

1 2 9 0



UNIVERSIDADE D  
COIMBRA

Tiago Alexandre Cunha Rocha

**A ECOGRAFIA *VERSUS* ELASTOGRAFIA NO  
ESTUDO DE NÓDULOS DA TIROIDE**

**VOLUME 1**

**Dissertação no âmbito do Mestrado em Engenharia  
Biomédica, especialidade em Instrumentação Biomédica  
orientada pelo Prof. Dr. Jaime Santos, pelo Prof. Dr. José  
Silvestre Silva e pela Prof. Dra. Alexandra André e  
apresentada à Universidade de Coimbra**

Setembro de 2023

**A ECOGRAFIA *VERSUS* ELASTOGRAFIA NO  
ESTUDO DE NÓDULOS DA TIROIDE**

Tiago Alexandre Cunha Rocha

**VOLUME 1**

**Dissertação no âmbito do Mestrado em Engenharia  
Biomédica, especialidade em Instrumentação Biomédica  
orientada pelo Prof. Dr. Jaime Santos, pelo Prof. Dr. José  
Silvestre Silva e pela Prof. Dra. Alexandra André e  
apresentada à Universidade de Coimbra**

1 2 9 0



**UNIVERSIDADE DE  
COIMBRA**

Setembro de 2023





## **Agradecimentos**

Expresso os meus sinceros agradecimentos a todas as pessoas que contribuíram para a execução e conclusão desta dissertação de mestrado. Primeiramente, quero agradecer aos meus orientadores, Jaime Santos, José Silvestre Silva e Alexandra André, pelo apoio e compromisso ao longo de todo este processo.

Também quero agradecer à minha família pelo apoio inabalável e encorajamento durante todos os meus anos de estudo.

Aos meus colegas e amigos, obrigado por compartilharem ideias, sugestões e experiências que enriqueceram tanto este trabalho como todo o meu percurso.

## Resumo

Os nódulos da tiroide, embora se apresentem como lesões discretas, são uma patologia bastante comum na população mundial. A deteção precoce e atempada é de grande importância para a saúde dos pacientes, visto que tal proporcionará um acompanhamento médico permanente com o consequente controlo da evolução da patologia. Os nódulos da tiroide são frequentemente de origem benigna, no entanto, este tipo de patologias pode evoluir para condições graves, como é o caso do cancro da tiroide.

Neste estudo, foram implementadas técnicas supervisionadas baseadas em algoritmos de inteligência artificial para analisar imagens médicas obtidas por ecografia e elastografia. O objetivo principal foi avaliar a capacidade de identificação e classificação de imagens quanto à presença de nódulos da tiroide. Tendo em vista o objetivo proposto, foram estudados e comparados dois métodos distintos com vista à extração de *features* e dos padrões mais importantes das imagens em questão. Inicialmente, foi implementado um algoritmo de *machine learning* e, posteriormente foi realizada uma abordagem baseada em *deep learning*.

Assim, os resultados obtidos com *deep learning* superaram consistentemente os resultados de *machine learning*, para ambos os tipos de imagens. Destaca-se a excelente capacidade da primeira técnica para extrair as *features* mais relevantes das imagens médicas, tendo-se obtido melhores resultados para as imagens provenientes da ecografia. Com *machine learning*, foi alcançado um *F1-Score* de 97,20% para as imagens de ecografia, de 75,40% para as imagens de elastografia e 75,64% quando se combinaram os dois tipos de imagens. A abordagem de *deep learning* alcançou *F1-Scores* de 98,85%, 89,15% e 93,70% para as imagens ecográficas, elastográficas e para a combinação das duas técnicas, respetivamente.

Os resultados obtidos neste estudo são promissores e apresentam um potencial significativo no campo da identificação de nódulos da tiroide. A capacidade demonstrada por ambas as abordagens abre portas para aplicações futuras que podem melhorar substancialmente o diagnóstico da patologia da tiroide.

**Palavras-chave:** nódulos da tiroide, ecografia, elastografia, classificação, *machine learning*, *deep learning*

## **Abstract**

Thyroid nodules, although they appear as discrete lesions, are a fairly common pathology in the world's population. Early and timely detection is of great importance to patients' health, as this will provide ongoing medical monitoring and consequent control of the pathology's progression. Thyroid nodules are often benign in origin; however, this type of pathology can develop into serious conditions, such as thyroid cancer.

In this study, supervised techniques based on artificial intelligence algorithms were implemented to analyse medical images obtained by ultrasound and elastography. The main objective was to evaluate the ability to identify and classify thyroid nodules in the images. With the proposed objective in mind, two different methods were studied and compared in order to extract the most important features and patterns from images of the organ in question. Initially, a machine learning algorithm was implemented, followed by an approach based on deep learning.

The results obtained with deep learning consistently surpassed those obtained with machine learning, for both types of images. The excellent ability of the first technique to extract the most relevant features from medical images stands out, with the best results being obtained for ultrasound images. With machine learning, an F1-Score of 97.20% was achieved for ultrasound images, 75.40% for elastography images and 75.64% when both types of images were combined. Deep learning achieved F1-Scores of 98.85%, 89.15% and 93.70% for ultrasound, elastography and the combination of the two techniques, respectively.

The results obtained in this study are promising and show significant potential in the field of thyroid nodule identification. The ability demonstrated by both approaches opens the door to future applications that could substantially improve the diagnosis of thyroid pathology.

**Keywords:** thyroid nodules, ultrasound, elastography, classification, machine learning, deep learning



# Índice

Agradecimentos .....	iv
Resumo .....	v
Abstract .....	vi
Lista de Figuras .....	x
Lista de Tabelas .....	xii
Lista de abreviaturas, siglas, acrónimos e símbolos .....	xiii
1 Introdução .....	1
1.1 Enquadramento e motivação .....	1
1.2 Objetivos .....	1
1.3 Estrutura da dissertação .....	2
1.4 Contribuição científica .....	2
1.5 Comissão de Ética .....	3
2 Conceitos teóricos .....	4
2.1 A tiroide .....	4
2.2 Doenças da tiroide .....	6
2.3 Técnicas imagiológicas .....	9
2.3.1 Ecografia .....	9
2.3.2 Elastografia .....	11
2.3.3 Tomografia computadorizada (TAC) .....	12
2.3.4 Ressonância magnética (RM) .....	13
2.3.5 Tomografia por Emissão de Positrões (PET) .....	13
3 Estado da arte .....	14
3.1 Métodos tradicionais .....	14
3.2 Métodos recentes .....	15
3.3 Métodos híbridos .....	17
3.4 Bases de dados .....	19
4 Metodologia .....	20
4.1 <i>Machine Learning</i> .....	20
4.1.1 Pré-processamento .....	20

4.1.2	Processamento.....	22
4.1.3	Classificação .....	23
4.1.3.1	<i>k-Fold-cross-validation</i> .....	24
4.2	<i>Deep Learning</i> .....	25
4.2.1	Pré-processamento .....	25
4.2.1.1	Aumento de dados .....	25
4.2.1.2	<i>Transfer Learning</i> .....	26
4.2.2	Processamento.....	27
4.3	Métricas utilizadas .....	29
5	Resultados .....	31
5.1	Imagens .....	31
5.2	Implementação em <i>Python</i> .....	33
5.3	<i>Machine Learning</i> .....	33
5.3.1	Desempenho dos classificadores.....	34
5.3.2	Tamanho do passo.....	35
5.3.3	Tamanho da janela .....	36
5.3.4	Normalização e seleção de <i>features</i> .....	37
5.3.5	<i>k-Fold Cross-Validation</i> .....	38
5.3.6	Número de <i>features</i> .....	38
5.3.7	Ecografia <i>versus</i> Elastografia .....	39
5.4	<i>Deep Learning</i> .....	41
5.4.1	Tamanho do conjunto de dados .....	41
5.4.2	Número de iterações .....	42
5.4.3	Otimização .....	43
5.4.4	Desempenho da VGG16 com o mecanismo de atenção .....	44
5.5	<i>Machine Learning versus Deep Learning</i> .....	45
6	Conclusão.....	47
7	Referências.....	49
A.	Anexo A .....	54
B.	Anexo B .....	55
C.	Anexo C .....	56

## Lista de Figuras

Figura 2.1 - Anatomia da tiroide (vista anterior) (adaptado de [4]).....	5
Figura 2.2 - Detalhes histológicos da glândula tiroideia (adaptado de [2]).....	6
Figura 2.3 - Representação de um nódulo da tiroide (adaptado de [6]).....	8
Figura 2.4 - Imagem ecográfica da tiroide (vista transversal) com legenda (adaptado de [9]).....	9
Figura 2.5 - Representação da onda acústica num referencial.....	10
Figura 4.1 - Esquema representativo do funcionamento da janela deslizante .....	21
Figura 4.2 - Aplicação da janela deslizante a uma imagem ecográfica da tiroide.....	21
Figura 4.3 - Funcionamento do k-Fold-Cross-Validation (adaptado de [38]).....	24
Figura 4.4 - Aumento de dados através de rotações e inversões da imagem original .....	25
Figura 4.5 - Esquema representativo do uso do Transfer Learning.....	26
Figura 4.6 - Arquitetura de uma rede neural VGG16 com legenda das camadas constituintes (adaptado de [40]).....	28
Figura 5.1 - Aquisição de imagens ecográficas pelo profissional de saúde.....	31
Figura 5.2 - Ecografia (esquerda) e elastografia (direita) de um nódulo da tiroide e configurações de aquisição. O código de cores apresentado refere-se à rigidez do tecido: o vermelho corresponde a tecido mais rígido e o azul a tecido menos rígido. ....	32
Figura 5.3 - ROIs: ecografia (esquerda) e elastografia (direita) .....	32
Figura 5.4 – Resultados do desempenho dos classificadores (SVM, KNN, RF e DT) para variações de tamanho da janela e tamanho do passo (e.g., 5x5 7 – 5x5 corresponde ao tamanho da janela (píxeis) e 7 ao tamanho do passo (píxeis)). As métricas de avaliação incluem a exatidão, precisão, F1-Score e AUC. ....	34
Figura 5.5 - Desempenho do classificador SVM em função da variação do tamanho do passo (tamanho da janela = 17x17 píxeis) .....	35

Figura 5.6 - Desempenho do classificador SVM em função da variação do tamanho da janela (tamanho do passo = 9 píxeis).....	36
Figura 5.7 - Desempenho do classificador SVM relativamente a diferentes combinações das funções de normalização com as funções de seleção de features.....	37
Figura 5.8 - Comparação do desempenho do classificador SVM utilizando 5-fold cross-validation e 10-fold cross-validation .....	38
Figura 5.9 - Desempenho (F1-Score) do classificador SVM em função da variação do número de features .....	39
Figura 5.10 – Desempenho do classificador SVM (F1-Score) para imagens de ecografia, elastografia e para a combinação das duas modalidades de imagem.....	40
Figura 5.11 – Desempenho da VGG16 para diferentes tamanhos do conjunto de dados	41
Figura 5.12 - Desempenho da VGG16 para diferentes números de iterações .....	43
Figura 5.13 -Desempenho (F1-Score) da VGG16 para diferentes otimizadores.....	44
Figura 5.14 – Desempenho (F1-Score) da VGG16 para diferentes tipos de imagens (ecografia, elastografia e combinação das duas). Estudo da influência do mecanismo de atenção .....	44
Figura 5.15 - Comparação dos resultados obtidos utilizando machine learning e deep learning .....	45

## Lista de Tabelas

Tabela 3.1 – Informações sobre as bases de dados públicas.....	19
Tabela A.1 - Bibliotecas e respectivos extratores de features .....	54
Tabela B.1 - Número de imagens sem nóculo e com nóculo, tamanho do passo e rácio (imagens sem nóculo/imagem com nóculo), para um tamanho da janela de 17x17 píxeis. ....	55
Tabela B.2 - Número de imagens sem nóculo e com nóculo, tamanho da janela e rácio imagens sem nóculo/imagem com nóculo), para um tamanho do passo de 9 píxeis. ....	55

## Lista de abreviaturas, siglas, acrónimos e símbolos

2D	Bidimensional
3D	Tridimensional
Adam	<i>Adaptive Moment Estimation</i>
Adagrad	<i>Adaptive Gradient Algorithm</i>
AUC	<i>Area Under the ROC Curve</i>
CAD	<i>Computer-Aided Design</i>
CNN	Rede neural convolucional
CPU	Unidade Central de Processamento
DDTI	<i>Digital Database of Thyroid Ultrasound Images</i>
DT	<i>Decision Tree</i>
FN	Número de amostras negativas incorretamente classificadas
FP	Número de amostras positivas incorretamente classificadas
Fps	<i>Frames per second</i>
GPU	Unidade de Processamento Gráfico
KNN	<i>K-nearest neighbors</i>
LDA	Análise de discriminação linear
MCDLM	<i>Multi-task cascade deep learning mode</i>
PACS	Sistema de arquivo e comunicação de imagens
PET	Tomografia por Emissão de Positrões
ReLU	<i>Rectified Linear Unit</i>
RF	<i>Random Forest</i>
RFE	<i>Recursive Feature Elimination</i>
RGB	<i>Red, Green, Blue</i>

RM	Ressonância Magnética
RMSprop	<i>Root Mean Square Propagation</i>
ROC	<i>Receiver Operating Characteristic</i>
ROI	Região de Interesse
SGD	<i>Stochastic Gradient Descent</i>
SVM	<i>Support Vector Machine</i>
SWE	Elastografia por onda <i>shear</i>
sSWE	Elastografia Sintética por onda <i>shear</i>
TAC	Tomografia Computadorizada
TN	Número de amostras negativas corretamente classificadas
TP	Número de amostras positivas corretamente classificadas
TSH	Hormona tiro-estimulante
US	<i>Ultrasound</i> (Ecografia)

# 1 Introdução

## 1.1 Enquadramento e motivação

A glândula da tiroide localiza-se na zona inferior do pescoço e desempenha um papel crucial na regulação do metabolismo do corpo humano. Os nódulos da tiroide são uma patologia comumente associada a esta glândula, afetando uma grande fatia da população mundial. O diagnóstico precoce desta patologia permite antecipar um tratamento adequado e prevenir a existência de riscos acrescidos desnecessários. A análise e diagnóstico dos nódulos da tiroide é efetuada primordialmente pela ecografia, tendo sido introduzida nos últimos anos uma nova técnica, a elastografia. Estas técnicas fornecem informações relevantes acerca da forma, tamanho e composição dos nódulos. Porém, a identificação e classificação das imagens nem sempre é linear, depende da interpretação do observador, o que pode levar a resultados dúbios e imprecisos.

Recentemente, as abordagens associadas à inteligência artificial têm apresentado bons resultados e com elevado potencial em diferentes contextos clínicos, como no caso do estudo de imagens médicas. Assim, com o desenvolvimento deste estudo pretende-se criar um mecanismo complementar e viável para auxiliar os profissionais de saúde na deteção dos nódulos da tiroide.

## 1.2 Objetivos

Neste trabalho, pretende-se levar a efeito um estudo comparativo envolvendo duas abordagens de inteligência artificial, *machine learning* e *deep learning*, com vista à classificação de imagens quanto à existência de nódulos da tiroide. As imagens são adquiridas por ecografia e elastografia. Com as metodologias de classificação propostas, pretende-se obter uma caracterização eficaz das imagens da tiroide, constituindo, assim, um importante contributo para o estado da arte como ferramenta de auxílio no diagnóstico médico supervisionado.

### 1.3 Estrutura da dissertação

A presente dissertação apresenta uma estrutura organizada de modo a introduzir e encadear de forma coesa todos os aspetos relevantes. No capítulo 2, são apresentados os conceitos teóricos relativos à tiroide, as patologias associadas e as diferentes técnicas imagiológicas utilizadas no diagnóstico dos nódulos da tiroide. No capítulo 3, é exibida uma revisão da literatura, sendo relatados os avanços mais recentes relativamente à classificação de imagens de ecografia e elastografia de nódulos da tiroide, através de algoritmos de inteligência artificial. O estado da arte encontra-se dividido em três secções. Inicialmente, são apresentados os métodos de classificação tradicional, *machine learning*, posteriormente, os métodos baseados nas técnicas mais recentes de *deep learning* e, por fim, é feita uma análise aos estudos que empregam uma combinação dos dois métodos anteriormente referidos, ou seja, uma abordagem híbrida. No quarto capítulo, são apresentadas as metodologias adotadas neste estudo, realçando as etapas de pré-processamento e processamento das imagens, bem como a implementação dos algoritmos de classificação. O capítulo 5, contém todos os resultados obtidos durante o estudo e respetiva discussão. O sexto e último capítulo, é dedicado às conclusões, onde são apresentados os avanços alcançados na temática em estudo. Adicionalmente, são tecidas considerações finais sobre as técnicas e abordagens utilizadas, assim como delineadas perspetivas futuras no contexto do diagnóstico de nódulos da tiroide.

