

Classificação entre Doença de Graves e Bócio Multinodular Tóxico por Imagens de Cintilografia

Anna Caroline Bozzi, Mateus Felipe de Cássio Ferreira

Departamento de Informática – Universidade Federal do Paraná (UFPR)
Curitiba – PR – Brasil

{acb17,mfcf17}@inf.ufpr.br

Abstract. *Graves' Disease and Toxic Multinodular Goiter are two manifestations of hyperthyroidism, a disease in which there is an excessive production of thyroid hormones and can generate several symptoms in patients. The correct identification of the forms of appearance of this disease can give agility in the beginning of the process of adequate treatment with an endocrinologist for the hormonal control of the thyroid gland. This work intends to use a machine learning approach to classify these forms, however, the amount of available images was a limiting factor in obtaining better classification results.*

Resumo. *Doença de Graves e Bócio Multinodular Tóxico são duas formas de manifestação do hipertireoidismo, uma doença em que há uma produção excessiva dos hormônios tireoidianos e pode gerar vários sintomas nos pacientes. A identificação correta das formas de aparição dessa doença pode dar uma agilidade no início do processo de tratamento adequado junto a um endocrinologista para o controle hormonal da glândula tireóide. Este trabalho pretende utilizar uma abordagem de aprendizagem de máquina para a classificação dessas formas, no entanto, a quantidade de imagens disponíveis foi um fator limitante na obtenção de melhores resultados de classificação.*

1. Introdução

A tireóide é uma glândula que está localizada na porção anterior do pescoço e é uma das principais glândulas para o funcionamento normal do metabolismo de um organismo e é responsável pela produção dos hormônios T3 e T4. Sabe-se que a produção excessiva dos hormônios tireoidianos (chamado, nesse caso, de hipertireoidismo), promove aumento do metabolismo e pode gerar vários sintomas, sendo necessário diagnóstico preciso e tratamento adequado junto ao endocrinologista para promover o controle hormonal (EINSTEIN, 2022).

Estima-se que cerca de 1% a 2% das pessoas irão desenvolver hipertireoidismo em algum momento de suas vidas, sendo que já existem dados coletados nos Estados Unidos da América que estimam que 500 mil novos casos são detectados a cada ano dessa doença (EINSTEIN, 2022).

A causa mais frequente do hipertireoidismo é a Doença de Basedow-Graves, também chamada de Bócio Difuso Tóxico (BDT). Esse é um distúrbio autoimune em que a síntese e secreção excessiva de hormônios da tireóide são desencadeados pelo

ataque autoimune a receptores específicos da glândula e de outros tecidos do corpo (JALEKO, 2022).

As outras formas de hipertireoidismo são relacionadas ao surgimento de um ou mais nódulos tireoidianos que se tornam autônomos, ou seja, não necessitam de estímulo para funcionar e, dependendo da quantidade de células autônomas, podem produzir hipertireoidismo (EINSTEIN, 2022). O quadro em que há a presença de vários nódulos é chamado de Bócio Multinodular Tóxico (BMNT) e é responsável por cerca de 10% dos casos de hipertireoidismo (EINSTEIN, 2022).

O objetivo deste trabalho é o de analisar imagens cintilográficas que foram obtidas da tireóide de pacientes diagnosticados com hipertireoidismo. A cintilografia é um nome dado ao procedimento de diagnóstico por imagem na especialidade da medicina nuclear, utilizado para avaliar a função de diversos órgãos. Assim, ao final deste trabalho, pretende-se utilizar algoritmos e técnicas de aprendizagem de máquina na tentativa de classificar um conjunto de 12 pacientes nas duas variações de hipertireoidismo abordadas anteriormente: BDT e BMNT. A Figura 1 apresenta duas imagens de cintilografia das duas classes que serão estudadas neste trabalho.

