

Redes neurais convolucionais para a classificação de nódulos tireoidianos através de ultrassonografia

Igor Machado Seixas¹, Alexei Manso Correa Machado^{1,2}

¹Departamento de Ciência da Computação - PUC Minas

² Departamento de Anatomia e Imagem - Universidade Federal de Minas Gerais

igormseixas@hotmail.com, alexeimcmachado@gmail.com

Abstract. *Premature detection of malignant nodes in the thyroid is critical for effective treatment. In this study a computer-aided diagnosis system is proposed to classify malign and benign nodes of the thyroid based on ultrasound images, as well as in the scale of the Thyroid Imaging Reporting and Data System (TI-RADS). The experiments implement 5 convolutional network and 3 support vector machines applied to a public dataset. Preliminary results indicate the MobileNet as the best binary classifier with 89% of accuracy and the DenseNet121 with 56% of accuracy for the 4 TI-RADS categories.*

Resumo. *A detecção precoce de linfonodos malignos é crítica para o tratamento do câncer de tireoide. Neste estudo, um sistema de diagnóstico é proposto para classificar nódulos malignos com base em imagens de ultrassom, bem como na escala do Thyroid Imaging Reporting and Data System (TI-RADS). Os experimentos implementam 5 redes convolucionais e 3 máquinas de vetores de suporte aplicadas a um conjunto de dados público. Os resultados preliminares indicam o MobileNet como o melhor classificador binário com 89% de acurácia e o DenseNet121 com 56% de acurácia para as 4 categorias TI-RADS.*

1. Introdução

Disfunções da tireoide são as patologias mais frequentes na endocrinologia e o câncer na glândula tireoidiana é o mais prevalente. Embora a maioria dos nódulos encontrados sejam benignos, 10% deles podem ser malignos (Guth et al., 2009). A classificação precoce desses nódulos pode reduzir o risco de morte do paciente além de reduzir de forma significativa os custos na realização de cirurgias ou biopsias.

A ultrassonografia da região do pescoço é o método menos invasivo utilizado por médicos na identificação de nódulos tireoidianos. Eles classificam o nódulo de acordo com a escala *Thyroid Imaging Reporting and Data System* (TI-RADS) que varia de 1 a 5 de acordo com sua suspeição de malignidade (Kwak et al., 2011). A detecção, no entanto, não é simples, demanda tempo e algumas vezes gera dúvidas entre médicos, uma vez que a correta classificação está sujeita à experiência profissional ou qualidade do equipamento (Lingam et al., 2013). Dessa forma, um sistema de diagnóstico computadorizado ajudaria a aumentar a confiabilidade, acurácia, reprodutividade e eficiência dos diagnósticos.

As imagens obtidas através do ultrassom apresentam diversos problemas relativos à resolução, variação de tons de cinza, contraste, ruídos e anotações presentes nas

imagens disponibilizadas pelos aparelhos ou realizada pelos médicos, sendo consideradas umas das mais difíceis modalidades para interpretação. Não obstante, devido à sua importância e baixo custo, a análise de imagens de ultrassom motivou diversos trabalhos na literatura. Um estudo proposto por Ding et al. (2011) selecionou subconjuntos, características e estatísticas de textura que foram extraídos de exames de imagem da tireoide com a intenção de classificar os nódulos em malignos ou benignos utilizando máquinas de vetores de suporte (SVM). Um sistema computadorizado automático foi proposto por Gopinath and Shanthi (2013), analisando imagens microscópicas obtidas por citologia aspirativa por agulha fina com o auxílio de uma SVM como classificador para identificar nódulos cancerígenos. Uma classificação utilizando redes convolucionais (CNN) foi proposta por Spanhol et al. (2016) para detecção de câncer de pulmão utilizando imagens. Outro modelo baseado em CNN foi proposto por Li et al. (2016) que utilizava esse tipo de rede para identificação de nódulos pulmonares. Uma CNN utilizando ajuste fino na Rede GoogLeNet, técnicas de pré-processamento e aumento de dados foi proposta por Chi and Walia (2017) para análise da malignidade dos nódulos da tireoide. Um modelo de CNN misto foi proposto por Park et al. (2019) que comparou o desempenho da rede proposta com radiologistas iniciantes e experientes na classificação dos nódulos. A rede ResNet-50 foi a escolha de Moussa et al. (2020) para um novo classificador com técnicas de ajuste fino que superou a VGG-19 na mesma tarefa. Uma pontuação de malignidade dos nódulos da tireoide foi apresentada por Stib and Pan (2020), utilizando uma CNN para classificação na escala de TI-RADS. Mais recentemente, Hang (2021) propuseram uma Rede Generativa Adversária Residual (Res-GAN) que separa as imagens entre prováveis benignas e prováveis malignas e depois realiza a comparação com os resultados obtidos pela rede convolucional residual ResNet18.

