Metodologia para localização de nódulos malignos da tireoide a partir de imagens infravermelhas

Maira Beatriz Hernandez Moran¹, Aura Conci¹

¹ Instituto de Computação
Universidade Federal Fluminense (UFF) – Niterói, RJ – Brazil

{mhernandez, aconci}@ic.uff.br

Abstract. *Thyroid anomalies have high prevalence and their early identification is crucial for a more effective treatment. Thermograms can be used in this process, since nodules tend to be more vascularized, resulting in a temperature increase. This work presents a methodology for determining thyroid nodules. We evaluate parameters that would allow to segment possibly nodular regions in thermographs. Convolutional neural networks (CNN) are used to classify these regions, identifying which ones refer to nodules. The good results of CNN in the classification (with 96% accuracy), show that the viability of the proposed methodology depends on the success of the segmentation.*

Resumo. *Anomalias na tireoide têm alta prevalência e sua identificação precoce é essencial para um tratamento mais eficaz. Termogramas podem ser usados nesse processo, já que nódulos tendem a ser mais vascularizados, resultando num aumento de temperatura na região. Este trabalho apresenta uma metodologia para localizar nódulos na tireoide. São estudados parâmetros que permitiriam segmentar regiões possivelmente nodulares em termografias. Redes neurais convolucionais (CNN) são usadas para classificação dessas regiões, identificando quais se referem a nódulos. Os resultados obtidos apontam que, devido ao bom resultado da CNN na classificação (com acurácia de 96%), a viabilidade da metodologia proposta depende do sucesso da segmentação.*

1. Caracterização do problema

Os nódulos da tireoide podem ser considerados problemas comuns na área clínica. São a principal manifestação de várias doenças da tireoide, cujo comportamento funcional é essencial nessas doenças e representa uma base para o diagnóstico delas. O diagnóstico precoce permite um tratamento eficaz. A medida que crescem, os nódulos dessa região podem ser encontrados através de palpação, devido a sua superficialidade.

Os exames de imagem são uma ferramenta importante no processo de diagnóstico, já que permitem avaliar precisamente as características dos nódulos. No entanto, a tomografia computadorizada (TC) pode expor o paciente a radiação ionizante, o que é um potencial risco para a saúde [Brenner and Hall 2007] e a ressonância magnética (RM) é cara, restringindo o acesso da população em geral [Goldbach-Mansky et al. 2003]. Neste contexto, as termografias são uma alternativa interessante no diagnóstico de doenças da tireoide, uma vez que tem um custo extremamente baixo quando comparado aos outros exames de imagem, além de não oferecerem nenhum efeito colateral [Brioschi et al. 2010]. Estudos indicam que elas podem ser utilizadas como exame complementar para a

detecção de anormalidades em várias regiões do corpo, já que lesões geralmente apresentam aumento no fluxo sanguíneo, o que conseqüentemente pode alterar a temperatura cutânea na região, especialmente nas lesões mais superficiais [Brioschi et al. 2010].

Apesar do uso de termografias para identificação de nódulos tireoidianos ter sido proposta em trabalhos anteriores, nota-se uma escassez de propostas metodológicas práticas, predominando na literatura estudos de caso e observações qualitativas, baseadas inclusive em características visuais das imagens. Isso demonstra o potencial do uso das termografias, mas não especifica um meio cientificamente válido de usá-las.