### 1.4 Contribuição científica

O mérito científico deste estudo reside numa avaliação e comparação de duas técnicas de classificação de imagens ecográficas e de elastografia, com vista ao estudo e análise de nódulos da tiroide. Para tal, são utilizados algoritmos de *machine learning* e *deep learning* para classificar as imagens dos nódulos. Como resultado deste trabalho, foi realizada uma submissão à conferência RECPAD2023 (29ª Conferência Portuguesa de Reconhecimento de Padrões) (consultar anexo C).

## **1.5 Comissão de Ética**

O presente estudo foi submetido à análise e avaliação da Comissão de Ética do Instituto Politécnico de Coimbra. Após uma análise cuidadosa, a Comissão de Ética emitiu o Parecer N.º 88\_CEIPC/2023 que fundamenta a aprovação do estudo conduzido no âmbito deste trabalho, sendo essencial para a inclusão e apresentação dos resultados nesta dissertação de mestrado.

A decisão da Comissão de Ética do Instituto Politécnico de Coimbra foi deferida, por unanimidade, no dia 28 de junho de 2023. Esta decisão reflete o compromisso desta pesquisa em seguir os pressupostos éticos em todas as fases da recolha, análise e divulgação dos dados.

## **2 Conceitos teóricos**

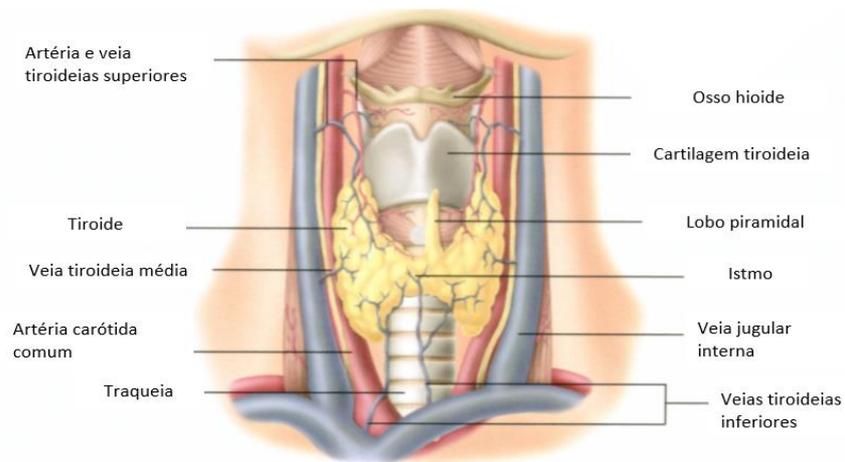
Neste capítulo, são explorados os princípios teóricos relativos à tiroide e às respectivas condições patológicas associadas frequentemente a esta glândula, com destaque para os nódulos da tiroide que são o foco principal deste estudo. As diversas abordagens imagiológicas utilizadas para analisar e diagnosticar os nódulos são também abordadas, com ênfase nas técnicas baseadas nos ultrassons, a ecografia e a elastografia, as quais desempenham um papel fundamental neste estudo.

### **2.1 A tiroide**

A tiroide é a glândula exclusivamente endócrina de maior dimensão, com um peso variável entre 20 e 34 gramas [1]. O tamanho da glândula tiroideia pode variar de pessoa para pessoa tendo em conta fatores como a hereditariedade, a nutrição e o ambiente [2].

Estruturalmente, a tiroide divide-se em dois lobos principais ligados através do istmo (pequeno filamento de tecido tiroideu). Cada lobo tem um comprimento de aproximadamente 4 a 7 centímetros [3].

O tecido tiroideu apresenta uma coloração mais intensa quando comparado com os tecidos envolventes devido à abundante irrigação sanguínea para suprir as necessidades das células glandulares. O aporte de sangue relativo a esta glândula é realizado através da artéria tiroideia superior (ramo da artéria carótida externa) e pela artéria tiroideia inferior (ramo do tronco tirocervical) (Fig. 2.1). Por outro lado, a drenagem sanguínea é realizada através das veias tiroideias médias e da veia tiroideia superior (conduzindo o sangue venoso até às veias jugulares internas) e pelas veias tiroideias inferiores, fazendo o sangue chegar às veias braquiocefálicas [2].



*Figura 2.1 - Anatomia da tiroide (vista anterior) (adaptado de [4])*

A tiroide encontra-se localizada na parte anterior do pescoço, sendo que os dois lobos constituintes se encontram posicionados lateralmente à porção superior e anterior da traqueia [5]. Devido à sua localização anatômica, a glândula da tiroide é facilmente palpável quando existe suspeita de alguma alteração da mesma, podendo até mesmo tornar-se mais saliente em relação ao que a rodeia [2].

A tiroide é constituída por diversos folículos ocos com forma esférica, cujo espaço que os separa é formado por tecido conjuntivo (tecido constituído por uma grande quantidade de células e fibras) bastante irrigado por capilares. Os folículos são formados por uma parede de epitélio simples cuboide com tirócitos T na sua constituição e a parte mais central (lúmen) é constituída por uma proteína designada tiroglobulina (Fig. 2.2) [2], [5]. Esta proteína é sintetizada pelas células foliculares e está intimamente ligada ao armazenamento das hormonas da tiroide [1]. A triiodotironina (também designada por  $T_3$ ) e a tetraiodotironina (conhecida igualmente como  $T_4$  ou tiroxina) são as principais hormonas produzidas pela tiroide (10% e 80%, respetivamente) e a sua produção depende da hormona tiro-estimulante (TSH) sintetizada pela adeno-hipófise. A adeno-hipófise corresponde ao segmento anterior da hipófise, uma glândula endócrina localizada na base do encéfalo.

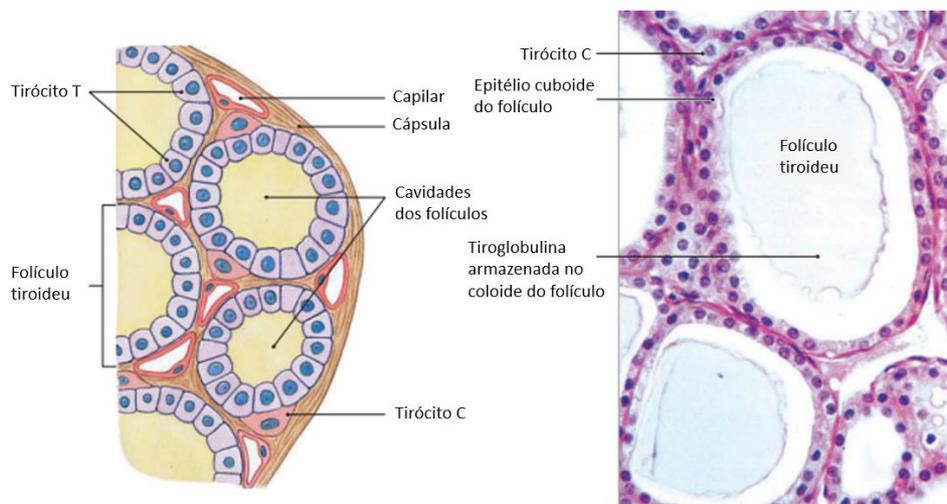


Figura 2.2 - Detalhes histológicos da glândula tiroideia (adaptado de [2])

Entre os folículos e as células que constituem as suas paredes estão localizadas as células parafoliculares (tirócitos C ou células C (Fig.2.2)) que produzem a calcitonina cuja função é reduzir a concentração de cálcio quando é necessário [1], [2].

A produção de calcitonina pela tiroide é contrariada pelas hormonas produzidas pela paratiroide. A produção destas hormonas contribui para incrementar a estimulação de osteoblastos e osteoclastos (células encontradas no tecido ósseo) de modo a aumentar os níveis de absorção de cálcio. Este aumento ocorre devido à redução das perdas de cálcio ao nível da urina e ao aumento da absorção no intestino delgado (na presença da hormona calcitriol produzida pelos rins) [2].

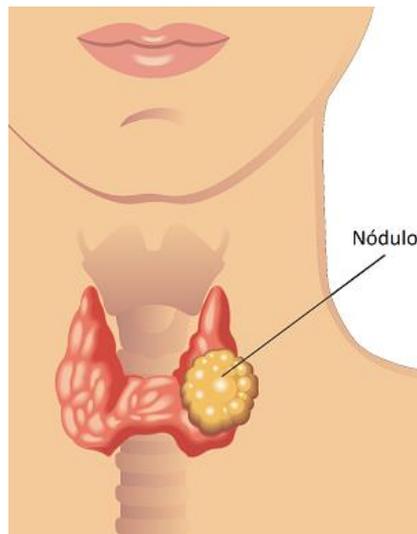
Em suma, estas hormonas (a calcitocina e as hormonas paratiroideias) atuam sobre o mecanismo do cálcio pelo que estão intimamente ligadas à deposição e fixação de minerais nos ossos e à regulação dos níveis de cálcio no corpo [5].

## 2.2 Doenças da tiroide

O diagnóstico de doenças associadas à tiroide pode prevenir consequências graves para a saúde. O hipertiroidismo, o hipotiroidismo e os nódulos constituem as principais patologias associadas à tiroide [5].

O hipertiroidismo consiste no excesso de produção de hormonas pela glândula da tiroide e pode originar alterações a nível do coração, do metabolismo (maior produção de calor, provocando transpiração excessiva) e do sistema nervoso (distúrbios psicológicos). Para além disso, o hipertiroidismo pode originar alguma vulnerabilidade muscular e perda de peso [1], [5]. Um dos exemplos mais comuns de hipertiroidismo é a doença de Graves (predominantemente no sexo feminino, meia-idade). Nesta patologia autoimune, o sistema imunitário produz anticorpos idênticos ao TSH o que provoca um aumento da produção de hormonas da tiroide. São sintomas desta doença uma anormal perda de peso, uma alta taxa metabólica, o aumento da transpiração, nervosismo excessivo ou alta frequência cardíaca. Em casos mais graves, pode ocorrer a projeção dos olhos devido à inflamação dos músculos e dos tecidos adjacentes ao olho [1], [5]. Pelo contrário, o hipotiroidismo representa uma condição de défice de produção e secreção hormonal. Esta patologia provoca intolerância ao frio, diminuição da síntese de proteínas, bradicardia, entre outras. Podem-se observar ainda distúrbios no sistema nervoso central com diminuição dos reflexos [5]. Uma das doenças provocadas pelo hipotiroidismo na idade adulta é o mixedema, que representa uma patologia autoimune caracterizada pela destruição do tecido da tiroide por parte de anticorpos. Por outro lado, quando o hipotiroidismo resulta de uma deficiente quantidade de iodo na alimentação, pode haver um aumento substancial da tiroide, provocando um inchaço na zona anterior do pescoço (bócio endémico). Nas crianças, o défice de produção das hormonas da tiroide prejudica o crescimento e desenvolvimento físico e mental [1], [2], [5].

Os nódulos da tiroide (Fig. 2.3) são comuns, tendo uma incidência em cerca de 60% da população geral. Estes nódulos são comumente lesões esféricas e podem apresentar três configurações: nódulos sólidos, líquidos ou mistos. Os nódulos benignos podem dividir-se em diferentes categorias: nódulos macro foliculares (sólidos), adenomas coloides (líquidos), quistos complexos (mistos), bócios e tiroidites de Hashimoto. Os nódulos benignos são, usualmente, monitorizados através da ecografia. Estes também podem produzir hormonas levando a condições de hipertiroidismo pelo que devem ser tratados, quando necessário, com iodo radioativo [1], [2], [5].



*Figura 2.3 - Representação de um nódulo da tireoide (adaptado de [6])*

Quanto aos nódulos malignos, identificam-se os carcinomas papilar, folicular, medular ou anaplásico e o linfoma da tireoide como as principais patologias. Em geral, quando se verifica a existência deste tipo de nódulos recorre-se à cirurgia para a sua extração ou até mesmo à remoção de toda a glândula. O tratamento com iodo deve também ser utilizado para eliminação de células tumorais residuais ainda existentes [7]. Normalmente, numa fase inicial os nódulos não provocam sintomas, pelo que a deteção ocorre quando são nódulos grandes, suscetíveis de serem visíveis ou palpáveis. Quando aparecem, os sintomas provocados pelos nódulos da tireoide consistem na sensação de estreitamento do pescoço, rouquidão, tosse, disfagia, dificuldade no processo de respiração, dores de pescoço e/ou ouvidos e, em casos extremos, os pacientes podem apresentar o sinal de Pemberton, ou seja, verifica-se um congestionamento sanguíneo e respiratório quando o paciente eleva os braços acima da cabeça [8]. Utilizam-se os exames de ultrassonografia para confirmar a presença de um nódulo na tireoide. Existe também a possibilidade de realizar uma biópsia dependendo de uma avaliação prévia por parte do profissional de saúde, tendo em consideração o tamanho do nódulo, a existência de características suspeitas nas imagens, os níveis de TSH, entre outros fatores clínicos [7].

## 2.3 Técnicas imagiológicas

Os nódulos da tiroide são lesões discretas que ocorrem no interior da glândula e que podem ser detetadas num exame de rotina à tiroide. Existem diversas modalidades de imagiologia médica que permitem a deteção desses nódulos. A ecografia e, mais recentemente, a elastografia apresentam-se como as técnicas de imagem mais utilizadas na deteção de nódulos da tiroide. No entanto, técnicas como a Tomografia Computorizada (TAC), a Ressonância Magnética (RM) ou a Tomografia por Emissão de Positrões (PET) são também utilizadas na deteção de nódulos mais complexos [7].

### 2.3.1 Ecografia

O desenvolvimento tecnológico das últimas décadas no campo da eletrónica e do processamento de imagem, conduziu a uma grande expansão e intensiva aplicação dos ultrassons no diagnóstico médico. Apresenta diversas vantagens em relação a outros métodos de imagem, pelo facto de ser um método rápido, pouco dispendioso, cómodo e de fácil utilização. Para além disso, é um método não invasivo e inócuo, não necessitando de instalações específicas com condições de proteção radiológica [9].

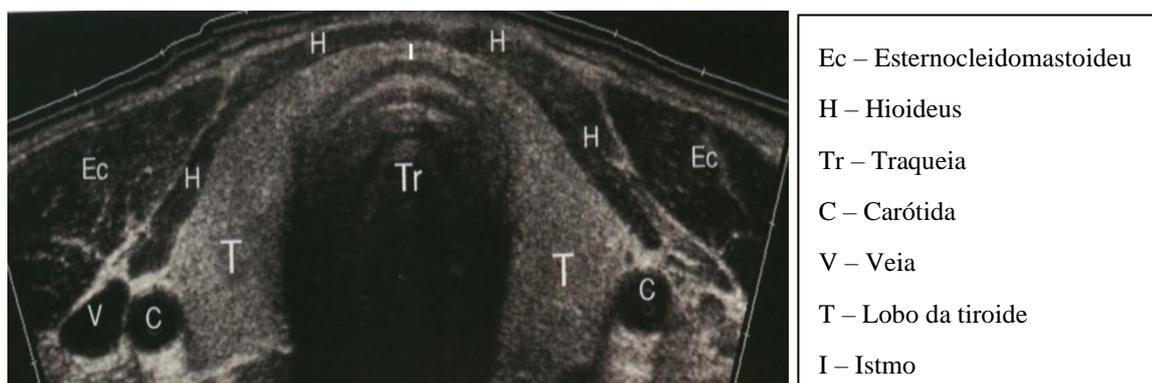


Figura 2.4 - Imagem ecográfica da tiroide (vista transversal) com legenda (adaptado de [9])

A ecografia permite obter imagens do corpo humano com boa resolução, em especial, de tecidos moles, como é exemplo a imagem ecográfica da tiroide ilustrada na Figura 2.4.

Esta faz uso de ondas acústicas que se propagam ao longo do corpo humano, refletem nas fronteiras dos tecidos e regressam ao emissor na forma de ecos. Conhecendo a velocidade de propagação das ondas acústicas no meio e o tempo de propagação das mesmas, é possível calcular a profundidade das fronteiras que originaram as reflexões [9].

No diagnóstico médico, são usadas ondas com frequências entre os 3 MHz e os 50 MHz.