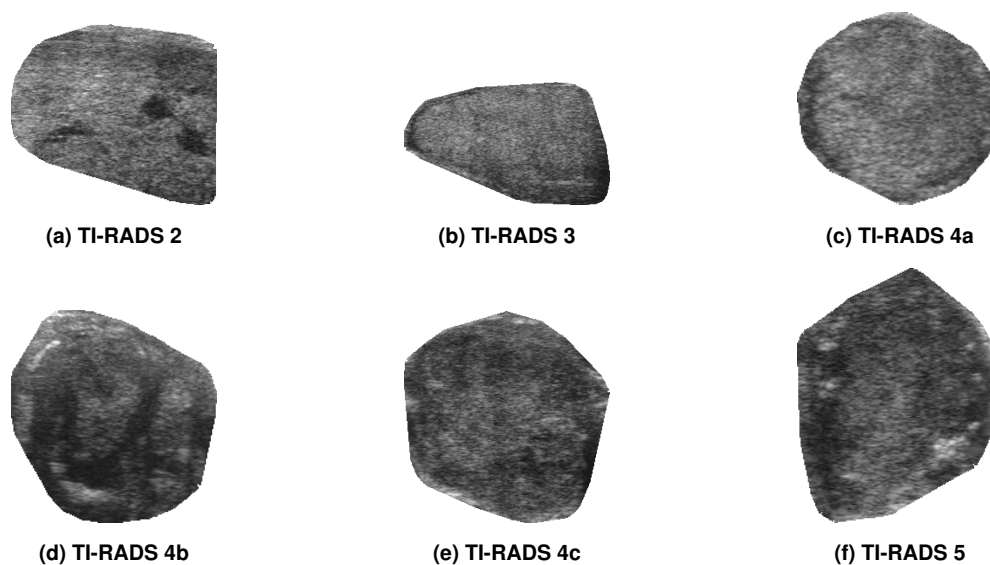
Os trabalhos encontrados propõem uma série de técnicas para segmentação, pré-processamento e classificação entre nódulos benignos e malignos. Entretanto, observa-se um menor número de métodos eficazes na classificação da escala TI-RADS o que ajudaria no estadiamento da lesões. O objetivo deste trabalho é propor um classificador eficaz no diagnóstico de nódulos tireoidianos malignos e benignos, mas também capaz de realizar a estratificação na escala TI-RADS. São avaliadas 5 arquiteturas de redes convolucionais, juntamente com técnicas de segmentação, pré-processamento e aumento de dados. Os resultados preliminares obtidos pelas redes são comparados com soluções baseadas em máquinas de vetores de suporte.

2. Materiais e Métodos

A base de dados utilizada é disponibilizada publicamente pela Universidade Nacional de Colômbia (Pedraza, 2015). Ela possui 347 imagens classificadas, com resolução de 560x360 pixels, 3 canais e 8 bits por canal. Dentre as imagens, 61 são benignas (TI-RADS 2 e 3) e 286 imagens malignas (TI-RADS 4a, 4b, 4c e 5). A base também possui um arquivo XML contendo informações dos pacientes e os pontos que compõem o polígono circunscritor da região do nódulo na tireoide. Os conjuntos de treino com 70% das imagens e o de teste com 30% foram escolhidos de forma aleatória para cada classe. As regiões contendo os nódulos foram recortadas de cada imagem através da descrição contida no arquivo XML. A Figura 1 mostra exemplos de cada escala TI-RADS.

Três sessões de treinamento foram testadas: (a) com imagens de resolução 360x360 contendo a área recortada sobre fundo verde; (b) contendo o retângulo circuns-

Figura 1. Exemplo das imagens segmentadas em função do nódulo tireoidiano.



critor mínimo do recorte; e (c) contendo quadrados com dimensões de 160, 256 e 272 pixels obtidos da área recortada. As CNNs foram testadas usando todas as possibilidades de conjuntos de dados. Os melhores resultados foram encontrados em imagens de 360x360 pixels segmentadas e com fundo verde. Técnicas para aumento de dados consistiram em rotações sucessivas de 45° e espelhamento. Foram implementadas 5 das CNNs mais presentes na maioria dos trabalhos da literatura: MobileNet Large, MobileNet Small, DenseNet121, GogLeNet e AlexNet. Para o treinamento das SVMs foram utilizados descritores de texturas de Haralick e momentos invariantes de Hu. A transformação log-modular proposta por John and Draper (1980) foi aplicada para se evitarem grandes variações de magnitude entre os descritores.