1.1. Estado da Arte

Dentre as propostas que apresentam abordagem quantitativa, algumas se destacam. Aqui apresentamos algumas delas, em ordem cronológica. Em [Alves et al. 1988] estudam-se termografias de 110 pacientes, 98 com nódulos únicos e 12 com nódulos múltiplos, seleção de nódulos tireoidianos com indicação cirúrgica. Foram calculados: a máxima diferença de isotermas, o índice térmico e o perfil térmico. O valor 0,9C foi usado como limite para definir a natureza da lesão. Usando esse critério, as medidas apontaram uma acurácia média de 92%. Em [Brioschi et al. 2010], os autores realizam um estudo de possíveis parâmetros a serem usados na análise de anomalias da tireoide. O estudo apresenta uma proposta de como as termografias podem ser incluídas como exame complementares no processo de diagnóstico de doenças da tireoide. Os parâmetros propostos são usados neste trabalho. Em [Mahajan and Madhe 2014] média e desvio padrão dos valores da região da lesão de termografias de 14 pacientes, 5 saudáveis e 9 com tireoides hiperfuncionantes, são calculados. Os autores propuseram valores limites para essas características que definiriam se a tireoide apresenta alguma disfunção.

No trabalho apresentado em [Vaz 2014], termografias são usadas para identificar hipotireoidismo e hipertireoidismo. Uma classificação das termografias, considerando características computadas, é realizada usando um classificador Multilayer Perceptron (MLP). Os experimentos usando validação cruzada 3-fold e 10-fold levaram a sensibilidade de 93,12% e 89,62%, e especificidade de 91,80% e 89,70%.

1.2. Motivação

A proposta deste trabalho é verificar a viabilidade de usar termografias como um exame complementar a ser utilizado na rotina de avaliação dos pacientes, de forma a identificar mais precocemente possíveis nódulos na região da tireoide. Assim, a demanda de exames menos acessíveis pode ser diminuída, minimizando a exposição dos pacientes aos efeitos colaterais deles. A termografia poderia ser uma opção interessante para pacientes em que os exames tradicionais não são indicados, como crianças pequenas e mulheres grávidas.

1.3. Objetivos

Este trabalho tem como objetivo principal avaliação da viabilidade da utilização de parâmetros baseadas em termografias para a identificação de nódulos tireoidianos. Como objetivos secundários surgem: o desenvolvimento de uma metodologia para a detecção de nódulos, utilizando técnicas de segmentação e algoritmos de redes neurais convolucionais em imagens infravermelhas; estudo de características presentes nos termografias que possibilitem a identificação de regiões nodulares; aplicação de técnicas de processamento de imagens para segmentação de regiões possivelmente nodulares; e reconhecimento de nódulos dentre regiões extraídas das termografias.

2. Contribuições do trabalho

O estudo realizado neste trabalho permitiu avaliar a viabilidade do uso de diversas características previamente propostas na literatura médica [Brioschi et al. 2010]. Através deste trabalho foi possível também testar a capacidade de usar algoritmos CNNs para classificar possíveis nódulos tireoidianos, demonstrando que existem possíveis padrões que permitem diferenciar esse tipo de anomalia.

2.1. Materiais e métodos

Os dados usados neste estudo foram adquiridos por uma câmera FLIR SC620, com sensibilidade de 0,04 C, seguindo o protocolo descrito em [González 2017]. A fim de verificar se as temperaturas na região da tireoide são influenciadas pela presença de nódulos, sejam eles malignos ou benignos, foram consideradas termografias de dezoito pacientes com nódulos malignos e os oitenta e cinco pacientes com nódulos benignos. A informação da localização dos nódulos foi fornecida por especialistas médicos, que se basearam na palpação e em dados de ultrassonografias para defini-las. Para cada uma delas, os seguintes parâmetros estatísticos de temperatura foram calculados: mínima, média, máxima e mediana. Para o cálculo desses valores considerou-se apenas as pixels que correspondiam a região da tiroide próxima a cada termografia, desconsiderando as regiões de fundo e outras partes do corpo que não pertenciam ao pescoço.

Para verificar as hipóteses de que tireoides saudáveis, com nódulos benignos e com nódulos malignos podem ser diferenciadas através dessas características, esses parâmetros foram calculados para as termografias dos pacientes aqui considerados. Os valores encontrados foram avaliados usando Tamanho do Efeito e o Teste Wilcoxon Bicaudal para amostras independentes com nível de confiança 95% , a fim de verificar se eram de fato estatisticamente diferentes. Tais avaliações foram escolhidas devido ao volume baixo, os quais poderiam não apresentar distribuições normais.

