estes são obstáculos a serem superados. A metodologia proposta tem como contribuição reduzir a quantidade de falsos positivos utilizando os índices de diversidade filogenética.

3 METODOLOGIA PROPOSTA

O método proposto compreende em quatro etapas, sendo: a aquisição das imagens de Tomografia Computadorizada (TC) da base LIDC-IDRI, extração de características utilizando

textura baseado nos índices de diversidade filogenéticos *Phylogenetic Diversity* (PD) (FAITH, 1992), *Mean Nearest Neighbor Distance* (MNND) (WEBB, 2000), *Sum of Phylogenetic Distances* (SPD) (WEBB, 2000), *Phylogenetic Species Variability* (PSV) (HELMUS, 2007) e *Phylogenetic Species Richness* (PSR) (HELMUS, 2007), a classificação utilizando o *Random Tree* (GEURTS *et al.*, 2006), *Random Forest* (BREIMAN, 2001), *Multilayer Perceptron* (HAYKIN *et al.*, 2008) e J48 e a validação dos resultados foram feitos a partir de algumas estatísticas como *kappa*, curva ROC, acurácia, sensibilidade e especificidade. A Figura 1 apresenta a ordem do método utilizado, iniciando com a aquisição da base de imagens e terminando com a validação dos resultados.

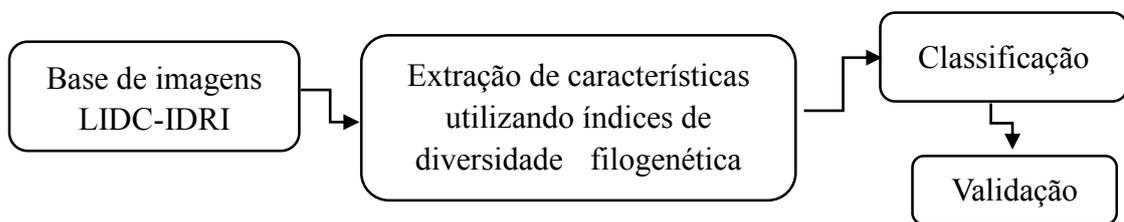


Figura 1. Metodologia proposta para redução de falsos positivos.

3.1 Base de Imagens

A base de imagem utilizada foi a LIDC-IDRI, um repositório público de imagens tomográficas, disponibilizado na internet pelo *National Cancer Institute of USA* (NCI) contendo 1018 casos.

Cada exame da base possui informações e avaliações de 4 especialistas, que são disponibilizadas em um arquivo XML (ARMATTO *et al.*, 2011). Entretanto, foram utilizados apenas 833 exames, pois, alguns problemas como divergências nos arquivos de marcação e informações presentes nos cabeçalhos dos exames tornaram inviáveis seu uso.

3.2 Índices de Diversidade Filogenética

Na maioria dos sistemas CAD a fase de extração de características é baseada na forma e na textura. A textura descreve os aspectos relacionados aos níveis de cinza que possui cada região de interesse, essa técnica possibilita analisar os exames de imagens e extrair informações a partir das variações em valores de *pixels*, que se repetem de maneira regular na imagem, originando um padrão.

A metodologia utiliza os índices de diversidade filogenética PD, MNND, SPD, PSV e PSR para a extração de características nas imagens de TC baseados na análise de textura, sendo

que os cálculos destes índices se baseiam em árvores filogenéticas para a redução de falsos positivos.

A diversidade é um termo extremamente utilizado na área da ecologia, sua finalidade é informar a variedade de espécies presentes em uma comunidade ou área (OLIVEIRA, 2013) e a filogenia é um ramo da biologia responsável pelo estudo das relações evolutivas entre as espécies, pela verificação dos relacionamentos entre elas, a fim de determinar possíveis ancestrais comuns (ARAÚJO, 2003).

A filogenia é representada por árvores filogenéticas, onde essas árvores devem expressar similaridade, ancestralidade ou parentescos evolutivos entre espécies. Nelas as folhas representam os organismos e os nós internos correspondem aos seus ancestrais hipotéticos, como apresenta a Figura 2.