3. Resultados preliminares

O treinamento foi realizado com cada uma das CNNs. Dentre as imagens selecionadas para treinamento foi efetuada outra subdivisão entre, treinamento (representado por 75% do total de treino) e validação (representado por 25% do total de treino). Uma característica do dataset é seu desbalanceamento, uma vez que imagens categorizadas como TI-RADS 2 e 3 estão presentes em menor quantidade (25% da base de dados). Todas as redes foram treinadas a partir de pesos randomizados ou com técnicas de aprendizado por transferência. O uso de redes pré-treinadas apresentou resultados melhores que o uso de pesos randômicos em 95% dos casos.

Foram escolhidos dois critérios de parada para se evitar *overfitting*: o número de épocas para o treinamento que foi configurado para no máximo 3000 e o valor da função de perda na validação ser menor ou igual a 0,001. No entanto a função de perda foi o critério que prevaleceu em todos os casos. Adicionalmente foi ajustada a taxa de aprendizado de 0,01 para 0,001 objetivando a suavização no gradiente descendente estocástico de aprendizado para que o peso dos modelos pré-treinados não sofressem uma distorção acelerada. As redes com melhores acurácias foram salvas ao longo do treinamento de acordo com cada um dos três tipos de classificação: binária (2+3 contra 4+5), TI-RADS

de categoria 4 (4a, 4b e 4c) e TI-RADS individuais (2 a 5).

Inicialmente foram treinadas as CNNs binárias para classificação entre nodos benignos e malignos. A MobileNet Large apresentou os melhores resultados com 89% de acurácia, 89% de sensibilidade, 97% de especificidade e F1-score de 87%. Em seguida foram testadas redes para a classificação dos TI-RADS de categoria 4 (4a, 4b e 4c) com o objetivo de verificar se os modelos eram capazes de discriminar nódulos mais semelhantes e que geram maior dúvida no diagnóstico. O melhor resultado obtido foi de 56% de acurácia utilizando-se a DenseNet121.

O terceiro conjunto de testes avaliou as redes na classificação de todas as categorias TI-RADS individualmente. O baixo número de exemplos, principalmente na categoria 3, causou menor eficácia de resultados sendo que a MobileNet Small foi o melhor modelo, com 41% de acurácia.

Os resultados da Tabela 1 permitem observar que as redes MobileNet e DenseNet121 apresentaram melhores desempenhos em todas as classificações quando comparadas com as demais. Por se tratar de uma rede mais leve, a MobileNet poderá ser usada como modelo inicial no prosseguimento da pesquisa, quando será projetado um modelo original para o problema da classificação de TI-RADS utilizando imagens de ultrassom. Para a mesma base de dados foi verificado que a SVM obteve um desempenho inferior às melhores CNNs em todos os casos.

Tabela 1. Melhores resultados de acurácia (A), sensibilidade (S), precisão (P) e F1-scores obtidos em cada problema de classificação dentre as CNNs propostas, utilizando imagens com fundo verde, 360x360, segmentadas, rotacionadas, espelhadas com taxa de aprendizado 0,001 para Binária, Categoria 4 e taxa de aprendizado 0,0001 para Individual.

Class.	CNN	A	P	S	F1
Binária	MobileNet Large	89%	89%	97%	87%
Binária	MobileNet Large	88%	86%	92%	82%
Binária	MobileNet Small	87%	86%	91%	82%
Binária	DenseNet121	86%	85%	90%	82%
Categoria 4	DenseNet121	56%	56%	56%	55%
Categoria 4	MobileNet Small	56%	54%	54%	54%
Categoria 4	MobileNet Small	55%	53%	53%	53%
Categoria 4	GoogLeNet	48%	47%	47%	46%
Individual	MobileNet Small	41%	39%	38%	38%
Individual	AlexNet	38%	35%	35%	34%
Individual	DenseNet121	37%	34%	33%	33%
Individual	AlexNet	37%	32%	30%	30%

Por último foram implementadas SVMs com objetivo de se compararem os resultados obtidos com os métodos de aprendizado profundo. Para os problemas de classificação com mais de 2 classes, foram escolhidas três perspectivas de solução: um contra um, um contra todos e cascata. Os melhores valores de acurácia dentre todos os

testes foram de 84% na classificação binária, 50% na classificação entre os TI-RADS de categoria 4 e 32% na classificação individual dos TI-RADS, todos a partir da estratégia *um contra um*, como pode ser visto na Tabela 2.

Tabela 2. Melhores resultados de acurácia (A), sensibilidade (S), precisão (P) e F1-scores obtidos em cada problema de classificação dentre as SVMs propostas.