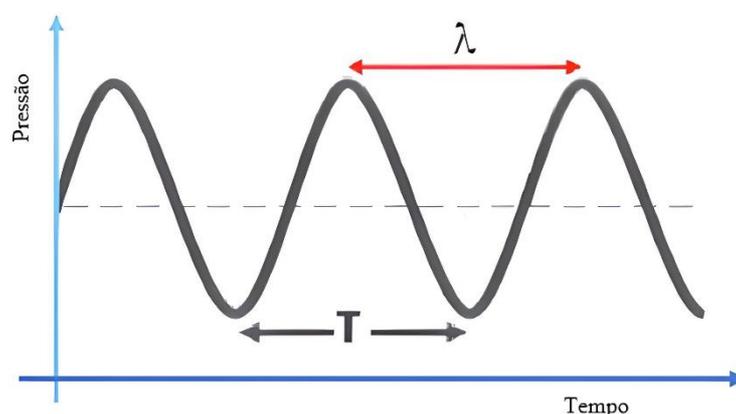


Figura 2.5 - Representação da onda acústica num referencial

Uma onda acústica comporta-se como uma onda sinusoidal, podendo ser representada num referencial bidimensional (Fig. 2.5). O período ( $T$ ), a frequência ( $f$ ) (inverso do período, representa o número de oscilações por segundo, em *Hertz*) e o comprimento de onda ( $\lambda$ ) (distância entre dois pontos correspondentes da onda sinusoidal) são características das ondas acústicas.

Existem ondas longitudinais (ou de compressão) e ondas transversais (ou *shear*). Assim, no caso das ondas longitudinais, o movimento das partículas ocorre na mesma direção da propagação da onda (no caso dos fluidos e tecidos, por exemplo), enquanto nas ondas transversais ocorre perpendicularmente à direção da propagação [9].

Os dispositivos de ultrassons são complexos e englobam três componentes básicos: um gerador-recetor de sinais, um transdutor (sonda) e um dispositivo para visualização dos sinais/ imagens.

O gerador-recetor tem como função a excitação da sonda, através do envio de um pulso elétrico. Os ecos originados ao longo do meio de propagação são recebidos e amplificados pelo mesmo dispositivo, agora em modo receção.

O transdutor (sonda) converte a energia elétrica em energia mecânica e vice-versa. Os transdutores criam as vibrações mecânicas através do efeito piezoelétrico. Ao aplicar uma diferença de potencial ao transdutor, este gera ondas acústicas e, em seguida, converte os ecos provenientes das reflexões nas fronteiras em sinais elétricos [9], [10].

A visualização dos sinais pode ser efetuada recorrendo a diversos modos de exibição de imagem. O modo A constitui uma representação da amplitude do sinal em função do tempo (profundidade). Este modo de visualização é usado em oftalmologia. O modo B, resulta do uso de um transdutor multi-elemento que forma uma imagem em tempo real que representa as estruturas numa escala de brilho, ou seja, as estruturas mais refletivas apresentam maior brilho. O modo M, associado ao modo B, exhibe simultaneamente a amplitude do sinal na forma de brilho bem como o movimento ao longo do tempo. Este modo permite analisar o movimento das fronteiras pelo que é usado em cardiologia, especialmente na análise da frequência cardíaca de fetos, na determinação da dimensão das cavidades cardíacas e avaliação das respetivas válvulas [9]. Existe ainda o modo *Doppler* através do qual se consegue adquirir informações sobre o fluxo sanguíneo. Por último, a elastografia, que mostra dados qualitativos acerca da elasticidade dos tecidos através da medição de propriedades mecânicas dos mesmos, inferindo acerca da existência de um tecido maligno ou inflamado [11].

As imagens ultrassonográficas são usadas frequentemente quando se suspeita da existência de nódulos da tiroide. Assim, obtém-se informação sobre as propriedades dos nódulos, sobre a anatomia da tiroide e respetivas estruturas [7].

### **2.3.2 Elastografia**

A elastografia é uma modalidade de imagiologia por ultrassons recente e em expansão. Beneficia de um *software* avançado que permite avaliar a elasticidade dos tecidos do corpo humano, fornecendo informações importantes para diagnósticos médicos. Desta forma, determina a existência de tumores ou inflamações que alteram as propriedades estruturais dos tecidos e, conseqüentemente, a rigidez dos mesmos. Estas informações são exibidas em imagens em escalas de cores [9], [10].

A elastografia possibilita a identificação de áreas rígidas anormais, ou seja, permite substituir a palpação convencional, ajudando a analisar certos nódulos de tamanho reduzido e/ou difícil acesso no exame físico.

A elastografia apresenta duas vertentes diferentes para o estudo do comportamento dos tecidos. Uma delas é a elastografia de deformação, caracterizada pela medição do deslocamento dos tecidos. Para realizar a elastografia de deformação, é necessário posicionar o transdutor de modo a aplicar uma pressão controlada sobre o tecido. Isso permite a captura de leituras antes e depois da compressão. Uma vez que a pressão efetuada de forma manual não pode ser medida diretamente, os níveis de deformação têm de ser comparados com os tecidos circundantes. Isto significa que, a rigidez do tecido em análise é apresentada relativamente a outros tecidos saudáveis. A outra técnica é a elastografia por onda transversal ou *shear* (SWE), onde o transdutor gera pulsos que resultam na formação de ondas transversais que interferem continuamente com os tecidos. Estas ondas são analisadas e acompanhadas através de imagens de alta resolução proporcionando um bom nível de detalhe na avaliação da elasticidade dos tecidos [9].

### **2.3.3 Tomografia computadorizada (TAC)**

A tomografia computadorizada, também conhecida como TAC ou TC, é uma técnica baseada na utilização de raios-X que permite adquirir imagens bastante detalhadas do corpo humano (pulmões, tecidos moles, esqueleto, coração, cérebro, etc.). Este método é usado no diagnóstico e no seguimento de certas patologias e como ferramenta de auxílio a outras técnicas (radioterapia).

A TAC apresenta uma fonte de raios-X que roda em torno do paciente. Através de milhares de detetores é medida a transmissão de raios-X em diversos planos, conhecidos como planos de *gantry*. Assim, através do pós-processamento da informação bidimensional (2D) adquirida em vários planos é possível formar uma imagem tridimensional (3D) (a cores ou em escala de cinza) e, se necessário e pertinente, podem também ser vistas as imagens 2D (cortes) [12].

#### **2.3.4 Ressonância magnética (RM)**

A ressonância magnética constitui uma técnica não invasiva, indolor e desprovida de radiação ionizante, que funciona através de ondas de rádio associadas a um campo magnético. A ressonância magnética baseia-se na interação entre partículas (com *spin* e carga) e um campo magnético produzido pelo equipamento. Este campo magnético faz com que os prótons localizados nos tecidos se alinhem. Ao fazer-se incidir um pulso de ondas rádio nos mesmos esse alinhamento é destabilizado, havendo a libertação de energia que é detetada pelo aparelho na forma de sinais.

É uma técnica muito utilizada na imagiologia de diagnóstico médico e permite a visualização de órgãos e tecidos em vários planos de modo a diagnosticar e acompanhar patologias principalmente associadas ao sistema nervoso central, articulações e tecidos moles. As imagens obtidas são de alta resolução e podem ser secções verticais ou horizontais da área em estudo [10], [13].

#### **2.3.5 Tomografia por Emissão de Positrões (PET)**

A PET é uma técnica que permite diagnosticar diversas doenças a nível oncológico, cardiológico e neurológico. Esta técnica permite analisar a atividade bioquímica das células e, conseqüentemente, o funcionamento de tecidos e órgãos. Para tal, são injetados componentes radioativos no corpo do paciente acoplados a uma molécula biológica ativa (ex.: glicose) de modo a serem transportados, habitualmente, ao longo da circulação sanguínea.

Assim, decorrido o tempo necessário após a injeção do material radioativo, é feito um varrimento de modo a obter imagens da distribuição do componente radioativo ao longo do corpo. Esta técnica está muitas vezes associada à TAC [14].

### 3 Estado da arte

A incidência de nódulos da tireoide tem aumentado e o uso de técnicas imagiológicas baseadas nos ultrassons é muito importante para a sua detecção e classificação. Nos últimos anos, têm sido apresentados diversos trabalhos acerca da identificação e classificação dos nódulos da tireoide. Estes estudos têm como objetivo aprimorar progressivamente as diversas abordagens associadas às técnicas de ultrassonografia utilizadas na prática clínica, como é o caso da ecografia e, mais recentemente, da elastografia.

Na literatura, constata-se que os autores empregam diferentes metodologias para proceder ao reconhecimento e classificação dos nódulos da tireoide, através de técnicas de processamento de imagem que servem de base para os algoritmos de *machine learning* que fazem uso de classificadores tradicionais (*Support Vector Machine* (SVM), por exemplo) para conceber os modelos de classificação. Por outro lado, existem estudos nos quais são aplicadas metodologias modernas baseadas em algoritmos de *deep learning* associados a redes neurais para efetuar a análise das imagens e o processamento da informação. Para além disso, existem ainda estudos que utilizam uma metodologia híbrida resultante da conjugação dos dois tipos de metodologias anteriormente referidas.

#### 3.1 Métodos tradicionais

Um dos desenvolvimentos, com base em métodos tradicionais, foi o sistema CAD (*Computer-Aided Design*) que apenas utiliza características independentes da direção (não dependentes da orientação nem do ângulo de inclinação da sonda de ultrassons) [15]. A metodologia inclui uma *Two-Threshold Binary Decomposition* que transforma as imagens num conjunto de imagens binárias para extrair cerca de 70 *features* relevantes (como por exemplo: *ConvexArea*, *EquivDiameter*, *Filled Area*) utilizadas em classificadores como *Random Forest* (RF) e SVM, com *10-fold cross-validation*, para proceder à classificação dos nódulos. Neste estudo, Prochazka et al. [15] utilizaram 60 imagens de nódulos da tireoide (40 benignos e 20 malignos) guardadas num PACS (sistema de arquivo e comunicação de imagens). Os resultados obtidos para o classificador RF e para o classificador SVM no que diz respeito à exatidão foram 95% e 91,6%, respetivamente. O

sistema proposto permite ultrapassar uma limitação dos sistemas CAD comumente utilizados, os quais apenas eram capazes de analisar imagens estáticas num único plano axial ou longitudinal, enquanto o sistema proposto pode ser aplicado a ambos os planos.

Zhao et al. [16] apresentaram uma forma de incrementar o desempenho da diferenciação dos nódulos da tiroide através de *machine learning*, de modo a reduzir o número de biópsias desnecessárias. Foram implementados e comparados dois métodos distintos de *machine learning*: uma abordagem visual e uma abordagem tradicional de extração intrínseca de *features* das imagens. Para a primeira foram recolhidos 11 parâmetros de ambos os tipos de imagens (ex.: forma, composição, média da elasticidade) por dois radiologistas experientes e, para a segunda, foram extraídas as *features* das imagens de ecografia e elastografia. Foram usados classificadores tais como *Decision Tree* (DT), *Naive Bayes*, KNN (*k-nearest neighbors*), SVM, e RF. Para o estudo foram utilizadas cerca de 850 imagens de nódulos da tiroide de 822 pacientes. A abordagem visual obteve sempre melhores resultados de exatidão (81,6% (US) e 88,8% (US+SWE)) comparativamente à abordagem tradicional (73,1% (US) e 76,2% (US+SWE)).

### 3.2 Métodos recentes

Qin et al. [17] propuseram uma combinação das *features* presentes nas imagens recolhidas através da ecografia e da elastografia para o diagnóstico dos nódulos benignos e malignos da tiroide. Para tal, uma CNN foi pré-treinada na *ImageNet* e os parâmetros necessários para o processamento das imagens ecográficas foram obtidos por *transfer learning*. Esta informação foi posteriormente combinada com as imagens obtidas por elastografia. Para o estudo foram analisadas 1156 imagens de nódulos da tiroide provenientes de 233 pacientes, dos quais 539 eram benignos e 617 malignos tendo-se obtido no estudo uma exatidão de 94,7%. O método apresentado, segundo os autores, atinge melhores resultados comparativamente a métodos que se baseiam em *features* recolhidas exclusivamente através de uma única modalidade, ou seja, a combinação da ecografia com a elastografia permite aumentar a viabilidade da classificação dos nódulos da tiroide quando comparado com o uso exclusivo da ecografia.

Feigin et al. [18] propôs uma nova abordagem que se baseia nas propriedades dos tecidos e que visa superar as limitações das técnicas de elastografia com o objetivo de melhorar a

capacidade de detecção de patologias associadas à tiroide. Assim, uma CNN serve de alicerce ao método alternativo apresentado, o qual se fundamenta na inversão da velocidade longitudinal do som para tecidos moles. Em condições normais, a velocidade do som nos tecidos rígidos é superior à velocidade do som em tecidos moles. No entanto, em certas situações patológicas, a velocidade do som nas lesões ou tumores tende a superiorizar-se relativamente aos tecidos rígidos circundantes. Para o estudo, recorreram a dados obtidos através de simulações. Foram criadas aleatoriamente 6026 amostras de treino e 800 amostras de teste para desenvolver a CNN. Essas amostras simulam dados de elastografia para diferentes tipos de tecidos. Os autores, através da utilização da tecnologia de *deep learning* atingiram, em tempo real, graus de exatidão superiores ao esperado admitindo, no entanto, a necessidade de melhorias no desenvolvimento, calibração, validação e treino.

Wildeboer et al. [19] calcularam a velocidade de uma onda *shear* para determinar o módulo de Young (módulo de elasticidade). Desta forma, criaram imagens sintéticas de elastografia por ondas *shear* (sSWE) através de um equipamento capaz de obter imagens ecográficas de modo B, mas sem a capacidade para obter imagens SWE. As imagens sSWE foram geradas utilizando CNN profunda. Para além disso, os investigadores tentaram aplicar o método sSWE a diversos órgãos a partir do estudo de imagens SWE de nódulos da tiroide obtidas a partir de 215 pacientes. Os autores acreditam que esta técnica baseada nas *features* do modo B (como textura, forma, entre muitas outras) pode ser pioneira na caracterização de tecidos e das respetivas propriedades mecânicas sem necessidade de recorrer às complexas imagens obtidas por SWE, promovendo a criação e uso de sistemas de aquisição de imagens mais simples e baratos. A necessidade de treinar a CNN para diferentes órgãos e de uma certa padronização das configurações de imagem são as principais limitações desta técnica.

Foi desenvolvida por Xiang et al. [20] uma rede de difusão multimodal formada por imagens obtidas através de mais do que uma técnica ultrassonográfica, suportada por um modelo de *deep learning*, de forma a automatizar o diagnóstico clínico de nódulos benignos e malignos. Para tal, são utilizadas três modalidades: a ecografia, a elastografia de ondas *shear* e o efeito *doppler* colorido. Este último permite avaliar a irrigação sanguínea dos nódulos. Inicialmente, foram utilizadas três ResNet18s para extrair as *features* das imagens de cada modalidade e, de seguida, foram removidas as informações

comuns das três modalidades. As imagens foram posteriormente fundidas para se proceder à classificação dos nódulos da tiroide. Foram utilizadas 486 imagens de nódulos da tiroide (277 benignos e 209 malignos). Este método, alcançou resultados (precisão = 92,54%, exatidão = 89,79% e *F1-Score* = 88,65%) superiores comparativamente a outros métodos que recorrem a um menor número de modalidades para obter as imagens sendo, segundo os autores, um método rápido e preciso no diagnóstico dos nódulos da tiroide.

Hu et al. [21] criaram um método de diagnóstico do cancro da tiroide baseado em imagens de ecografia e de elastografia de ondas *shear* aplicada a regiões periféricas dos nódulos. Para além das imagens, foram recolhidos parâmetros quantitativos SWE como é o caso da elasticidade máxima dos nódulos e a média da elasticidade máxima de cinco pontos. Diversos modelos CNN foram desenvolvidos através da ResNet18. Foram utilizados 1747 nódulos da tiroide (de 1582 pacientes), dos quais 1247 foram usados para treinar o sistema e os restantes 500 para validação do mesmo.

O método CNN baseado nas imagens US+SWE obteve bons resultados para diferentes tamanhos de nódulos para a métrica AUC (*Area Under the ROC Curve*), demonstrando melhorias na previsão do cancro da tiroide.

### 3.3 Métodos híbridos

O trabalho realizado por Wang et al. [22] avaliou dois métodos de deteção e avaliação dos nódulos da tiroide através do uso de imagens obtidas pela ecografia, tendo comparado um método tradicional de classificação com um método de *deep learning* alcançado através de uma CNN. No processo de *machine learning* foram extraídas *features* comumente usadas, aplicando funções como a análise de discriminação linear (LDA) para realizar a seleção das *features* mais adequadas para o estudo. Posteriormente, um classificador SVM é aplicado para identificar e distinguir os nódulos benignos e malignos. Para a abordagem de *deep learning*, a CNN utilizada foi a VGG16, pré-treinada na base de dados *ImageNET* com 1,2 milhões de imagens. A arquitetura original da rede foi modificada, reduzindo de 13 para 5 as camadas de convolução, mantendo as 5 camadas de *max-pooling* e diminuindo de 3 para 2 as camadas totalmente conectadas que se comportam como camadas classificadoras. O conjunto de dados era constituído por 3120 nódulos da tiroide

(de 1040 pacientes). Dos resultados obtidos, salienta-se a exatidão de classificação de 66,8% para o método tradicional e, para *o deep learning*, atingiu-se 74,4% nas amostras de teste.