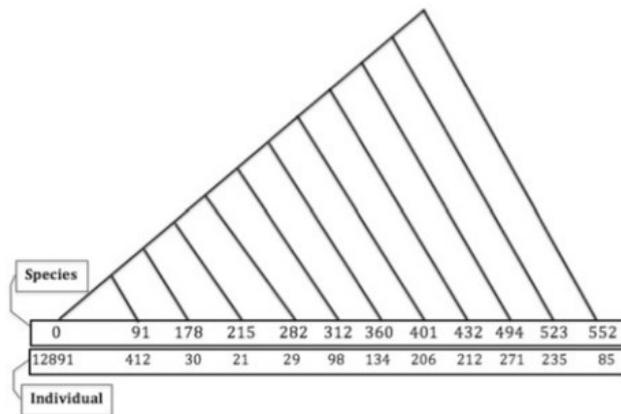


Figura 2. Árvore enraizada em forma de cladograma inclinada.
Fonte: Carvalho Filho, 2016.

As árvores filogenéticas, combinadas com o PD, MNND, SPD, PSV e PSR, são usadas na biologia para comparar os padrões de comportamento das espécies em diferentes áreas. A Tabela 1 mostra a comparação entre os termos de biologia e da metodologia proposta neste trabalho, para compreender como essas árvores filogenéticas são utilizadas em uma imagem de TC.

Tabela 1. Correspondência entre os termos da biologia e o trabalho

Biologia	Metodologia
Comunidade	Regiões de interesse (nódulo da imagem TC)
Espécies	Níveis de Unidade de Hounsfield da região
Indivíduos	Quantidade de <i>voxel</i> de cada Unidade de Hounsfield

3.3 Phylogenetic Diversity

O índice de diversidade filogenética PD, na biologia é obtido pela soma dos comprimentos dos braços da árvore filogenética das espécies de uma comunidade multiplicado pelo número de pares das espécies (FAITH, 1992), onde, B é o número de ramificações da árvore, S é número de espécies do conjunto focal, L_i é o comprimento do ramo i e A_i representa a abundância (indivíduos) média de espécies que compartilham i , conforme apresenta a Equação 1.

$$PD = B * \frac{\sum_i^S L_i A_i}{\sum_i^S A_i} \quad (1)$$

3.4 Mean Nearest Neighbor Distance

O índice MNND é a distância filogenética média do parente mais próximo de todas as espécies e é equivalente às taxas de espécies por gênero (WEBB, 2000). Pode ser calculado a partir da média ponderada da distância filogenética de cada vizinho mais próximo das espécies, com pesos iguais a abundância das espécies, onde S é número de espécies do conjunto focal, m e n são espécies, d_{mn} é a distância filogenética entre m e n e a_m é a abundância das espécies de m , conforme a Equação 2.

$$MNND = \sum_i^S \min(d_{mn}) a_m \quad (2)$$

3.5 Sum of Phylogenetic Distances

O índice de diversidade filogenética SPD é a soma das distâncias filogenéticas entre cada par de espécies multiplicado pelo número de pares das espécies, onde, S é número de espécies do conjunto focal, m e n são espécies, d_{mn} é a distância filogenética entre m e n , a_m é a abundância das espécies de m e a_n é a abundância das espécies de n , como apresenta a Equação 3.

$$SPD = \left(\frac{S(S-1)}{2} \right) * \frac{\sum \sum_{m < n} d_{mn} a_m a_n}{\sum \sum_{m < n} a_m a_n} \quad (3)$$

3.6 Phylogenetic Species Variability

Segundo Helmus (2007), a variabilidade quantifica o parentesco filogenético, diminuindo a variação das características compartilhadas por todas as espécies da comunidade. A Equação 4 demonstra a variabilidade de espécies filogenéticas, que resume o grau em que as espécies em

uma comunidade são filogeneticamente relacionadas. Na fórmula do PSV, o trC representa a soma dos valores da diagonal de uma matriz C , \sum é o somatório de todos os valores da matriz, n é o número de espécies e \bar{c} é a média dos elementos da diagonal de C .