Outro parâmetro a ser estudado é o índice termográfico (TI), inicialmente definido por [Collins et al. 1974] e usado em estudos similares, considerando outras partes do corpo [Brioschi et al. 2010]. Foi avaliado se as termografias de tireoides que apresentam nódulos têm valores diferentes de TI das termografias de tireoides saudáveis.

Para avaliar se apenas a informação da temperatura é suficiente para identificar os pontos referentes à região nodular, aplica-se posterizações usando níveis de cinza 4, 8 e 16 nas imagens a fim de observar visualmente agrupamentos de temperatura. Considerando que os tons nessas imagens refletem os valores capturados no mapa de temperatura, caso haja um padrão visual de tons que correspondem às regiões nodulares, então poderia haver um padrão de temperatura que indicaria a presença de nódulos.

Também é possível comparar pontos com outros de uma região considerada equivalente, que no caso de órgão horizontalmente simétricos, corresponde a mesma região no lado oposto em relação a um eixo de simetria. Essa análise pode ser feita através do parâmetro de assimetria (P_A). Como já comentado, a tireoide é um órgão bastante simétrico horizontalmente e suficientemente superficial para permitir a detecção de calor causado por hipermetabolismo de nódulos ou da glândula em si. Segundo [Brioschi et al. 2010], o parâmetro de assimetria é definido pela diferença de temperatura de uma região em relação ao seu lado contralateral, considerando um eixo vertical de simetria do corpo. Valores do P_A acima de 0,3°C podem indicar alguma disfunção, sendo

que, de forma geral, se o valor ultrapassar 1°C algum problema significativo pode estar ocorrendo na região [Brioschi et al. 2010].

Para verificar se P_A é adequado para identificar regiões nodulares, testes estatísticos Wilcoxon foram aplicados. Para cada um dos casos estudados, todos os pixels da região da tireoide nas termografias tiveram seus P_A calculados. Para cada paciente, os valores de P_A são agrupados em dois conjuntos: P_{AN} - valores de P_A dos pontos que estão na região nodular; e P_{AS} - valores de P_A dos demais pontos.

Após avaliar os parâmetros a serem usados na segmentação de regiões possivelmente nodulares, analisou-se o melhor CNN para a classificação dessas regiões. Para isso, algumas variações de CNNs foram criadas, considerando diferentes parâmetros e dados de entrada. Já que é apontado pela literatura que um volume maior de dados no treinamento de CNNs leva a melhores resultados, foram usados, além do conjunto de dados inicial (DT1), outros 4 datasets (DT2, DT3, DT4 e DT5), obtidos ao aplicar o processo de data augmentation em DT1. As operações de espelhamento e mudança de tons foram usadas na criação de dados sintéticos. Cada um desses datasets apresentam o dobro de imagens em relação ao anterior. Ressalta-se que este se trata de um estudo preliminar, já que o volume de dados disponível é muito restrito para levar a conclusões definitivas.

Neste estudo foram selecionados duas arquiteturas de CNNs que apresentam bons resultados em problemas similares: ResNet e GoogLeNet. Os valores de taxa de aprendizado usados foram: 0,1, 0,01 e 0,001. A técnica de cross-validation foi descartada devido ao alto custo computacional. Todas as redes foram treinadas por 50 épocas. Os subconjuntos de treinamento e teste correspondem a 80% e 20% de cada dataset.