Sun et al. [23] introduziram uma técnica que permite efetuar a distinção entre nódulos benignos e nódulos malignos através da fusão de dados recolhidos através do modo B e da técnica de elastografia por onda *shear*. Posteriormente, foram utilizados classificadores (SVM, por exemplo) para classificar os dados com base nas *features* extraídas das diferentes imagens, obtidas através de uma CNN (VGG16). Assim, as *features* recolhidas através da imagem do modo B e as *features* obtidas através da elastografia por onda *shear* são analisadas e comparadas pelos diferentes classificadores. Posteriormente, essas informações são fundidas e combinadas de modo a obter uma imagem mais completa. Segundo os investigadores, os resultados demonstram que o método apresentado permite uma quantificação e diferenciação eficaz dos nódulos da tiroide, alcançando uma taxa de exatidão de 86,5%.

O método proposto por Yang et al. [24] permite contornar algumas limitações do *deep learning*. Os autores propuseram um *multi-task cascade deep learning model* (MCDLM) que permite a identificação automática dos nódulos da tiroide através de imagens ultrassonográficas multimodais (obtidas através de imagens de ecografia e de imagens de elastografia). A informação de interesse passível de ser analisada para identificar os nódulos é obtida através da *U-net*. Foram quantificadas as *features* das imagens dos nódulos, de modo que, em conjunto com uma DScGAN fossem obtidas imagens com maior qualidade e com melhores atributos diferenciadores/discriminatórios. Seguidamente, um SVM foi treinado para classificar os nódulos da tiroide. A informação utilizada corresponde a 3090 imagens com nódulos de 1090 pacientes com idade superior a dezoito anos. Das 3090 imagens, 1489 correspondem a nódulos malignos e as restantes a nódulos benignos. O método utilizado foi aplicado a cerca de 35% das imagens inicialmente recolhidas e apresenta resultados de exatidão de 90,01% e AUC de 91,07%. Os investigadores concluíram que o modelo MCDLM permite alcançar bons resultados em imagens não etiquetadas, ou seja, imagens que não foram marcadas previa e manualmente pelos radiologistas. Tais resultados devem-se, segundo os autores, ao facto de ter sido implementada uma segmentação automática e posterior extração das *features* necessárias.

### 3.4 Bases de dados

Os artigos apresentados anteriormente recorrem a bases de dados privadas, maioritariamente a dados e imagens obtidas em hospitais locais. No entanto, existem algumas bases de dados públicas que podem ser consultadas. A DDTI (*Digital Database of Thyroid Ultrasound Images*) consiste numa base de dados publicada por Pedraza et al. [25], que contém imagens ecográficas acompanhadas de um diagnóstico efetuado por radiologistas e confirmado através de biópsias. A base de dados 3DThyroid, disponibilizada por Wunderling et al. [26] é também de acesso livre e apresenta imagens 3D das quais se podem retirar imagens 2D. Existe ainda a base de dados pública TN3K publicada por Gong et al. [27] que é composta por imagens de nódulos da tiroide de cerca de 2400 pacientes. Algumas informações relativas às bases de dados referenciadas são apresentadas na Tabela 3.1.

Tabela 3.1 – Informações sobre as bases de dados públicas

Nome	Ano	Nº de pacientes	Nº de imagens	Patologias	Equipamento utilizado	Resolução	Artigos
DDTI [25]	2015	270 mulheres 29 homens	400	Tiroidites, nódulos císticos, adenomas, cancro da tiroide	TOSHIBA Nemio 30	560 x 360 píxeis	[28],[29],[30], [31]
3DThyroid [26]	2017	-----	16	Nódulos da tiroide	GE Logiq E9 XDclear 2.0	240 x 120 píxeis	[31]
TN3K [27]	2021	2421 pacientes	3493	Nódulos da tiroide	-----	-----	[32]

## 4 Metodologia

*Machine learning* e *deep learning* são duas abordagens que possibilitam a resolução de problemas com elevado grau de complexidade e que processam uma grande quantidade de dados, padrões e informações. *Machine learning* é uma técnica mais simples, necessita de menos dados e a extração de *features* tem de ser previamente programada de forma manual. Por outro lado, *deep learning* tem por base o uso de CNNs e processos de extração de *features* automatizados, sendo utilizada em problemas de maior complexidade, associados a uma grande quantidade de dados. Neste capítulo, são abordados os diferentes passos da metodologia associada a ambas as técnicas, desde o pré-processamento dos dados recolhidos à escolha dos algoritmos e arquiteturas que melhor se adaptam, tendo como objetivo a classificação e avaliação dos resultados através de diferentes métricas. Parte da metodologia utilizada assenta na revisão da literatura apresentada no capítulo anterior (Estado da Arte). No entanto, a maioria dos estudos apresentados foca-se na classificação de nódulos benignos e malignos, ao passo que este estudo se centra na classificação das imagens quanto à presença ou ausência de nódulos.

### 4.1 *Machine Learning*

#### 4.1.1 Pré-processamento

O pré-processamento é uma etapa de enorme importância para que o modelo implementado alcance os melhores resultados. As imagens foram pré-processadas em três etapas distintas. Inicialmente, foi aplicada uma janela deslizante para aumentar de forma artificial o número de imagens. Seguidamente, as imagens passaram por processos de transformação para escala de cinzentos e, posteriormente, foram objeto de normalização.

A janela deslizante, é uma técnica utilizada no processamento computacional de imagens e permite que uma imagem seja percorrida através de uma janela com dimensões (comprimento e largura) escolhidas previamente. Para além disso, é possível definir o tamanho do passo, ou seja, a distância entre cada janela, o que corresponde ao número de píxeis de deslocamento entre cada janela. Quanto mais pequeno é o tamanho do passo, maior é a sobreposição das janelas e, conseqüentemente, das imagens obtidas. Pelo

contrário, um tamanho do passo maior corresponde a uma menor sobreposição ou à inexistência de sobreposição. Neste último caso, haverá regiões da imagem original que não estarão nas imagens posteriormente obtidas. Na Figura 4.1, ilustra-se o procedimento da janela deslizante.

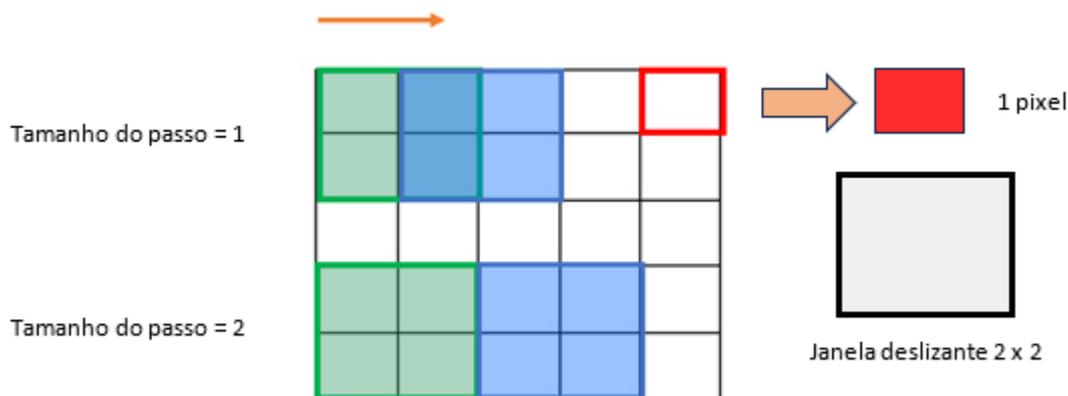


Figura 4.1 - Esquema representativo do funcionamento da janela deslizante

Esta técnica é frequentemente utilizada em diversas funções associadas a imagens, como na detecção de objetos e reconhecimento facial [33], [34]. No entanto, neste caso foi utilizada para aumentar o número de imagens (dados) para posterior análise. Deste modo, a cada ROI (região de interesse) selecionada pelo profissional de saúde (a partir da imagem original) foi aplicada uma janela deslizante, como ilustrado na Figura 4.2, proporcionando um incremento significativo do número de imagens para classificação.

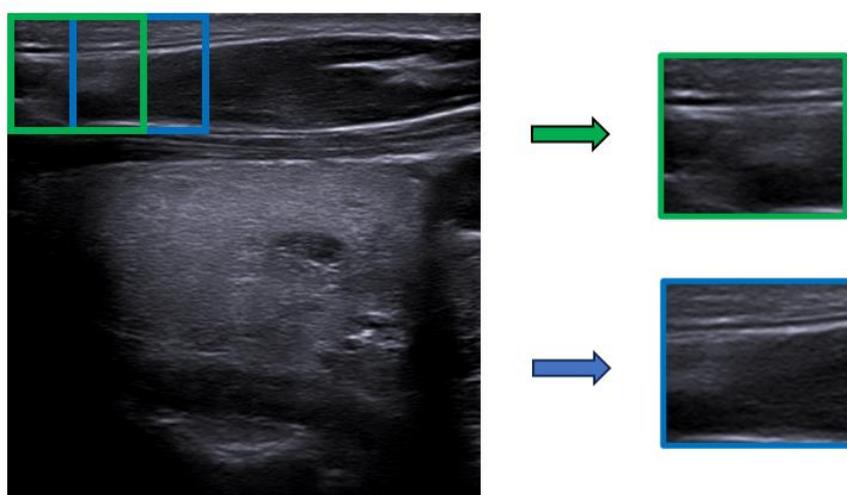


Figura 4.2 - Aplicação da janela deslizante a uma imagem ecográfica da tireoide

Apesar de visualmente as imagens serem em tons de cinzentos, o sistema de aquisição produz sempre uma imagem em RGB (três canais), sendo necessário converter computacionalmente essa imagem para a escala de cinzentos (dois canais).

Posteriormente, as imagens obtidas são sujeitas a uma normalização da intensidade de cada píxel, redimensionando os valores dos píxeis para uma escala de 0 a 1. Este passo é fundamental no pré-processamento de imagens de fontes distintas, evitando-se, neste caso, discrepâncias de intensidades entre as imagens obtidas pelo profissional e as imagens provenientes de bases de dados existentes, assegurando-se, assim, que as diferentes imagens apresentam escalas comparáveis.

#### **4.1.2 Processamento**

Após o pré-processamento das imagens segue-se o processo de extração das *features* das imagens. As *features* são atributos específicos das imagens e que contêm informações relevantes que são fornecidas ao algoritmo de *machine learning*, de modo que o modelo seja capaz de distinguir entre diferentes classes.

Após a extração de todas as *features*, estas são normalizadas e, de seguida, é executada a seleção das melhores *features* do conjunto de dados. As funções que desempenham este processo executam testes estatísticos univariados, ou seja, cada *feature* é analisada individualmente com o intuito de se perceber qual o seu possível contributo para a tarefa a realizar. Assim, reduz-se a dimensionalidade do conjunto de dados, melhorando o desempenho do modelo [35]. A seleção de *features* é uma etapa importante do processamento, excluindo as *features* mais irrelevantes do conjunto de dados e, desta forma, evitando problemas de *overfitting* que ocorrem quando o modelo se ajusta em demasia aos dados de treino.

### 4.1.3 Classificação

As *features* derivadas do processamento realizado, de acordo com o exposto anteriormente, são analisadas por diferentes classificadores para posterior comparação de resultados, no que se refere à classificação das imagens (com ou sem nódulo).

Os algoritmos de classificação supervisionada, implementados neste trabalho, foram o SVM, *Random Forest*, *Decision Tree* e KNN. Estes algoritmos são frequentemente utilizados em problemas de classificação de modo a prever, com base nas características e atributos, a categoria a que determinados dados pertencem.

O SVM é comumente usado em problemas de classificação supervisionada de forma a encontrar o hiperplano que permite dividir os diferentes tipos de dados em grupos bem definidos. A *Decision Tree* (árvore de decisão) é uma técnica eficaz que permite representar esquemática e graficamente todas as decisões possíveis e respectivas respostas. A árvore é composta por nós de raiz (*features*), de ramo (decisão) e de folha (resultado). Este classificador divide os conjuntos de entrada surgindo pequenas árvores (subconjuntos) dentro da árvore principal passíveis de serem submetidos ao processo de classificação. [36]. O *Random Forest*, recorre a várias árvores de decisão formadas a partir de vários conjuntos aleatórios de dados de treino e de subconjuntos de variáveis de previsão. Assim, para implementar e criar os modelos de previsão, este método associa cada variável a um valor (alto ou baixo) de modo a obter os resultados [37]. O KNN é um modelo de classificação que permite fazer uma rápida divisão e categorização dos dados tendo em conta as semelhanças dos mesmos com os seus vizinhos. Este avalia a distância (medida que se refere ao quão similares ou diferentes são dois pontos de dados) entre cada um dos dados de entrada e o conjunto total de dados. Para tal, armazena um conjunto de treino num espaço n-dimensional para comparar com o novo dado de entrada (conjunto de teste), selecionando os K exemplos mais próximos deste e, desta forma, colocando-os na mesma classe [36].

#### 4.1.3.1 *k-Fold-cross-validation*

O *cross-validation* (ou validação cruzada) é um método cuja utilização garante que o mesmo grupo de dados não é usado como conjunto de treino e como conjunto de teste, levando à generalização eficaz do modelo, para todo o conjunto de dados, sem exceção. A implementação do *cross-validation* é essencial nos casos em que os dados são obtidos a partir de diferentes sujeitos, havendo mais do que uma amostra por sujeito. Assim, se ao invés de se utilizar o *cross-validation*, se efetuar uma divisão padrão dos dados, existe uma probabilidade considerável do modelo ser treinado com demasiadas *features* dos dados de um único sujeito [15].

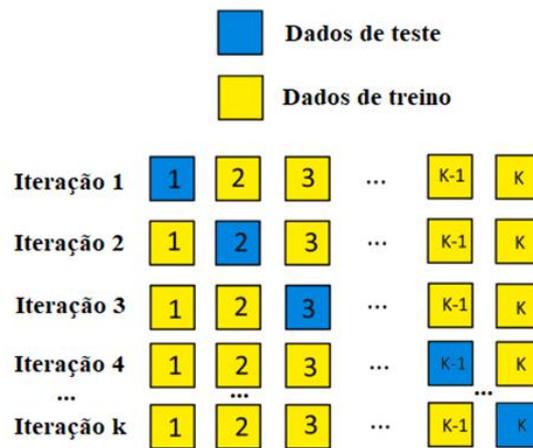


Figura 4.3 - Funcionamento do *k-Fold-Cross-Validation* (adaptado de [38])

O *k-Fold-cross-validation*, baseia-se na divisão do conjunto de dados em *k folds* (partes). O modelo é então treinado e testado com várias combinações de todos esses subconjuntos, garantindo um modelo de classificação mais robusto [38]. A Figura 4.3, ilustra o modo de aplicação do método.

## 4.2 *Deep Learning*

### 4.2.1 Pré-processamento

A complexidade das redes neurais utilizadas em *deep learning* e as suas variadas configurações, requerem a existência de um grande conjunto de dados de treino. No entanto, quando se aplica este método a problemas médicos, o tamanho do conjunto de dados torna-se um problema devido à insuficiência de imagens. Para contornar esta adversidade, as abordagens *transfer learning* e aumento de dados (também conhecido como *data augmentation*) são duas opções viáveis [17].

#### 4.2.1.1 Aumento de dados

O aumento de dados é utilizado para ultrapassar situações de défice de dados de treino de modo a incrementar a capacidade e o desempenho do modelo. O aumento de dados tem por base transformações simples ao nível das imagens como, por exemplo, espelhamentos, rotações, adição de ruído gaussiano (cria variações na intensidade dos píxeis), inversões, entre outros [17]. A Figura 4.4, mostra alguns exemplos de incremento de dados, através da aplicação de rotações e inversões.

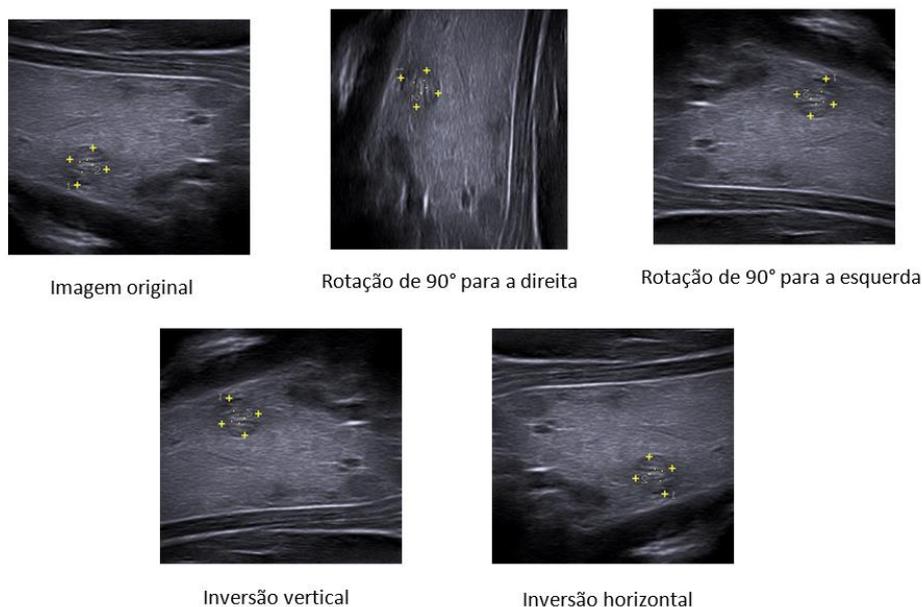


Figura 4.4 - Aumento de dados através de rotações e inversões da imagem original

Neste estudo, este método foi aplicado às imagens originais, com o intuito de aumentar o número das mesmas e, conseqüentemente, enriquecer o conjunto de dados.