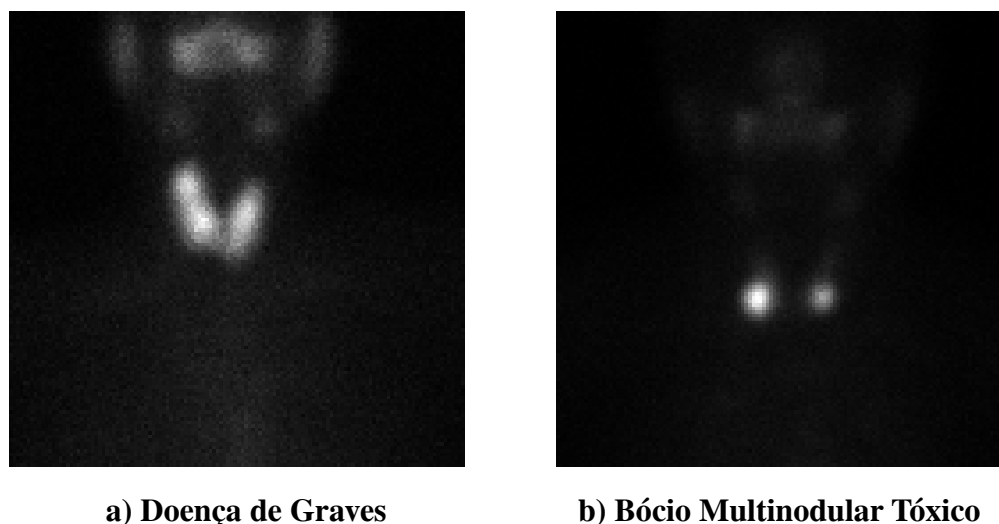


Figura 1 - Imagens cintilográficas de duas formas de hipertireoidismo.

2. Conjunto Amostral e Extração de Características

O conjunto amostral deste experimento consiste em 12 imagens de 12 pacientes diagnosticados com algum tipo de hipertireoidismo. Nesse caso, 6 pacientes foram diagnosticados com BDT e outros 6 com BMNT.

Para a extração das características das imagens, foi utilizado o *Local Binary Pattern* (LBP), que resume as estruturas locais na imagem gerando uma representação estatística dessas estruturas. Quando os *pixels* ao redor são todos pretos ou brancos, a região da imagem é plana, ou seja, sem características. Grupos de *pixels* pretos ou brancos contínuos são padrões “uniformes”, interpretados como cantos ou arestas, se os *pixels* alternarem entre pretos e brancos o padrão é considerado “não uniforme”. Foi utilizado a função *local_binary_pattern* da biblioteca *skimage.feature* em seguida

normalizado o histograma retornado, e por fim criado um arquivo com as 25 características extraídas de cada imagem juntamente com as suas classes.

3. Classificação

3.1 Classificadores

Os classificadores utilizados neste trabalho foram:

- *Random Forest*: ele cria uma floresta de um modo aleatório. A floresta que ele cria é uma combinação (*ensemble*) de árvores de decisão.
- *Decision Tree*: a árvore de decisão estabelece nós que se relacionam entre si por uma hierarquia. Existe o nó-raiz, que é o mais importante, e os nós-folha, que são os resultados finais.
- *AdaBoost*: é uma técnica de *boost* muito popular que combina vários classificadores fracos em um único classificador forte.
- *K-Nearest Neighbors (kNN)*: é um classificador onde o aprendizado é baseado no quão similar é um dado do outro.

3.2 Métodos

Como principal método de classificação foi usado uma técnica que se chama “*Leave One Patient Out*”. Esse método consiste em dos 12 pacientes deixar um disponível para testar, e os outros 11 pacientes serão usados para treino. E assim repete-se até que todos os pacientes tenham passado pelo teste. Assim temos então 12 resultados de classificação. Para os 12 resultados são extraídos as métricas de Acurácia, Precisão, *Recall* e F1 Score. Foi feita a soma e então a média e desvio padrão de cada uma dessas métricas para consolidar em um único valor para cada um dos 4 classificadores utilizados.

Além disso, foi utilizado a função *LeaveOneOut()* da biblioteca *sklearn.model_selection* do Python para aplicar a técnica de “*Leave One Patient Out*”.