$$PSV = \frac{trc - \sum c}{n(n-1)} = 1 - \bar{c} \quad (4)$$

3.7 Phylogenetic Species Richness

A riqueza de espécies quantifica o número de espécies em uma comunidade é apresentada nos parâmetros na Equação 5, onde o valor do PSR é a multiplicação do número espécies n pela variabilidade da comunidade (HELMUS, 2007).

$$PSR = nPSV \quad (5)$$

3.8 Classificação e Validação

A classificação é um processo que visa o reconhecimento automático de objetos, sendo que as entradas são as características extraídas das imagens TC do pulmão. Os classificadores testados, utilizaram o método *k-fold cross validation* para obter os resultados. Os dados foram divididos em 10 conjuntos, sendo 9 deles para treinamento e 1 para testes. Este processo é repetido 10 vezes e no final é gerado um resultado.

A ferramenta utilizada para classificação foi a *Waikato Environment for Knowledge Analysis* (WEKA) que é um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Contém ferramentas para pré-processamento de dados, classificação, regressão, *clustering*, regras de associação, e visualização (WEKA, 2016).

O desempenho obtido pelos algoritmos foi avaliado utilizando as seguintes métricas: *Kappa* (K) (PERROCA; GAIDZINSKI, 2003), área sob a curva *Receiver-Operating Characteristics* (ROC) (METZ, 1986), Sensibilidade (S) (MARTINEZ *et. al.*, 2003), Especificidade (E) (MARTINEZ *et. al.*, 2003) e Acurácia (A).

A sensibilidade é a proporção de imagens patológicas, ou seja, demonstra o quanto a metodologia foi boa para detectar a doença, a especificidade representa a proporção de imagens saudáveis, o quanto o método foi eficiente em detectar a ausência da doença, o *kappa* é uma medida de concordância entre observadores, a acurácia simula a taxa de acertos e a curva ROC descreve a habilidade de separar corretamente o conjunto de nódulos em duas classes baseado na fração verdadeiro positivo (especificidade) e na fração falso positivo (sensibilidade), quanto mais próximo de 1 significa que a classificação foi boa.

4 RESULTADOS

Os testes realizados neste trabalho foram efetuados com uma amostra de 24.127 imagens de TC da base LIDC-IDRI de um total de 833 exames, sendo 17.728 não-nódulos e 6.399 nódulos. As características foram extraídas a partir dos índices de diversidade filogenéticas PD, MNND, SPD, PSR e PSV. Os classificadores *MultiLayer Perceptron* (MLP), *Random Forest* (RF), *Random Tree* (RT) e J48 foram usados para classificação de nódulo e não nódulo em imagens de TC, através do *software* WEKA.

Tabela 2. Resultados da classificação usando os cinco índices.

Classificador	FP/i	K	ROC	S (%)	E (%)	A (%)
MLP	0,401	0,437	0,86	66,9	83,1	80
RF	0,008	0,9909	1	99,7	99,6	99,6
RT	0,013	0,9816	0,99	99	99,4	99,3
J48	0,016	0,9812	0,995	99,3	99,3	99,3

A partir da Tabela 2 pode-se afirmar que os classificadores obtiveram bons resultados, tendo como evidência o RF que alcançou o melhor resultado, pois destacou-se no reconhecimento da doença com sensibilidade igual a 99,7%, reconhecimento da ausência da doença com especificidade igual a 99,6%, *kappa* de 0,9909, curva ROC igual a 100, taxa de falsos positivos (FP/i) igual 0,008 e acurácia de 99,6%. Enquanto que o MLP obteve o pior resultado, apresentando sensibilidade de 66,9%, especificidade de 83,1%, taxa de falsos positivos igual a 0,401 e acurácia de 80%.