Class.	SVM	A	S	P	F1
Binária	Um contra um	84%	50%	42%	46%
Categoria 4	Um contra um	50%	49%	51%	49%
Individual	Um contra um	32%	28%	27%	27%

3.1. Discussão

O baixo número de amostras em algumas categorias provou ser um desafio para as redes na classificação de nódulos tireoidianos entre malignos, benignos, TI-RADS categoria 4 e individual. É possível que em uma base com maior quantidade de dados, com pré-processamento mais específico e com ajuste nas redes os resultados possam ser significativamente melhorados. Foi verificado que a classificação de nódulos já segmentados apresentou melhor resultado que com imagens contendo informações de fundo, portanto as redes parecem não serem afetadas pelo fundo verde. Técnicas de aumento de dados poderiam ser investigadas também com objetivo de evitar a necessidade de segmentação prévia da região de interesse.

Finalmente embora a classificação binária tenha obtido resultado próximo a 90%, buscar um modelo que melhore a classificação de TI-RADS de categoria 4 possivelmente aumentará os resultados na classificação individual. Cabe ressaltar que é nessa categoria que a presença de tumores malignos e sua classificação acarretam as maiores dúvidas no diagnóstico.

4. Conclusão

Neste artigo foi avaliado o resultado de diversas redes neurais convolucionais na classificação de nódulos tireoidianos através de imagens de ultrassom. As imagens foram selecionadas, distribuídas e segmentadas entre os seis tipos de TI-RADS disponíveis a partir das notações presentes nos arquivos XML. Foram aplicadas técnicas de aumento de dados, como rotação e espelhamento e de pré-processamento, pela extração de descritores de textura. Cinco tipos diferentes de CNN em três tipos de classificação foram utilizados no treinamento. Para a classificação binária entre nódulos malignos e benignos, o melhor resultado obtido foi de 89% de acurácia. Para a classificação TI-RADS de Categoria 4 que avaliou os TI-RADS 4a, 4b e 4c, o melhor resultado foi de 56% de acurácia e para a classificação individual, contemplando os seis tipos disponíveis de TI-RADS, o melhor resultado foi de 41%. A etapa seguinte deste trabalho prevê testar as melhores redes na escala de TI-RADS da *American College of Radiology*. Essa escala possui melhor definição na diferenciação dos nódulos além de reduzir o número de possibilidades em uma categoria. Pretende-se implementar novas redes e propor uma nova arquitetura inspirada na MobileNet para uso no problema específico do diagnóstico de câncer de tireoide através de imagens de ultrassom.

Referências

- J. Chi and E. Walia. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging*, 30(4):477–486, 2017.
- J. Ding, H. Cheng, C. Ning, J. Huang, and Y. Zhang. Quantitative measurement for thyroid cancer characterization based on elastography. *J Ultrasound Med*, 30(9):1259–66, 2011.
- B. Gopinath and N. Shanthi. Support vector machine based diagnostic system for thyroid cancer using statistical texture features. *Asian Pac J Cancer Prev*, 14(1):97–102, 2013.
- S. Guth, U. Theune, J. Aberle, A. Galach, and C. Bamberger. Very high prevalence of thyroid nodules detected by high frequency (13 mhz) ultrasound examination. *Eur J Clin Invest*, 39(8):699–706, 2009.
- Y. Hang. Thyroid nodule classification in ultrasound images by fusion of conventional features and res-gan deep features. *Journal of Healthcare Engineering*, 22:9917538, 2021.
- J. A. John and N. R. Draper. An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):190–197, 1980.
- J. Kwak, K. Han, and J. Yoon. Thyroid imaging reporting and data system for us features of nodules: a step in establishing better stratification of cancer risk. *Radiology*, 260(3):892–9, 2011.
- W. Li, P. Cao, D. Zhao, and J. Wang. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput Math Methods Med*, 216, 2016.
- R. Lingam, M. Qarib, and N. Tolley. Evaluating thyroid nodules: predicting and selecting malignant nodules for fine-needle aspiration (fna) cytology. *Insights Imaging*, 4:617–624, 2013.
- O. Moussa, H. Khachnaoui, R. Guetari, and N. Khlifa. Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network. *Int J Imaging Syst Technol*, 30(1):185–195, 2020.
- V. Park, K. Han, and Y. Seong. Diagnosis of thyroid nodules: Performance of a deep learning convolutional neural network model vs. radiologists. *Scientific Reports*, 9, 2019.
- L. Pedraza. An open access thyroid ultrasound-image database. *Proceedings of the SPIE*, 9287, 2015.
- F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567, 2016.
- M. Stib and I. Pan. Thyroid nodule malignancy risk stratification using a convolutional neural network. *Ultrasound Q.*, 36(2), 2020.