#### 4.2.1.2 *Transfer Learning*

Quando os recursos computacionais ou o conjunto de dados são insuficientes para que o treino do modelo de *deep learning* se processe de forma viável, o *transfer learning* é uma opção que permite melhorar os resultados através do conhecimento prévio adquirido relativamente a determinada tarefa [17]. Este método possibilita o uso de um modelo pré-treinado para uma tarefa mais geral, para posterior aplicação numa tarefa mais específica. Deste modo, ao invés do modelo ser treinado do zero com parâmetros aleatórios, recorre-se a um modelo já existente e que contém logo à partida uma grande quantidade de dados, sendo depois ajustado à tarefa em causa. A Figura 4.5 ilustra esquematicamente o funcionamento do *transfer learning*. A adaptação do modelo pré-treinado obtido por *transfer learning* é efetuada frequentemente através do *fine-tuning*, que ajusta as propriedades e características do modelo pré-treinado ao novo modelo mais específico. Ao recorrer ao *transfer learning* consegue-se poupar tempo e recursos computacionais e, simultaneamente, obter um modelo mais versátil e aprimorado ao nível do desempenho. Uma das formas mais usuais para pré-treinar um modelo é através da *ImageNet*, que é uma base de dados com milhões de imagens de milhares de categorias distintas [39].

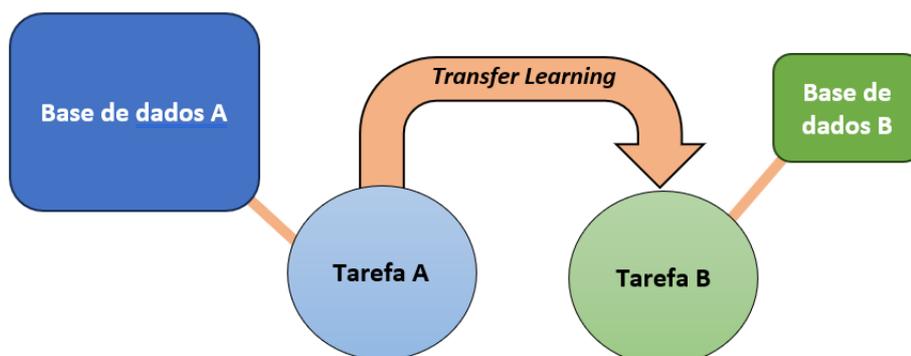


Figura 4.5 - Esquema representativo do uso do *Transfer Learning*

#### 4.2.2 Processamento

As CNNs, uma técnica popular de *deep learning*, têm sido cada vez mais utilizadas no âmbito médico e destacam-se numa variedade de tarefas das quais se salientam a deteção de objetos e a classificação de imagens. As CNNs têm a capacidade de extrair um número elevado de *features*, representando-as hierarquicamente. Existem várias CNNs, com diferentes camadas de processamento. Uma das camadas recebe os dados e outra funciona como camada de saída. Pelo meio estão camadas ocultas que recebem os dados, processam-nos e transmitem-nos para a próxima camada até atingir a última camada que produz os resultados. As principais camadas utilizadas numa CNN podem ser divididas em diferentes grupos: camadas de convolução, camadas de *max pooling* e camadas totalmente conectadas [40]. As camadas de convolução são as principais camadas destas redes e são responsáveis pela extração das *features* mais relevantes. As camadas de *max pooling* são comumente introduzidas entre as camadas de convolução de modo a reduzir a dimensionalidade da representação da rede, o que leva à redução de parâmetros e de posteriores cálculos, conservando sempre as *features* mais importantes. Uma camada totalmente conectada, ao contrário de uma camada de convolução, está totalmente conectada a todos os pontos da camada anterior e permite dividir e mapear os dados em diferentes classes. Para além disso, podem ainda existir camadas que permitem complementar uma CNN. Assim, pode-se recorrer a camadas de ativação para garantir a não-linearidade da rede para que esta não responda sempre de forma linear aos variados dados de entrada e consiga resolver problemas com maiores níveis de complexidade. *SoftMax* é um exemplo de uma camada de ativação, utilizada com frequência como última camada, para que os resultados possam ser avaliados sob a forma de uma distribuição de probabilidade [40], [41].

A VGG16 foi a rede neural escolhida para este estudo, também utilizada nos estudos [22], [23]. Esta CNN, criada pelo *Oxford Visual Geometry Group*, é vastamente utilizada em problemas de classificação de imagens e apresenta uma arquitetura com 13 camadas convolucionais, 3 camadas totalmente conectadas e 5 camadas de *pooling*. Aos primeiros dois tipos de camadas referidos anteriormente estão associadas funções de ativação (ReLU). A última camada é uma camada de ativação *SoftMax*. As ROIs, obtidas a partir das imagens originais, têm de ser redimensionadas para um tamanho de 224x224 píxeis

de modo a serem compatíveis com esta CNN. À medida que a rede se torna mais profunda, as camadas de *max pooling* reduzem progressivamente as dimensões das camadas [22], [23]. A arquitetura da referida rede é representada na Figura 4.6. A influência do número de iterações e do tamanho do conjunto de dados é crucial ao utilizar a VGG16.

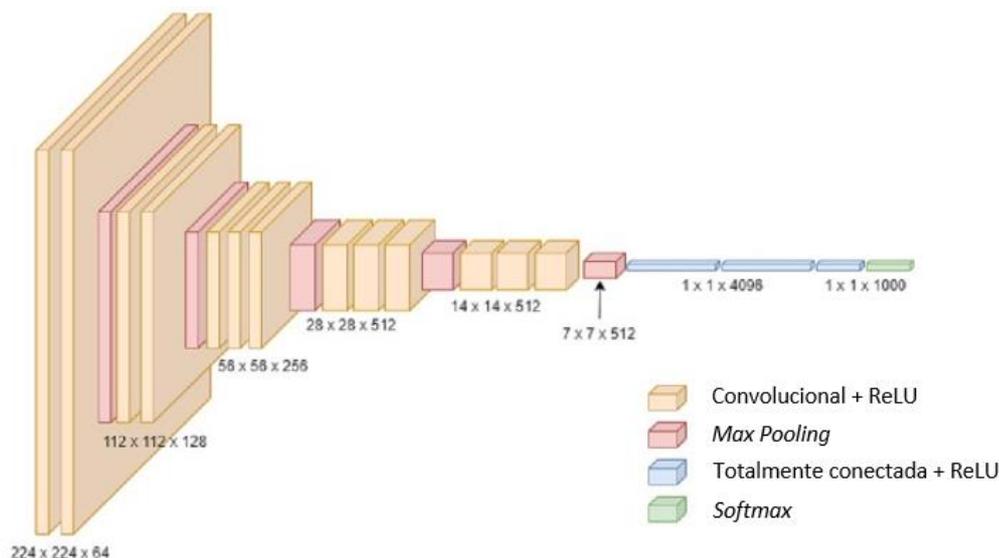


Figura 4.6 - Arquitetura de uma rede neural VGG16 com legenda das camadas constituintes (adaptado de [40])

Na arquitetura ilustrada pela Figura 4.6, a terceira dimensão da camada inicial é 64. Essa dimensão corresponde ao número de filtros utilizados na camada para criar os mapas de *features*. O número de filtros varia consoante a camada.

Para melhorar a performance de classificação foi implementado um mecanismo de atenção designado por *Multi-Head Attention* (ou Atenção de Múltiplas Cabeças). Este é usado principalmente em arquiteturas de CNNs em áreas de visão computacional como a classificação de imagens e detecção de objetos. O mecanismo auxilia os modelos de *deep learning* a focarem-se em diferentes partes de uma imagem ao mesmo tempo. Ele usa várias "cabeças" de atenção para adquirir informações e padrões de diferentes regiões da imagem, permitindo que o modelo melhore a eficácia de classificação [42].

### 4.3 Métricas utilizadas

As métricas utilizadas para a avaliação dos resultados obtidos através das abordagens de *machine* e *deep learning* foram a precisão, exatidão, *F1-Score* e o AUC, cujas fórmulas são apresentadas abaixo.

Nas expressões seguintes são usadas abreviaturas com o seguinte significado: TP representa o número de amostras positivas corretamente classificadas, FP o número de amostras positivas incorretamente classificadas, FN o número de amostras negativas incorretamente classificadas e TN está associado ao número de amostras negativas corretamente classificadas.

A precisão, Equação 1, traduz a relação das amostras positivas corretamente classificadas com o número total de amostras positivas pelo que, quando o modelo consegue classificar corretamente grande parte das amostras positivas o valor da precisão é elevado.

$$\text{Precisão} = \frac{TP}{(TP + FP)} \quad (1)$$

A exatidão mede a capacidade do modelo para fazer previsões corretas em relação ao número total de amostras avaliadas (Eq. 2). É uma medida geral de desempenho que considera as amostras positivas e negativas corretamente classificadas.

$$\text{Exatidão} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

O *F1-Score*, representado pela Equação 3, é uma medida que determina se o modelo consegue prever os valores positivos de forma correta e resulta da combinação da precisão e da sensibilidade. A sensibilidade está associada à capacidade de o modelo prever valores positivos incorretamente [43].

Esta métrica é utilizada, neste estudo, para obtenção dos resultados mais avançados de ambas as abordagens de *machine* e *deep learning*.

$$F1 - Score = \frac{2TP}{2TP + FN + FP} \quad (3)$$

A AUC (Eq. 4) é uma métrica que avalia a capacidade do modelo em distinguir entre as amostras positivas e negativas. A AUC é calculada através do integral da curva ROC (*Receiver Operating Characteristic*) ao longo de todo o intervalo [0, 1]. A curva ROC é uma representação gráfica da taxa de amostras positivas corretamente classificadas em função da taxa de amostras positivas incorretamente classificadas, para diferentes valores de limiar de classificação. Quanto maior a área sob a curva melhor é a capacidade do modelo em fazer essa distinção [22].

$$AUC = \int_0^1 ROC(x) dx \quad (4)$$

## 5 Resultados

### 5.1 Imagens

No total, foram utilizadas 207 imagens, 145 das quais foram recolhidas diretamente por um profissional de saúde, como se encontra representado na Figura 5.1. Foram obtidas 76 imagens de ecografia e 69 imagens de elastografia com e sem nódulos. As restantes 62 imagens foram extraídas de uma base de dados pública, a DDTI (*Digital Database of Thyroid Ultrasound*) [25], que contém imagens ecográficas de nódulos da tiroide.



Figura 5.1 - Aquisição de imagens ecográficas pelo profissional de saúde

As imagens foram adquiridas com um equipamento de ultrassom da *Siemens*, modelo *ACUSON NX3 Elite* e uma sonda *VF12-4*. Durante o processo de aquisição de imagens foram mantidas as seguintes configurações: um ganho de sinal entre 0 e 5 dB, sonda com frequência central de 8.0 MHz, uma gama dinâmica de 65 dB, uma taxa de atualização de 15 fps (*frames per second*) e uma escala vertical (*V-Scale*) de 65%. Para além disso, as imagens de elastografia foram recolhidas com um fator de qualidade (QF) superior a 65. Estas características de aquisição de imagem, estão ilustradas na Figura 5.2.

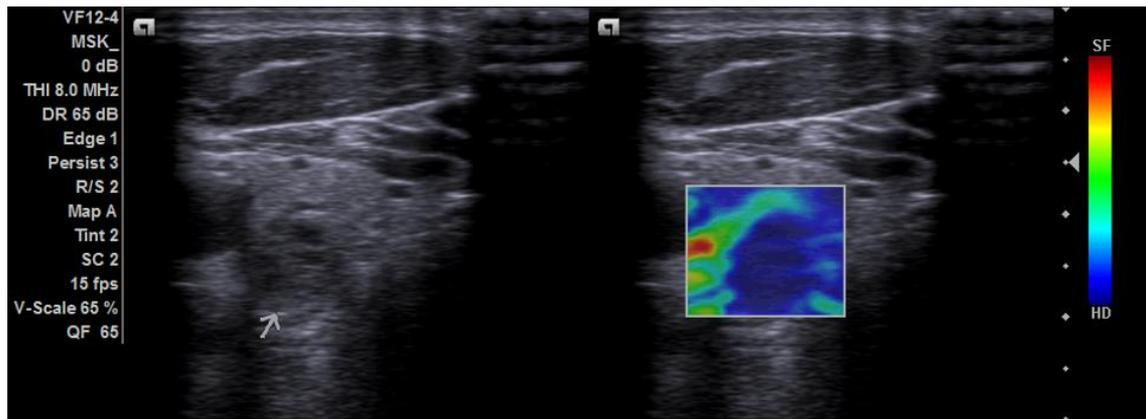


Figura 5.2 - Ecografia (esquerda) e elastografia (direita) de um nódulo da tiroide e configurações de aquisição. O código de cores apresentado refere-se à rigidez do tecido: o vermelho corresponde a tecido mais rígido e o azul a tecido menos rígido.

Após a aquisição de imagens, estas foram agrupadas e categorizadas pelo profissional de saúde em duas classes: imagens com nódulos e imagens sem nódulos. Em seguida, de cada imagem original foi extraída uma ROI. Para as imagens de ecografia sem nódulos foi selecionada uma região saudável bem definida. Para as imagens ecográficas com nódulos criou-se uma ROI de forma a abranger o máximo de região nodular possível. No caso da elastografia, o próprio equipamento permite obter uma imagem com a seleção da região retangular desejada (ver Figura 5.2) pelo que, as ROIs posteriormente extraídas, tanto com nódulos como sem nódulos, correspondem a essa região delimitada na imagem original proveniente do equipamento. Na Figura 5.3, encontram-se representadas ROIs extraídas da Figura 5.2.

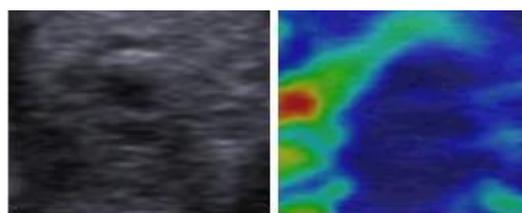


Figura 5.3 - ROIs: ecografia (esquerda) e elastografia (direita)

As ROIs criadas foram objeto de processos diferentes de pré-processamento. Como foi descrito na metodologia (capítulo 4), para o *machine learning* foi aplicado o método da janela deslizante, enquanto para o *deep learning* procedeu-se ao aumento de imagens através de rotações e inversões das imagens (ROIs).

## 5.2 Implementação em *Python*

O algoritmo de *machine learning* foi processado por um computador do modelo *HP Pavilion 15-bc000np* com um processador *Intel(R) Core (TM) i7-6700HQ CPU @ 2.60GHz*, com 8.00 GB de memória RAM instalada e uma *GPU NVIDIA GeForce GTX 950M*. A implementação e processamento do método *deep learning*, usando uma rede neural VGG16, efetuou-se num computador com melhores recursos, contendo uma processador *Intel(R) Core (TM) i5-10600K CPU @ 4.10 GHz*, 16 GB de memória RAM e uma *GPU NVIDIA GeForce RTX 3060 Ti*. Ambas as abordagens foram implementadas na linguagem de programação *Python*<sup>1</sup> através do *Jupyter Notebook*<sup>2</sup>. Para aceder a esta plataforma foi utilizado o *Anaconda Navigator*<sup>3</sup>.