Tabela 3. Comparação da metodologia com os trabalhos relacionados

Trabalhos	FP/i	Base	S (%)	E (%)	A (%)
Santos (2014)	1,17	LIDC-IDRI	90,6	85	88,4
Eskandarian (2015)	3,9	LIDC-IDRI	89,9	-	97
Setio (2016)	1-4	LIDC-IDRI	85,4	90,1	93,1
Carvalho Filho (2016)	-	LIDC-IDRI	98	98,92	99,2
Metodologia	0,008	LIDC-IDRI	99,7	99,6	99,6

De acordo com Tabela 3, a metodologia proposta utilizando os índices de diversidade filogenética em comparação com os trabalhos relacionados apresenta o melhor resultado, pois há um equilíbrio entre as três métricas (sensibilidade, especificidade e acurácia). Assim, os resultados mostraram-se promissores em relação aos trabalhos recentes na literatura relacionados à redução de falsos positivos em regiões de nódulo e não nódulo pulmonar.

5 CONCLUSÃO E TRABALHOS FUTUROS

O câncer de pulmão é uma doença silenciosa e normalmente é detectado nos estágios mais avançados. A detecção precoce aumenta as taxas de sobrevivência dos pacientes, por isso sistemas auxiliados por computador é um assunto relevante e pode fornecer uma segunda opinião aos especialistas.

Este trabalho apresentou uma metodologia de redução de falsos positivo em imagens de TC, utilizando os índices de diversidade filogenética PD, MNND, SPD, PSV e PSR. A partir dos resultados encontrados, pode-se afirmar que a utilização dos índices baseados na textura para a redução de falsos positivo são promissores. Os índices apresentados nos resultados da análise demonstraram um bom desempenho para a extração de características de regiões em nódulos ou não nódulos pulmonares.

Para trabalhos futuros pretende-se utilizar outras bases de imagens e outros índices de diversidade filogenética, utilizando análise de textura para otimizar os resultados e constatar a eficácia da metodologia.

7 PUBLICAÇÕES

Trabalho Original

SIMOES, T. O.; SANTOS, O. S.; CARVALHO FILHO, A. O.; SOUSA, A. D. **Redução de falsos positivos baseado nas relações filogenéticas entre espécies.** In: Congresso Brasileiro de Informática em Saúde CBIS, 2016, Goiânia, Goiás, 2016.

Outras Publicações

SIMOES, T. O.; SANTOS, O. S.; CARVALHO FILHO, A. O.; DRUMOND, P. M. L. L.; SOUSA, A. D. **Redução de falsos positivos em imagens de tomografia computadorizada, usando índice de diversidade filogenético e SVM.** In: Escola Regional de Informática do Piauí, 2016, Teresina. Anais da I Escola Regional de Informática do Piauí, 2016.

SANTOS, O. S.; SIMOES, T. O.; SOUSA, A. D.; CARVALHO FILHO, A. O.; DRUMOND, P. M. L. L. **Diferenciação dos padrões de malignidade e benignidade em imagens de tomografia computadorizada usando índice de diversidade filogenético e SVM.** In: Escola Regional de Informática do Piauí, 2016, Teresina. Anais da I Escola Regional de Informática do Piauí, 2016.

SANTOS, O. S.; SIMOES, T. O.; NASCIMENTO, L. M.; SOUSA, A. D.; CARVALHO FILHO, A. O. **Análise filogenética para diferenciação entre nódulos malignos e benignos.** In: Congresso Brasileiro de Informática em Saúde CBIS, 2016, Goiânia, Goiás, 2016.