3. Resultados obtidos e discussão

Ao analisar os parâmetros estatísticos, os p-valores encontrados indicam que as temperaturas mínima, máxima, média e mediana não podem ser utilizadas para diferenciar pacientes saudáveis dos que apresentam nódulos malignos ou benignos considerando apenas os casos deste estudo. Para complementar esta análise, calculou-se o tamanho do efeito para essas estatísticas considerando os mesmos grupos e os valores encontrados para TDE apontam que nenhuma das estatísticas demonstram uma alta diferenciação para todos os grupos. Assim, o uso de apenas esses parâmetros para a identificação de regiões nodulares neste estudo está descartado.

A fim de avaliar se os valores de TI para cada um dos grupos são significativamente diferentes entre eles, realiza-se o teste Wilcoxon. O p-valor resultante desse teste é 0,26 para os grupos de controle e com nódulos malignos. Já para os grupos de controle e com nódulos benignos o p-valor encontrado foi de 0,97. O p-valor para o teste com os grupos com nódulos malignos e com nódulos benignos foi de 0,99. Isso indica que o TI não pode ser usado para diferenciar pacientes de diferentes grupos considerando um nível de confiança de 95%. Os valores obtidos para os TDEs apontam também que o TI não é apropriado para diferenciar os grupos.

Passando à análise de parâmetros para segmentação, ao avaliar visualmente os resultados da posterização das termografias de todos os pacientes com nódulos, observa-se que para nenhum dos agrupamentos de tons usados foi possível definir um grupo que indique regiões que se aproximassem das identificadas pelos especialistas como nodulares

para maior parte das termografias avaliadas, tanto as de nódulos malignos, quanto as de nódulos benignos. Isso sugere que apenas a informação de temperatura não é suficiente para identificar tais regiões.

O outro parâmetro de segmentação analisado é o P_A . Verifica-se com o teste de Wilcoxon, se os valores de P_A das regiões nodulares (P_{AN}) são significativamente superiores em relação aos das demais regiões (P_{AS}) considerando os pacientes que apresentam nódulos malignos. Para a maioria dos casos malignos, os p-valores concordam, ou se aproximam da hipótese de que P_A pode ser utilizado na detecção de regiões referentes a nódulos malignos. Dessa forma, o P_A foi o parâmetro escolhido para detecção de regiões nodulares malignas neste trabalho. Para os oitenta e cinco pacientes com nódulos benignos, 7 apresentaram p-valor que concorda com a hipótese de que as regiões nodulares apresentam maior temperatura. Isso está de acordo com o esperado, já que nódulos benignos não apresentam características que levariam a um aumento de temperatura.

Assim, o processo de segmentação dos casos de nódulos malignos foi realizado usando diferentes valores de limiar iguais a: 0,31, 0,55, 0,78, 1,02, 1,25 e 1,48. O resultado das segmentações foi avaliado por um especialista médico, que comparou visualmente esses resultados com os dados da ultrassonografia e as termografias com marcadores indicando a região dos nódulos. Para três pacientes não foi possível encontrar regiões que apresentassem alguma correspondência com nódulos apontados pelos especialistas com nenhum valor usado. De forma geral, para as termografias analisadas neste estudo, o limiar usado permitiu, para grande parte delas, identificar regiões com localização correspondente ao indicado pelos especialistas, e dimensões próximas ao indicado nas ultrassonografias. Porém, não houve um único valor que gerasse essas regiões para todos os casos, o que indica que a diferença de temperatura está relacionada a cada caso, considerando o paciente e as características do nódulo em si.

Na análise dos resultados dos algoritmos CNNs na classificação foram observados os valores de acurácia (accuracy) de cada CNN alcançados ao longo das épocas durante o treinamento. Observou-se que, de forma geral, a rede ResNet mantém uma tendência crescente e apresenta melhores resultados que a GoogLeNet. Para a rede GoogLeNet não é possível identificar uma tendência. Os melhores resultados são os obtidos com a modelo ResNet com o dataset DT4, o que leva a concluir que o acréscimo de imagens do conjunto DT4 para o DT5 não introduz melhora no desempenho da rede. Os cinco maiores valores de acurácia alcançados nos experimentos se referem a CNNs do modelo ResNet. Dentre os melhores resultados da GoogLeNet estão os do experimento com DT4 com taxa 0,01. Assim, de forma geral, o conjunto que permite melhor desempenho é o DT4. Os resultados com DT1, DT2 e DT3 são bastante oscilantes e baixos, o que é esperado devido à quantidade de dados da entrada.