## 5.3 *Machine Learning*

Diversas bibliotecas foram utilizadas para a implementação, em *Python*, deste algoritmo tradicional: *OpenCV*<sup>4</sup> para processamento de imagens; *Numpy*<sup>5</sup> para cálculos matemáticos e manipulação de *arrays*; *Pandas*<sup>6</sup> para manipulação de dados e criação de *dataframes*; *Scikit-Learn*<sup>7</sup> para pré-processamento de dados, modelos de classificação e seleção de *features* e *SimpleITK*<sup>8</sup> para processamento de imagens médicas. O passo primordial de um algoritmo de *machine learning* é a extração de *features*. Para tal, foram utilizados extratores de *features* provenientes de três bibliotecas distintas: *Pyradiomics* [44], *Pyfeats*<sup>9</sup> e *Scikit-image* [45] (consultar anexo A). Depois da implementação destes extratores, foi aplicada às *features* representadas na forma de um ou mais *arrays* uma função com onze variáveis básicas: média, máximo, mínimo, mediana, variância,

---

<sup>1</sup> <https://www.python.org/>

<sup>2</sup> <https://jupyter.org/>

<sup>3</sup> <https://www.anaconda.com/>

<sup>4</sup> <https://pypi.org/project/opencv-python/>

<sup>5</sup> <https://numpy.org/>

<sup>6</sup> <https://pandas.pydata.org/>

<sup>7</sup> <https://scikit-learn.org/>

<sup>8</sup> <https://simpleitk.org/>

<sup>9</sup> <https://pypi.org/project/pyfeats/>

*skewness*, *kurtosis*, desvio, moda, *range* (máximo-mínimo) e entropia. Obtiveram-se, assim, um total de 606 *features* (um extrator permite obter mais do que uma *feature*), ainda que algumas destas sejam obtidas de forma indireta, como foi explicado anteriormente.

### 5.3.1 Desempenho dos classificadores

Começando com um número fixo de *features* (20) estabelecido de forma arbitrária, variaram-se, igualmente de forma aleatória, as configurações da janela deslizante. Foram então explorados tamanhos da janela de 5x5, 9x9, 13x13, 17x17, 21x21, 25x25 e 29x29 píxeis para passos de 7, 11, 15, 19 e 23 píxeis (Fig. 5.4).

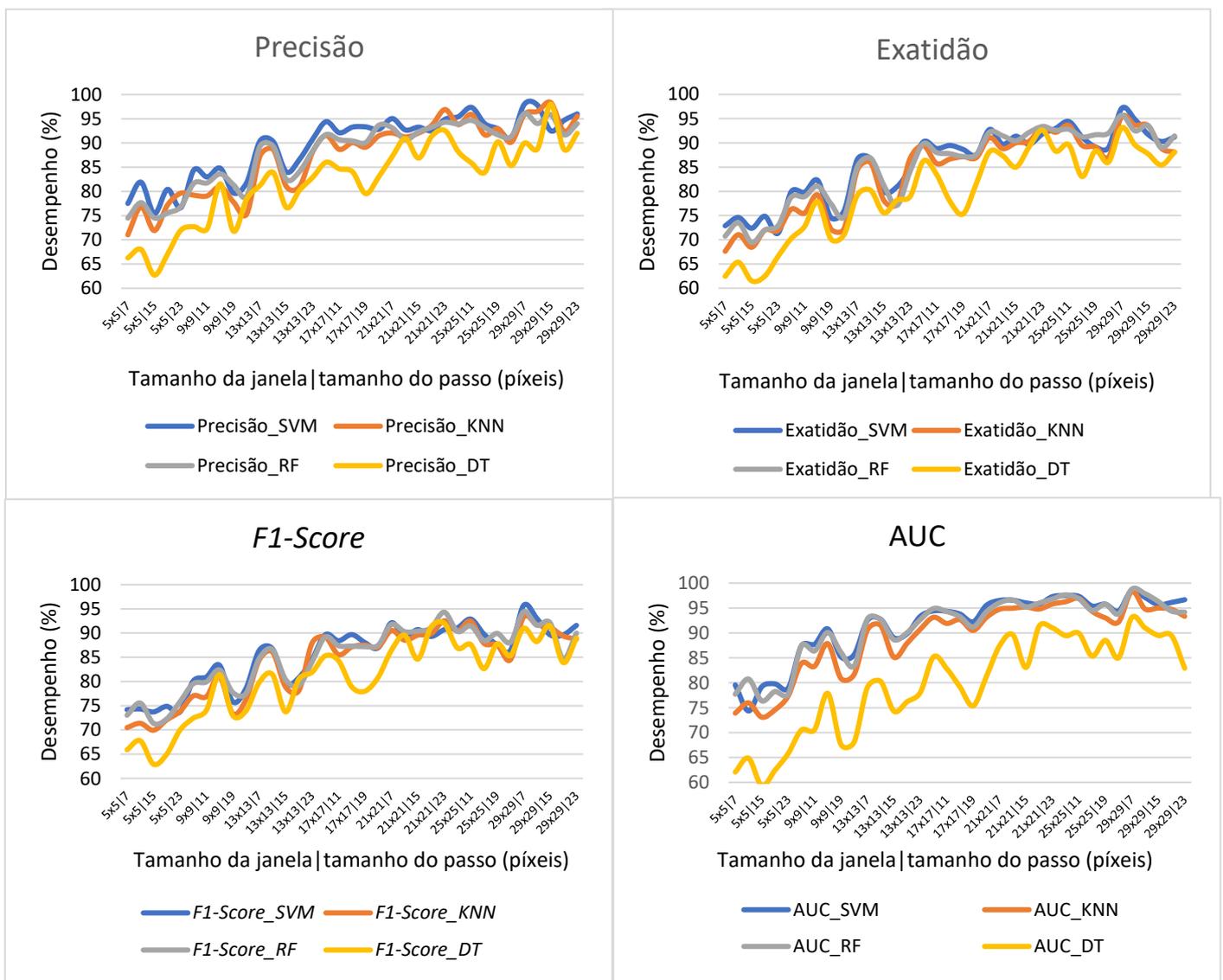


Figura 5.4 – Resultados do desempenho dos classificadores (SVM, KNN, RF e DT) para variações de tamanho da janela e tamanho do passo (e.g., 5x5|7 – 5x5 corresponde ao tamanho da janela (píxeis) e 7 ao tamanho do passo (píxeis)). As métricas de avaliação incluem a exatidão, precisão, F1-Score e AUC.

Os resultados do desempenho dos quatro classificadores (SVM, KNN, RF e DT), com base nas métricas precisão, exatidão, *F1-Score* e AUC, são apresentados na Figura 5.4. Todos os testes foram efetuados para conjuntos de treino de 80% e 20% de teste.

Observa-se que o classificador SVM apresenta um melhor desempenho global de classificação, embora no gráfico da métrica AUC se verifique uma ligeira superiorização de desempenho por parte do classificador RF. Por conseguinte, o classificador SVM foi utilizado para obter os resultados subsequentes, devido à sua boa performance geral em comparação com os outros classificadores.

### 5.3.2 Tamanho do passo

Baseado na literatura, em que investigadores usaram um tamanho da janela deslizante de 17x17 píxeis no seu estudo [10], partindo-se desta janela, fez-se variar o tamanho do passo, de modo a perceber qual permitia conduzir a um melhor desempenho por parte do classificador SVM. O comportamento do classificador em função do tamanho do passo é ilustrado na Figura 5.5. Para todos os tamanhos do passo utilizados, a relação entre imagens sem nódulos e imagens com nódulos (ver Tabela B.1) foi mantida dentro de uma faixa de 60%, o que corresponde a um rácio de 0,67 a 1,5 (representado na Equação 5).

$$\frac{40}{60} \leq \frac{\text{Imagens sem nódulos}}{\text{Imagens com nódulos}} \leq \frac{60}{40} \quad (5)$$

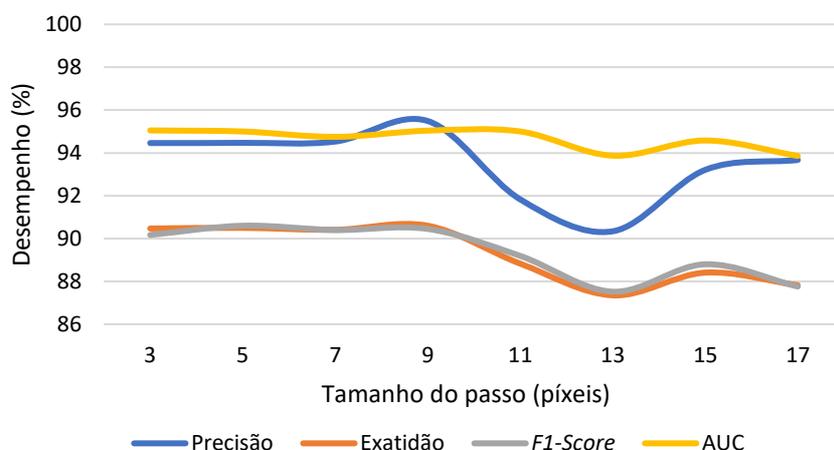


Figura 5.5 - Desempenho do classificador SVM em função da variação do tamanho do passo (tamanho da janela = 17x17 píxeis)

Da análise do gráfico, conclui-se que, para o tamanho da janela utilizado, existe um decréscimo geral das diferentes métricas à medida que se vai aumentando o tamanho do passo. Para um tamanho do passo igual 9 píxeis, o desempenho do classificador atingiu o seu máximo para a globalidade das métricas.

### 5.3.3 Tamanho da janela

No seguimento dos resultados anteriores, foram realizados testes com vista a analisar se para o tamanho do passo verificado existia alguma janela que permitia atingir um desempenho ainda melhor. Assim, foram aplicadas janelas deslizantes com diferentes tamanhos para um passo de 9 píxeis. Foram excluídos (indicados a vermelho na Tabela B.2 do anexo B) os tamanhos da janela deslizante que conferiam uma relação entre imagens sem nódulos e imagens com nódulos inferior a 60%, ou seja, que não verificavam as condições da equação 5.

Os resultados do desempenho do classificador SVM em função do tamanho da janela são apresentados na Figura 5.6.

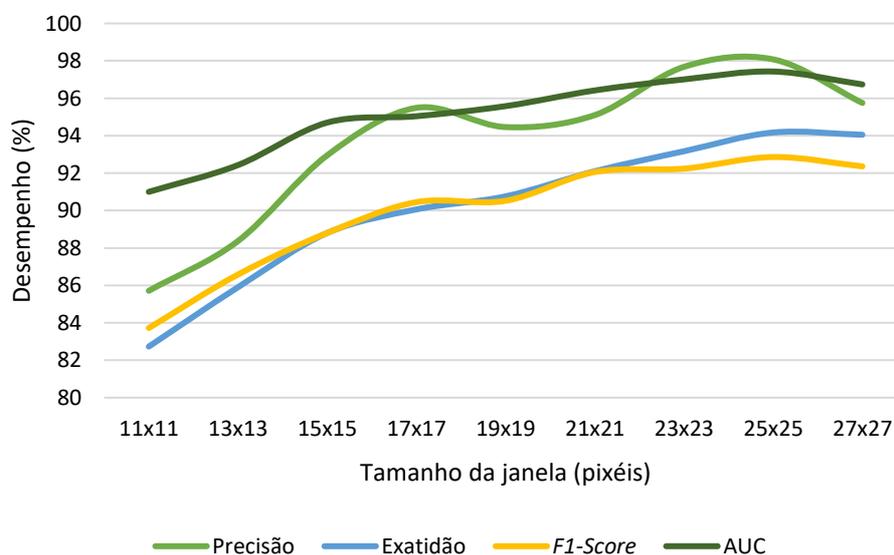


Figura 5.6 - Desempenho do classificador SVM em função da variação do tamanho da janela (tamanho do passo = 9 píxeis)

Em conclusão, verifica-se que o classificador consegue atingir desempenhos superiores ao verificado aquando do uso de uma janela 17x17 píxeis. Mais concretamente, uma janela

com dimensões de 25x25 píxeis conduziu a um melhor desempenho por parte do classificador SVM, para todas a métricas utilizadas.

### 5.3.4 Normalização e seleção de *features*

Tendo definidos o tamanho da janela e o tamanho do passo para os quais o desempenho do classificador é maximizado, analisou-se a influência das funções utilizadas para fazer a normalização e posterior seleção de *features*. No caso da normalização de *features*, foram estudadas e comparadas as funções *StandardScaler* e *MinMaxScaler*. Para além disso, foram introduzidas nesta comparação as funções de seleção de *features* *SelectKBest* e *Recursive Feature Elimination* (RFE). Os resultados são apresentados Figura 5.7.

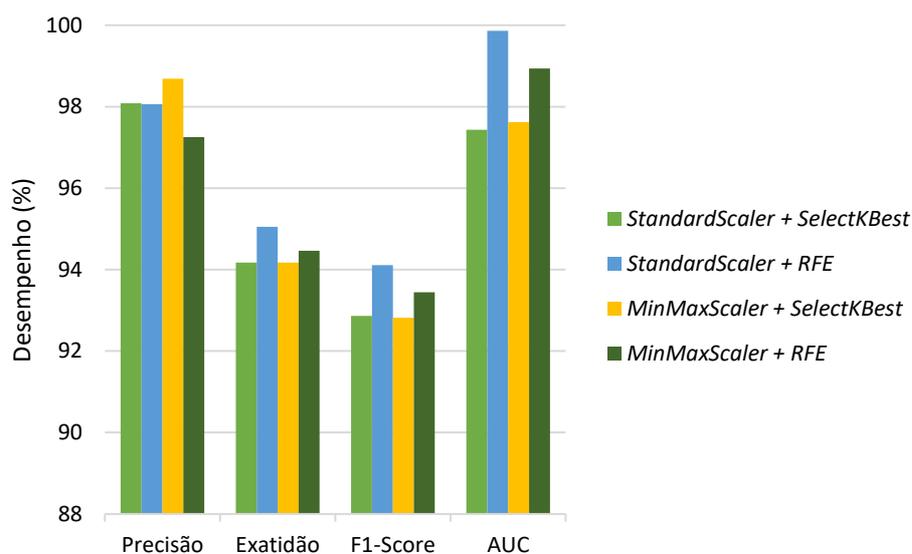


Figura 5.7 - Desempenho do classificador SVM relativamente a diferentes combinações das funções de normalização com as funções de seleção de *features*

Das combinações de funções de normalização com funções de seleção de *features* chega-se à conclusão que a melhor performance (para as métricas exatidão, *F1-Score* e AUC) ocorre quando se associa a função *StandardScaler* à função *Recursive Feature Elimination*. Apesar de, para a precisão, a melhor combinação ser *MinMaxScaler* com a *SelectKBest*, a diferença é insignificante, na ordem das décimas, sendo pouco

significativa. Assim, a primeira combinação referida (*StandardScaler* + RFE) é a que tem mais relevância para a continuação do estudo.

### 5.3.5 *k-Fold Cross-Validation*

Na Figura 5.8 são ilustrados os resultados proporcionados pela aplicação de *5-fold cross-validation* e de *10-fold cross-validation*, com o intuito de avaliar as diferenças de desempenho.

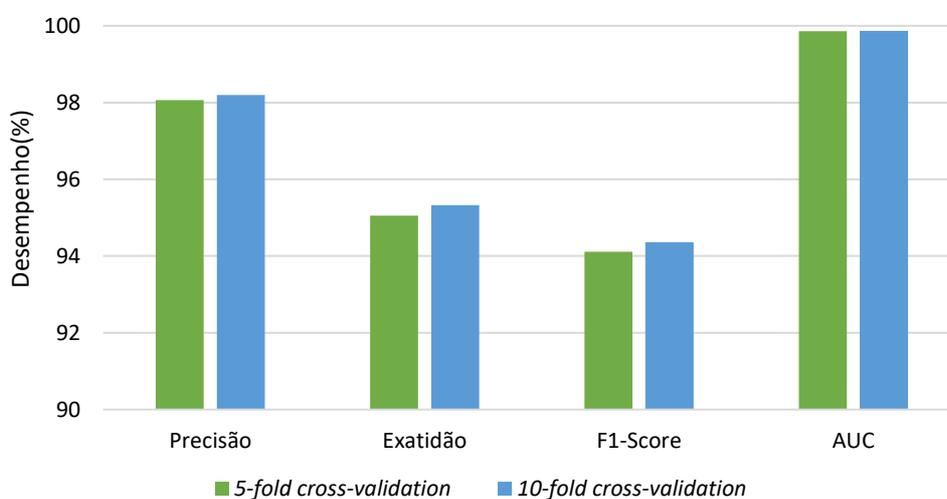


Figura 5.8 - Comparação do desempenho do classificador SVM utilizando *5-fold cross-validation* e *10-fold cross-validation*

Constata-se que o desempenho do classificador é ligeiramente superior quando se utiliza *10-fold cross-validation*, ainda que a vantagem desta abordagem seja reduzida relativamente ao uso de um conjunto de *5 folds*.

### 5.3.6 Número de *features*

O desempenho do modelo de classificação depende também do número de *features* selecionadas, como demonstra a evolução representada na Figura 5.9. Neste teste, é

apenas apresentada como métrica o *F1-score*, de modo a ter uma análise mais detalhada do número de *features* ideal.