4. Resultados e Discussões

Nesta seção serão apresentados os resultados obtidos nos experimentos realizados. Esses resultados estão sumarizados, basicamente, em quatro grandes tabelas. A Figura 2 apresenta os resultados obtidos pelas métricas avaliadas. Assim, para cada um dos quatro classificadores, foram obtidas medidas relativas ao F1-Score, Acurácia, Precisão e *Recall*, além de apresentar, por meio da matriz de confusão, as medidas de Sensibilidade e Especificidade.

A partir da análise dessa figura é possível perceber que o classificador que obteve uma maior medida de acurácia das classificações realizadas foi o kNN, com uma média de 0,583 e desvio padrão de 0,515 (nota-se, nesse caso, que todos os classificadores apresentaram medidas de desvio padrão significativas para essa métrica). Em contrapartida, o classificador que apresentou a pior medida para a acurácia foi o *Decision Tree*.

Além disso, uma outra discussão importante a ser feita diz respeito à sensibilidade e especificidade de cada um dos classificadores avaliados, a partir da

análise da matriz de confusão de cada um dos testes realizados. Sabe-se que a medida de sensibilidade está relacionada a quanto o classificador acerta o problema a ser resolvido, ou seja, a classificação correta entre as duas classes envolvidas. Por outro lado, a medida da especificidade diz respeito a quanto o classificador acerta o não problema, ou seja, quão bom um classificador é para determinar a normalidade dada a uma determinada classe. Assim, para a área médica, um sistema de classificação robusto deve ser o mais sensível e o mais específico possível, uma vez que falsos positivos e falsos negativos são bastante prejudiciais durante um diagnóstico.

Nota-se que, nesse caso, o classificador que obteve maiores valores das métricas de sensibilidade e especificidade também foi o kNN, com valores de 0,6 e 0,571 respectivamente.

AdaBoost				
MEDIDA	F1-Score	ACURÁCIA	PRECISÃO	RECALL
MÉDIA	0,250	0,417	0,250	0,250
DESVIO PADRÃO	0,452	0,515	0,452	0,452
Sensibilidade			0,428	
Especificidade			0,400	

Decision Tree				
MEDIDA	F1-Score	ACURÁCIA	PRECISÃO	RECALL
MÉDIA	0,167	0,333	0,167	0,167
DESVIO PADRÃO	0,389	0,492	0,389	0,389
Sensibilidade			0,333	
Especificidade			0,333	

kNN				
MEDIDA	F1-Score	ACURÁCIA	PRECISÃO	RECALL
MÉDIA	0,250	0,583	0,250	0,250
DESVIO PADRÃO	0,452	0,515	0,452	0,452
Sensibilidade			0,600	
Especificidade			0,571	

Random Forest				
MEDIDA	F1-Score	ACURÁCIA	PRECISÃO	RECALL
MÉDIA	0,250	0,500	0,250	0,250
DESVIO PADRÃO	0,452	0,522	0,452	0,452
Sensibilidade			0,500	
Especificidade			0,500	

Figura 2 - Resultados obtidos dos experimentos de classificação realizados.

5. Conclusão

Nesse sentido, nota-se que o conjunto pequeno de imagens para serem utilizadas por algoritmos de aprendizagem de máquina foi um fator limitante na obtenção de resultados de classificação melhores para o problema proposto. Apesar disso, este trabalho apresentou um direcionamento de um classificador que poderia ser utilizado para o problema proposto e, com uma quantidade maior de imagens, poderia obter resultados significativos para a classificação de imagens entre Doença de Graves e Bócio Multinodular Tóxico.

Referências

- EINSTEIN, Hospital Israelita Albert. Hipertireoidismo. 2022. Disponível em: <<https://www.einstein.br/Pages/Doenca.aspx?eid=146>>. Acesso em: 29 abr. 2022.
- JALEKO. Doença de Graves. 2022. Disponível em: <https://jaleko-files.s3-sa-east-1.amazonaws.com/apostila-web/5f735a6d8b857_Ebook%20-%20Doen%C3%A7a%20de%20Graves.pdf>. Acesso em: 29 abr. 2022.