Ao observar também as demais métricas de avaliação do desempenho das redes (acurácia de validação, perda e perda de validação) conclui-se que as redes ResNet com DT4 e taxas de aprendizado 0,1 e 0,01 apresentam os melhores resultados, já que os valores de acurácia de validação se aproximam dos valores de acurácia, e os valores de perda e perda de validação são relativamente baixos e se aproximam, o que aponta um bom *fitting*. Observa-se em geral, que diversas variações testadas nos experimentos apresentam valores de perda altos. Também foi observado que diversos testes levaram a casos de *overfitting* e *underfitting*, sendo esse modelos considerados portanto inadequados para o

problema aqui apresentado.

Os resultados encontrados ao longo deste estudo foram apresentados em diversos artigos publicados em meios científicos de grande alcance [Moran et al. 2018b, Moran et al. 2018a, González et al. 2018]. Agradeço a fundação CAPES pelo apoio financeiro a mim concedido que me permitiu realizar este trabalho

Referências

- Alves, M. L. D., Andrade, J., Cherri, J., Moriya, T., Piccinato, C. E., Okano, N., Llorach Velludo, M. A., and Iazigi, N. (1988). Papel da termografia da seleção de nódulos tireoideanos de indicação cirúrgica. *Arq. bras. endocrinol. metab*, 32(4):97–9.
- Brenner, D. J. and Hall, E. J. (2007). Computed tomography—an increasing source of radiation exposure. *New England Journal of Medicine*, 357(22):2277–2284.
- Brioschi, M., Teixeira, M., Silva, F., and Colman, D. (2010). Medical thermography textbook: principles and applications. *Editores e Livrarias Andreoli*.
- Collins, A., Ring, E., Cosh, J., and Bacon, P. (1974). Quantitation of thermography in arthritis using multi-isothermal analysis. i. the thermographic index. *Annals of the rheumatic diseases*, 33(2):113.
- Goldbach-Mansky, R., Woodburn, J., Yao, L., and Lipsky, P. E. (2003). Magnetic resonance imaging in the evaluation of bone damage in rheumatoid arthritis: a more precise image or just a more expensive one? *Arthritis & Rheumatism*, 48(3):585–589.
- González, J., Pupo, Y., Hernandez, M., Conci, A., Machenry, T., and Fiirst, W. (2018). On image registration for study of thyroid disorders by infrared exams. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICIP)*, pages 151–158. The Steering Committee of The World Congress in Computer Science, Computer
- González, J. R. (2017). TUm estudo sobre a possibilidade do uso de imagens infravermelhas na análise de nódulos de tireoide. Master’s thesis, Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil.
- Mahajan, P. and Madhe, S. (2014). Morphological feature extraction of thermal images for thyroid detection. *Int. J. Electron. Comput. Eng*, 5(4):11–14.
- Moran, M., Conci, A., Gonzalez, J., Saade, D., Fiirst, W., Araújo, A., Damião, C., Lima, G., Cruz Filho, R., and Resmini, R. (2018a). Monomodal image registration by tensor analysis: going beyond the brightness constancy assumption. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–8. IEEE.
- Moran, M. B. H., Conci, A., Gonzalez, J. R., Araujo, A. S., Fiirst, W., and Damiao, C., L. G. F. R. (2018b). Identification of thyroid nodules in infrared images by convolutional neural networks. In *31st International Joint Conference on Neural Networks (IJCNN)*. IEEE.
- Vaz, V. A. S. (2014). Diagnosis of hypo and hyperthyroid using mlpn network. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(7):14314–23.