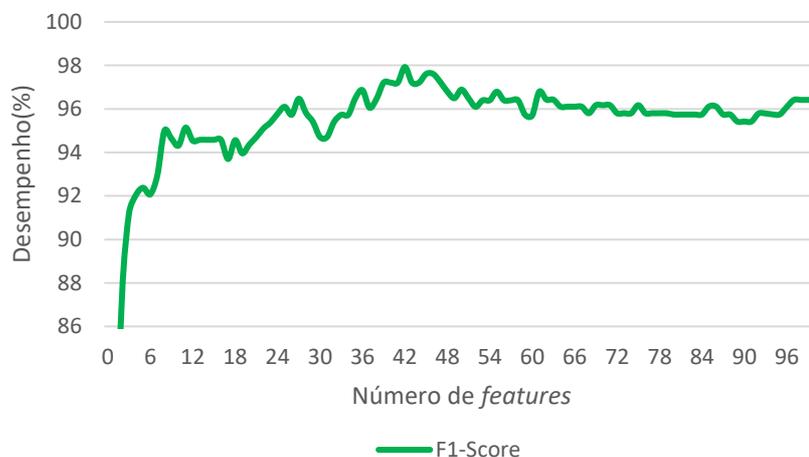


Figura 5.9 - Desempenho (F1-Score) do classificador SVM em função da variação do número de features

O traçado da Figura 5.9 evidencia um rápido incremento do desempenho do classificador até o número de *features* ser 10. A partir deste valor, ocorre uma subida menos acentuada com posterior estabilização do desempenho. O classificador SVM apresenta máximo desempenho para 40 *features* selecionadas.

### 5.3.7 Ecografia versus Elastografia

Com o classificador SVM devidamente caracterizado em termos de desempenho, procedeu-se à classificação das diferentes categorias de imagens, nomeadamente, ecografia, elastografia e a combinação das duas modalidades.

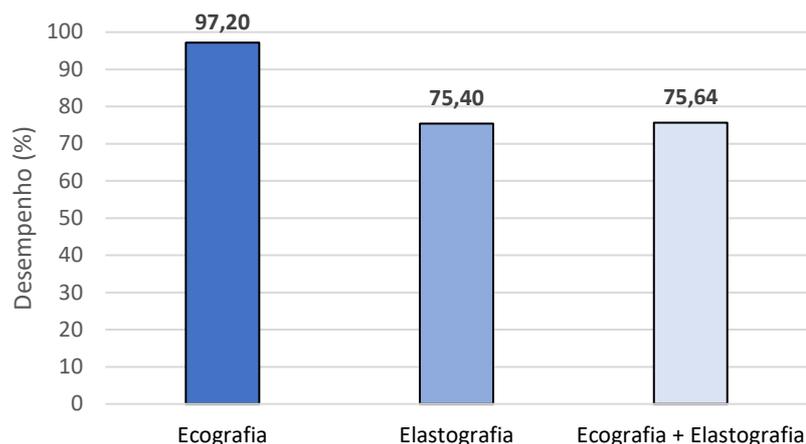


Figura 5.10 – Desempenho do classificador SVM (*F1-Score*) para imagens de ecografia, elastografia e para a combinação das duas modalidades de imagem

Os resultados do classificador, usando a métrica *F1-Score*, são ilustrados na Figura 5.10. Observa-se que o desempenho do modelo varia em função do grupo de imagens. Assim, para as imagens ecográficas, o modelo atingiu uma taxa de 97,20%. Para imagens de elastografia, o modelo obteve o valor de 75,40%. O uso combinado de imagens de ecografia e de elastografia conduziu a um *F1-Score* ligeiramente superior (75,64%) ao alcançado para o uso exclusivo de imagens elastográficas.

A discrepância nas taxas de *F1-score* entre as modalidades de imagem de ecografia e elastografia pode ser atribuída à diferença do número de imagens utilizadas. Para a combinação de ambos os tipos de imagens, este fator pode provocar um fraco balanceamento dos conjuntos de treino e teste. Além disso, o algoritmo de *machine learning* foi otimizado apenas para as imagens ecográficas, resultando em desempenhos mais reduzidos para as imagens provenientes da elastografia. As *features* extraídas das imagens de ecografia apresentam também grande importância na eficácia de classificação, podendo conter parâmetros e informações com diferentes relevâncias. Isso pode fazer com que o algoritmo de *machine learning* apresente comportamentos diferentes para as imagens de ecografia, elastografia e para a combinação das duas modalidades imagiológicas apesar da implementação de *cross-validation*.

## 5.4 Deep Learning

Nesta abordagem, foram utilizadas várias bibliotecas para proceder à implementação do algoritmo. Utilizou-se novamente a biblioteca *Numpy* para manipulação de matrizes e operações numéricas, a *Scikit-Learn* para efetuar a divisão do conjunto de dados em treino e teste, bem como para o cálculo das métricas. Além disso, fez-se uso da biblioteca *Keras*<sup>10</sup> que possui a *TensorFlow*<sup>11</sup> (biblioteca de código aberto). O *TensorFlow* permite a criação e treino dos modelos de *deep learning*, a implementação da rede neural VGG16 e do mecanismo de atenção.

### 5.4.1 Tamanho do conjunto de dados

Após o pré-processamento das imagens, estas começam por ser redimensionadas para 224x224 píxeis com posterior normalização dos valores dos píxeis entre 0 e 1. Foram testados diferentes parâmetros e configurações para produzir o melhor desempenho de classificação com dados divididos em 80% de treino 20% teste. Neste contexto, foi então selecionado um elevado número de iterações (15) para calcular o desempenho da VGG16 para diferentes tamanhos do conjunto de dados (16, 32, 64 e 128). Os resultados são mostrados na Figura 5.11.

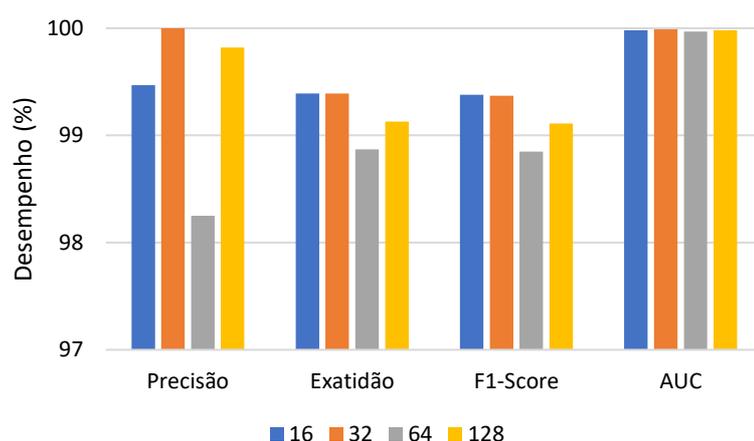


Figura 5.11 – Desempenho da VGG16 para diferentes tamanhos do conjunto de

<sup>10</sup> <https://keras.io/>

<sup>11</sup> <https://www.tensorflow.org/>

Observa-se, na Figura 5.11, que a métrica AUC apresenta, inesperadamente, resultados extremamente elevados para todos os tamanhos dos conjuntos de dados, pelo que foi descartada desta análise. O conjunto de dados de tamanho 16 apresenta uma precisão de 99,47%, ligeiramente inferior ao desempenho para um tamanho de 32. Além disso, as métricas *F1-score* e exatidão revelam que os conjuntos de dados de tamanho igual a 16, 32 ou 128 conduzem a bons resultados. O tamanho de dados igual a 64 é o que apresenta os piores resultados

Além dos resultados de desempenho, é necessário ter em consideração a eficiência computacional e os seus limites neste estudo. Conjuntos de dados menores, como 16, otimizam a utilização dos recursos computacionais disponíveis, ao minimizar o uso de memória e o tempo de treino do modelo. Assim, decidiu-se adotar o conjunto de dados menor, ou seja, o de tamanho 16.

#### **5.4.2 Número de iterações**

Utilizando o conjunto de dados com o tamanho previamente estabelecido, conduziu-se uma análise para determinar o número ideal de iterações necessárias. Os resultados são representados na Figura 5.12. Da análise do traçado infere-se, mais uma vez, que a métrica AUC apresenta resultados excessivamente otimistas, pelo que foi prontamente descartada. A métrica precisão aumenta à medida que se aumenta o número de iterações, atingindo o seu máximo quando se utilizam 10 iterações. As curvas da exatidão e do *F1-score* exibem comportamentos semelhantes, variando apenas na ordem das décimas, com picos de desempenho máximo a 4, 6 e 8 iterações.

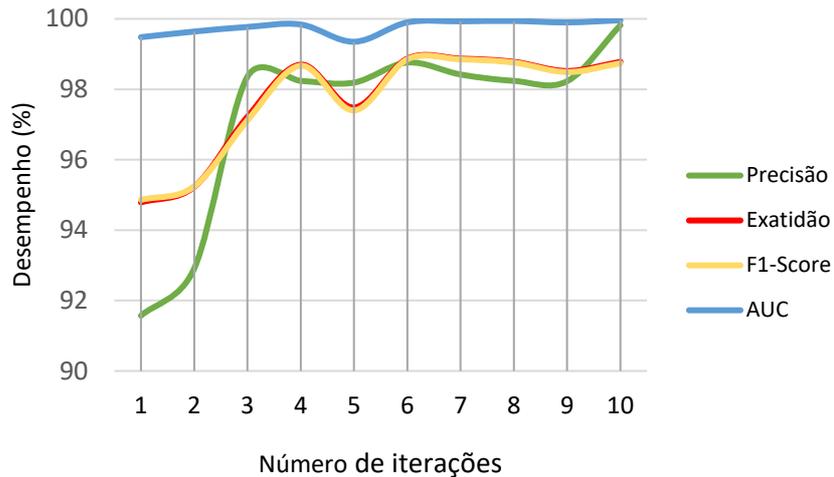


Figura 5.12 - Desempenho da VGG16 para diferentes números de iterações

Dadas as limitações dos recursos computacionais e considerando que as diferenças de desempenho entre os três diferentes números de iterações anteriormente referidos (4,6 e 8) são pouco significativas, optou-se por utilizar o menor número de iterações, ou seja, foram usadas 4 iterações nas etapas seguintes. Tal como no *machine learning*, a única métrica utilizada nos cálculos subsequentes do desempenho da VGG16 foi igualmente o *F1-Score*.

### 5.4.3 Otimização

O otimizador é um componente essencial no processo de *deep learning*. Foram comparados cinco otimizadores: o Adam (*Adaptive Moment Estimation*), o Adamax (variante do Adam), o SGD (*Stochastic Gradient Descent*), o Adragad (*Adaptive Gradient Algorithm*) e o RMSprop (*Root Mean Square Propagation*). Os resultados desta análise são apresentados na Figura 5.13.

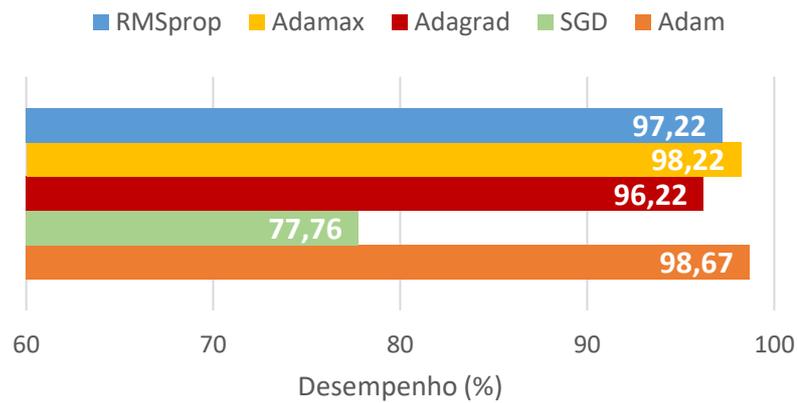


Figura 5.13 -Desempenho (F1-Score) da VGG16 para diferentes otimizadores

Entre os otimizadores avaliados, o SGD exibiu o pior desempenho. Os restantes apresentaram resultados muito elevados, com valores de *F1-Score* superiores a 95%, sendo o otimizador Adam o que proporcionou o melhor resultado (98,67%).

#### 5.4.4 Desempenho da VGG16 com o mecanismo de atenção

Finalmente, foi analisado o desempenho da VGG16 na classificação das imagens relativamente à presença de nódulos da tiroide, bem como a influência do mecanismo de atenção (*Multi-Head Attention*) no desempenho da rede neural (Fig. 5.14).

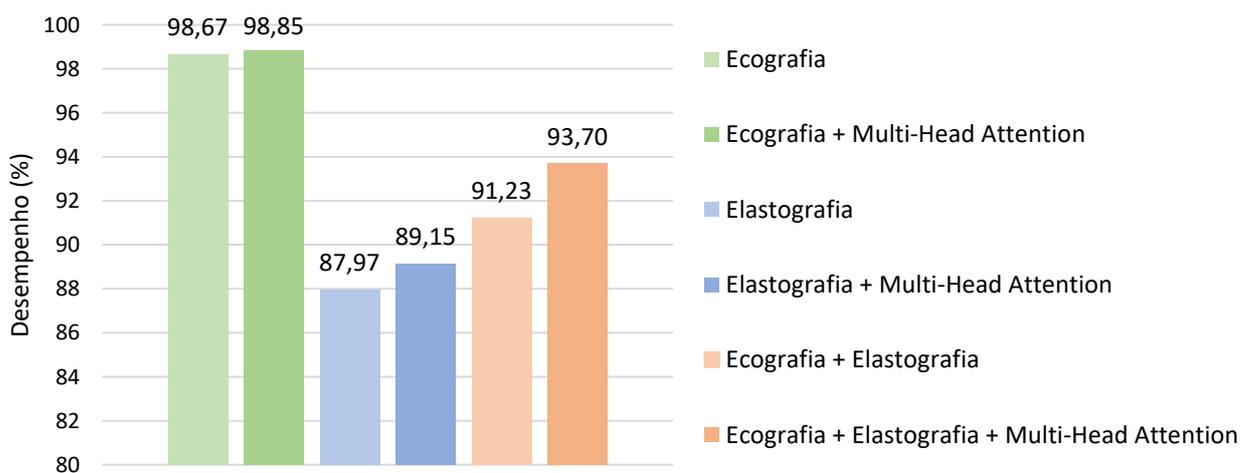


Figura 5.14 – Desempenho (F1-Score) da VGG16 para diferentes tipos de imagens (ecografia, elastografia e combinação das duas). Estudo da influência do mecanismo de atenção

Dos resultados ilustrados na Figura 5.14 infere-se que a classificação das imagens ecográficas apresenta valores de *F1-Score* de 98,67%. A aplicação do mecanismo de *Multi-Head Attention* provoca um ligeiro aumento desse valor para 98,85%. No que diz respeito às imagens provenientes da técnica de elastografia, verifica-se uma diferença mais notória, passando de 87,97% para 89,15%, evidenciando a eficácia do mecanismo de atenção na melhoria da performance deste processo de classificação. Quando se combinam os dois tipos de imagens obtém-se um valor inicial de 91,23% e 93,70% com o mecanismo de atenção.

### 5.5 *Machine Learning versus Deep Learning*

Após a análise dos resultados obtidos anteriormente, apresenta-se na Figura 5.15 uma síntese dos melhores desempenhos obtidos por ambas as metodologias adotadas, para os diferentes tipos de imagens estudadas.

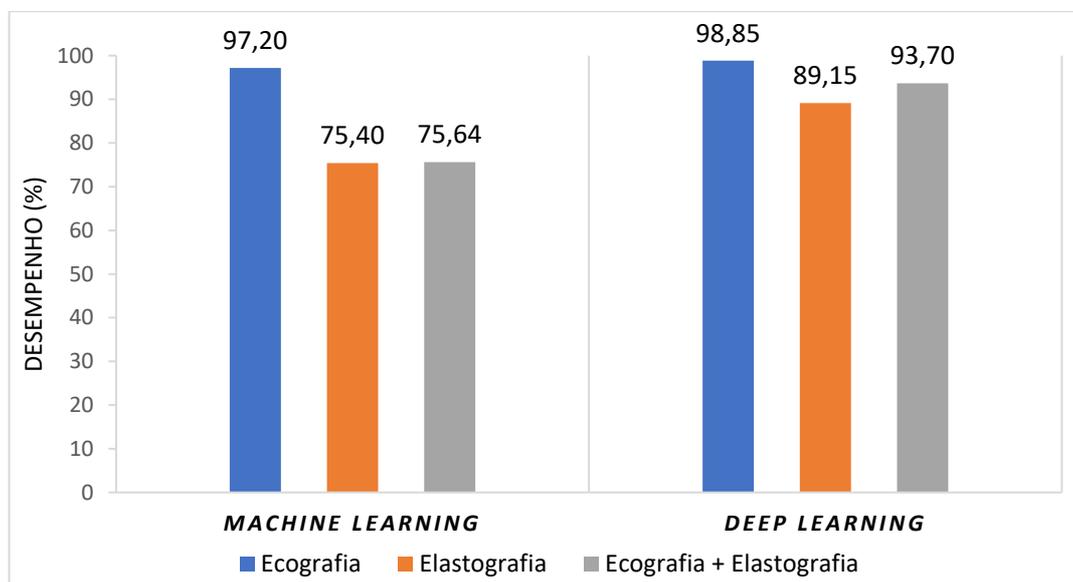


Figura 5.15 - Comparação dos resultados obtidos utilizando machine learning e deep learning

A Figura 5.15 revela os resultados da classificação das imagens utilizando abordagens de *machine learning* e *deep learning* em três modalidades de imagem: ecografia, elastografia e a sua combinação. Ao examinar os resultados, é evidente que o *deep learning* supera consistentemente o *machine learning* em todas as configurações. Por exemplo, na

modalidade de ecografia, o desempenho da VGG16 atinge um impressionante *F1-score* de 98,85%, em contraste com os 97,20% do SVM. A diferença é igualmente notável na modalidade de elastografia, onde o *deep learning* supera o *machine learning* com *F1-scores* de 89,15% e 75,40%, respetivamente. Vale a pena destacar que a combinação de ambas as modalidades de imagem resulta em melhorias no desempenho em relação ao uso isolado da elastografia, com *F1-scores* de 93,70% para o *deep learning* e 75,64% para o *machine learning*, embora o aumento no caso deste último seja menos significativo. As imagens ecográficas associadas a uma arquitetura CNN (VGG16) mostraram-se com as melhores capacidades para a classificação de imagens quanto à presença de nódulos.