REFERÊNCIAS

- ARAÚJO, G. S. **Filogenia de proteomas**. 2003, 73 f. Dissertação (Ciência da Computação) - Universidade Federal do Mato Grosso do Sul, MS, 2003.
- ARMATO III S.G.; MCLENNAN G.; BIDAUT L.; MCNITT-GRAY M.F.; MEYER C.R., REEVES A.P.; ZHAO B.; ABERLE D.R.; HENSCHKE C.I.; HOFFMAN E.A., *et al.* **The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans**, *Med. Phys.*, 38 (2011), pp. 915–931.
- BREIMAN, L. **Random forests**. *Machine learning*, v. 45, n. 1, p. 5-32, 2001.
- CARVALHO FILHO, A. O.; SILVA, A. C.; PAIVA, A. C.; GATTAS, R. A. N. M. **Lung-nodule classification based on computed tomography using taxonomic diversity indexes and an SVM**. *Journal of Signal Processing Systems for Signal, Image, and Video Technology*, vol. 83, may 2016.
- ESKANDARIAN P.; BAGHERZADEH J. **Computer-aided detection of Pulmonary Nodules based on SVM in thoracic CT images**, *IEEE Transactions on Medical Imaging*, p. 1-6, 2015.
- FAITH, D.P. **Conservation evaluation and phylogenetic diversity**. *Biol. Conserv.* 61(1):1-10, 1992.
- GEURTS, P.; ERNST, D.; WEHENKEL, L. **Extremely randomized trees**, *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.
- HAYKIN, S. **Neural Networks and Learning Machines**. 3^a ed. New Jersey: Prentice Hall, 936 p. 2008.
- HELMUS, M. R.; BLAND, T. J.; WILLIAMS, C. K.; IVES, A. R. **Phylogenetic measures of biodiversity**. *The American Naturalist*, v. 169, n. 3, p. E68-E83, 2007.
- Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). **Mistério da Saúde, O que é o câncer**. 2016. Disponível em: <www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pulmao>. Acesso em: 15 de fevereiro 2016.
- LANDIS, J. R.; KOCH, G. G. **The measurement of observer agreement for categorical data**. *Biometrics*, p. 159-174, 1977.
- MAGURRAN, A. E. **Measuring biological diversity**. John Wiley & Sons, 2013.
- MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, B. B. **A curva ROC para testes diagnósticos**. *Cad Saúde Coletiva*, v. 11, n. 1, p. 7-31, 2003.
- METZ, C. E. **ROC methodology in radiologic imaging**. *Investigative radiology*, v. 21, n. 9, p. 720-733, 1986.

National Cancer Institute - NCI. **Lung Cancer. U.S. National Institute of Health.** Disponível em: <[www.cancer.gov/about-cancer/what-is-cancer- What Is Cancer](http://www.cancer.gov/about-cancer/what-is-cancer-What-Is-Cancer)>. Acesso em: 17 fev. 2016.

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.** 2013.70f. Dissertação (Mestrado na área de Ciência da Computação) - Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2013.

PERROCA M. G; GAIDZINSKI R. R. **Avaliando a confiabilidade inteavaliadores de um instrumento para classificação de pacientes – coeficiente Kappa.** Ver. Esc. Enferm. 37 (1): 72-80., 2003.

SANTOS, A. M.; CARVALHO FILHO, A. O.; SILVA, A. C.; PAIVA, A. C.; NUNES, R. A.; GATTASS, M. **Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM,** Science Direct, vol. 36, p. 27–39, 2014.

SETIO, A.A.A; CIOMPI, F; LITJENS, P.; JACOBS, C.; RIEL, S.; WILLE, M.W.; SANCHEZ, C.; GINNEKEN. **Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks,** IEEE Transactions on Medical Imaging, p. 1, 2016.

WEBB, C. O. **Exploring the phylogenetic structure of ecological communities: an example for rain forest trees.** The American Naturalist, v. 156, n. 2, p. 145-155, 2000.

WEKA – **Machine Learning Group at the University of Waikato.** Disponível em:<<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em 15 fev. 2016.



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

Identificação do Tipo de Documento

- () Tese
() Dissertação
() Monografia
 Artigo

Eu, Thayane de Oliveira Simões
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 Redução de Falsos Positivos Baseado nas Relações Filogenéticas entre Espécies 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 março de 2017.

Thayane de O. Simões
Assinatura
Thayane de O. Simões
Assinatura