## 6 Conclusão

Este trabalho apresenta um estudo comparativo de abordagens com base em metodologias de *machine learning* e *deep learning*, aplicadas à classificação de imagens de ecografia e elastografia de nódulos da tiroide. Mais concretamente, pretendeu-se avaliar a eficácia dos algoritmos na discriminação e identificação de nódulos da tiroide em imagens ecográficas e elastográficas. Para a abordagem que se baseou no *machine learning*, o classificador SVM surgiu como o mais adequado para a classificação das imagens, apresentando um maior desempenho para as imagens obtidas por ecografia (*F1-score*, 97,20%) em contraste com as obtidas por elastografia (75,40%). A combinação dos dois tipos de imagens, proporcionou um desempenho de 75,64%. Por outro lado, o uso dos algoritmos de *deep learning*, empregando uma arquitetura de uma rede neural convolucional, caso da VGG16, conduziram a um *F1-Score* de 98,85% para imagens de ecografia, 89,15% para imagens de elastografia e 93,70% para a combinação das duas.

O *machine learning* e o *deep learning* são duas técnicas fundamentais para a resolução de problemas complexos caracterizados por um grande volume de dados. A classificação tradicional é menos complexa e necessita de menos dados de entrada, mas requer que a extração de *features* seja programada manualmente. Em contrapartida, o *deep learning* funciona com base em processos automatizados de extração de *features*. Esta diferença nos processos de aquisição de *features* está intrinsecamente relacionada com as diferenças de desempenho verificadas entre os dois modelos de inteligência artificial empregues. É possível que as *features* mais relevantes necessárias para uma classificação mais eficaz das imagens, pelo SVM, não tenham sido obtidas devido à falta de implementação dos respetivos extratores. Pelo contrário, os mecanismos automáticos de extração de *features* característicos do *deep learning* permitiram a obtenção de resultados superiores.

É de salientar que, apesar dos resultados obtidos por *machine learning* demonstrarem um desempenho inferior em relação ao *deep learning*, é evidente que os requisitos computacionais necessários também são consideravelmente menores. Assim, é viável implementar este algoritmo em computadores mais modestos comparativamente aos necessários para implementar uma metodologia de *deep learning*. Essa diferença nos requisitos de *hardware* pode tornar o *machine learning* uma opção atrativa para cenários em que os recursos computacionais são limitados.

Verifica-se, ainda, que ao usar exclusivamente imagens ecográficas, os resultados são superiores. Tal pode ser explicado através de diversos fatores. Primeiramente, devido à dificuldade de aquisição de imagens com nódulos, teve de se recorrer a uma base de dados pública para o efeito. Como o equipamento de ultrassons utilizado é diferente do nosso estudo, as imagens ecográficas obtidas podem ser diferentes das adquiridas pelo profissional, induzindo maior diferenciação (e um melhor desempenho) para este tipo de imagens, apesar da implementação de funções de normalização. As imagens elastográficas foram todas obtidas pelo profissional de saúde, não havendo este problema. Para além disso, a o conjunto de imagens de ecografia é superior ao da elastografia e a otimização do algoritmo, no caso de *machine learning*, foi efetuada especificamente para as imagens ecográficas, o que pode estar relacionado com as diferenças de desempenho entre as diferentes modalidades de imagem. As *features* extraídas das imagens de cada modalidade apresentam diferentes relevâncias e, conseqüentemente, comportamentos distintos por parte dos algoritmos de classificação. Apesar de haver essa discrepância, os resultados para as imagens de elastografia e para a combinação das imagens da ecografia com as das elastografia são satisfatórios comparativamente a outros estudos anteriormente realizados.

Desta forma, acredita-se que o uso de conjuntos de dados de maior dimensão e bastante diversificado é fundamental, não só para melhorar os resultados obtidos, como para testar as capacidades dos modelos de *machine learning* e *deep learning* implementados neste estudo.

Os resultados promissores deste estudo oferecem valiosas perspectivas para investigações futuras na área de identificação e classificação de nódulos da tiroide com base em imagens médicas. A contínua expansão de conjuntos de dados e o desenvolvimento de algoritmos de *deep learning* têm o potencial de revolucionar a deteção precoce e o diagnóstico desta patologia comum, melhorando a qualidade de vida dos pacientes e reduzindo os riscos associados a condições mais graves, como o cancro da tiroide. Portanto, este estudo marca um importante passo em direção a avanços significativos na área da medicina diagnóstica.

## 7 Referências

- [1] P. Tate, T. Stephens, and R. Seeley, *Anatomy and Physiology*, 6th ed. New York, USA: McGraw-Hill, 2003.
- [2] F. Martini, M. Timmons, and R. Tallitsch, *Fundamentals of Anatomy & Physiology*, 6th ed. San Francisco, USA: Pearson, 2009.
- [3] A. Viduetsky and C. L. Herrejon, “Sonographic Evaluation of Thyroid Size: A Review of Important Measurement Parameters,” *Journal of Diagnostic Medical Sonography*, vol. 35, no. 3. 2019. doi: 10.1177/8756479318824290.
- [4] A. André, “Apontamentos da unidade curricular Ecografia I.” 2020.
- [5] E. Marieb, P. Wilhem, and J. Mallat, *Marieb Hoehn Anatomy Physiology*, 6th ed. Boston, USA: Pearson, 2012.
- [6] “Thyroid Nodules: Symptoms, Causes, Diagnosis, & Treatment.” <https://www.healthcentral.com/condition/thyroid-nodules?legacy=ew> (accessed Sep. 15, 2023).
- [7] S. Tamhane and H. Gharib, “Thyroid nodule update on diagnosis and management,” *Clin Diabetes Endocrinol*, vol. 2, no. 1, 2016, doi: 10.1186/s40842-016-0035-7.
- [8] E. A. De Filippis, A. Sabet, M. R. M. Sun, and J. R. Garber, “Pemberton’s sign: explained nearly 70 years later,” *J Clin Endocrinol Metab*, vol. 99, no. 6, pp. 1949–1954, 2014, doi: 10.1210/JC.2013-4240.
- [9] C. Rumack and D. Levine, *Diagnostic ultrasound*, 5th ed. Philadelphia, USA: Elsevier, 2017.
- [10] “Science Education | National Institute of Biomedical Imaging and Bioengineering.” <https://www.nibib.nih.gov/science-education> (accessed Sep. 12, 2023).
- [11] Barr Richard G., *Elastography: A Practical Approach*, 1st ed. New York, USA: Thieme, 2017.

- [12] S. Edyvean and J. Gelijns, *Chapter 11: Computed Tomography*, vol. 44, no. 2. 2013.
- [13] C. M. Collins, *Electromagnetics in Magnetic Resonance Imaging*. in 2053-2571. Morgan & Claypool Publishers, 2016. doi: 10.1088/978-1-6817-4083-6.
- [14] B. K. Das, *Positron Emission Tomography : A Guide for Clinicians*. New Delhi, India: Springer, 2015.
- [15] A. Prochazka, S. Gulati, S. Holinka, and D. Smutek, "Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition," *Computerized Medical Imaging and Graphics*, vol. 71, 2019, doi: 10.1016/j.compmedimag.2018.10.001.
- [16] C. K. Zhao *et al.*, "A Comparative Analysis of Two Machine Learning-Based Diagnostic Patterns with Thyroid Imaging Reporting and Data System for Thyroid Nodules: Diagnostic Performance and Unnecessary Biopsy Rate," *Thyroid*, vol. 31, no. 3, pp. 470–481, Mar. 2021, doi: 10.1089/THY.2020.0305.
- [17] P. Qin, K. Wu, Y. Hu, J. Zeng, and X. Chai, "Diagnosis of Benign and Malignant Thyroid Nodules Using Combined Conventional Ultrasound and Ultrasound Elasticity Imaging," *IEEE J Biomed Health Inform*, vol. 24, no. 4, pp. 1028–1036, Apr. 2020, doi: 10.1109/JBHI.2019.2950994.
- [18] M. Feigin, D. Freedman, and B. W. Anthony, "A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound," *IEEE Trans Biomed Eng*, vol. 67, no. 4, pp. 1142–1151, Apr. 2020, doi: 10.1109/TBME.2019.2931195.
- [19] R. R. Wildeboer *et al.*, "Synthetic Elastography Using B-Mode Ultrasound through a Deep Fully Convolutional Neural Network," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 67, no. 12, pp. 2640–2648, Dec. 2020, doi: 10.1109/TUFFC.2020.2983099.
- [20] Z. Xiang *et al.*, "Self-supervised multi-modal fusion network for multi-modal thyroid ultrasound image diagnosis," *Comput Biol Med*, vol. 150, Nov. 2022, doi: 10.1016/J.COMPBIOMED.2022.106164.
- [21] L. Hu, C. Pei, L. Xie, Z. Liu, N. He, and W. Lv, "Convolutional Neural Network for Predicting Thyroid Cancer Based on Ultrasound Elastography Image of

- Perinodular Region,” *Endocrinology*, vol. 163, no. 11, Nov. 2022, doi: 10.1210/ENDOCR/BQAC135.
- [22] Y. Wang *et al.*, “Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images,” *IEEE Access*, vol. 8, pp. 52010–52017, 2020, doi: 10.1109/ACCESS.2020.2980290.
- [23] H. Sun, F. Yu, and H. Xu, “Discriminating the Nature of Thyroid Nodules Using the Hybrid Method,” *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/6147037.
- [24] W. Yang *et al.*, “Integrate domain knowledge in training multi-task cascade deep learning model for benign–malignant thyroid nodule classification on ultrasound images,” *Eng Appl Artif Intell*, vol. 98, p. 104064, Feb. 2021, doi: 10.1016/J.ENGAPPAI.2020.104064.
- [25] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, “An open access thyroid ultrasound image database,” in *10th International Symposium on Medical Information Processing and Analysis*, 2015. doi: 10.1117/12.2073532.
- [26] T. Wunderling, B. Golla, P. Poudel, C. Arens, M. Friebe, and C. Hansen, “Comparison of thyroid segmentation techniques for 3D ultrasound,” *Medical Imaging 2017: Image Processing*, vol. 10133, p. 1013317, Feb. 2017, doi: 10.1117/12.2254234.
- [27] H. Gong *et al.*, “Multi-task learning for thyroid nodule segmentation with thyroid region prior,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2021-April, pp. 257–261, Apr. 2021, doi: 10.1109/ISBI48211.2021.9434087.
- [28] R. Song, L. Zhang, C. Zhu, J. Liu, J. Yang, and T. Zhang, “Thyroid Nodule Ultrasound Image Classification through Hybrid Feature Cropping Network,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2982767.
- [29] H. A. U. Rehman, C. Y. Lin, and S. F. Su, “Deep Learning Based Fast Screening Approach on Ultrasound Images for Thyroid Nodules Diagnosis,” *Diagnostics*, vol. 11, no. 12, Dec. 2021, doi: 10.3390/DIAGNOSTICS11122209.

- [30] D. T. Nguyen, J. K. Kang, T. D. Pham, G. Batchuluun, and K. R. Park, "Ultrasound Image-Based Diagnosis of Malignant Thyroid Nodule Using Artificial Intelligence," *Sensors (Basel)*, vol. 20, no. 7, Apr. 2020, doi: 10.3390/S20071822.
- [31] D. T. Nguyen, J. Choi, and K. R. Park, "Thyroid Nodule Segmentation in Ultrasound Image Based on Information Fusion of Suggestion and Enhancement Networks," *Mathematics 2022, Vol. 10, Page 3484*, vol. 10, no. 19, p. 3484, Sep. 2022, doi: 10.3390/MATH10193484.
- [32] Z. Tao, H. Dang, Y. Shi, W. Wang, X. Wang, and S. Ren, "Local and Context-Attention Adaptive LCA-Net for Thyroid Nodule Segmentation in Ultrasound Images," *Sensors 2022, Vol. 22*, vol. 22, no. 16, p. 5984, Aug. 2022, doi: 10.3390/S22165984.
- [33] J. Lee, J. Bang, and S. Il Yang, "Object detection with sliding window in images including multiple similar objects," *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017*, vol. 2017-December, pp. 803–806, Dec. 2017, doi: 10.1109/ICTC.2017.8190786.
- [34] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on MTCNN," *Proceedings - 2020 International Conference on Computer Network, Electronic and Automation, ICCNEA 2020*, pp. 154–158, Sep. 2020, doi: 10.1109/ICCNEA50255.2020.00040.
- [35] D. A. Otchere, T. O. A. Ganat, J. O. Ojero, B. N. Tackie-Otoo, and M. Y. Taki, "Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions," *J Pet Sci Eng*, vol. 208, p. 109244, Jan. 2022, doi: 10.1016/J.PETROL.2021.109244.
- [36] V. Sheth, U. Tripathi, and A. Sharma, "A Comparative Analysis of Machine Learning Algorithms for Classification Purpose," *Procedia Comput Sci*, vol. 215, pp. 422–431, Jan. 2022, doi: 10.1016/J.PROCS.2022.12.044.
- [37] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst Appl*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/J.ESWA.2019.05.028.

- [38] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, “The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction,” *Clean Eng Technol*, vol. 15, p. 100664, Aug. 2023, doi: 10.1016/J.CLET.2023.100664.
- [39] F. J. Pacis, A. Ambrus, S. Alyaev, R. Khosravianian, T. G. Kristiansen, and T. Wiktorski, “Improving predictive models for rate of penetration in real drilling operations through transfer learning,” *J Comput Sci*, p. 102100, Jul. 2023, doi: 10.1016/J.JOCS.2023.102100.
- [40] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, “Comparison of CNN-based deep learning architectures for rice diseases classification,” *Artificial Intelligence in Agriculture*, Jul. 2023, doi: 10.1016/J.AIIA.2023.07.001.
- [41] “CS231n Convolutional Neural Networks for Visual Recognition.” <https://cs231n.github.io/convolutional-networks/> (accessed Jul. 18, 2023).
- [42] A. Vaswani *et al.*, “Attention Is All You Need,” *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: Sep. 04, 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762v7>
- [43] S. H. Kok, A. Azween, and N. Z. Jhanjhi, “Evaluation metric for crypto-ransomware detection using machine learning,” *Journal of Information Security and Applications*, vol. 55, p. 102646, Dec. 2020, doi: 10.1016/J.JISA.2020.102646.
- [44] J. J. M. Van Griethuysen *et al.*, “Computational Radiomics System to Decode the Radiographic Phenotype,” *Cancer Res*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
- [45] S. Van Der Walt *et al.*, “Scikit-image: Image processing in python,” *PeerJ*, vol. 2014, no. 1, p. e453, Jun. 2014, doi: 10.7717/PEERJ.453/FIG-5.

## A. Anexo A

Tabela A.1 - Bibliotecas e respectivos extratores de features

<b>Bibliotecas</b>	<b>Pyradiomics</b>	<ul style="list-style-type: none"> <li>- <i>Gray Level Run Length Matrix (GLRLM)</i></li> <li>- <i>Gray Level Cooccurrence Matrix (GLCM)</i></li> <li>- <i>Shape-based (2D)</i></li> <li>- <i>Gray Level Size Zone Matrix (GLSZM)</i></li> <li>- <i>Neighbouring Gray Tone Difference Matrix (NGTDM)</i></li> <li>- <i>Gray Level Dependence Matrix (GLDM)</i></li> </ul>	
	<b>Pyfeats</b>	<ul style="list-style-type: none"> <li>- <i>Law's Texture Energy Measures (LTE/TEM)</i></li> <li>- <i>Fractal Dimension Texture Analysis (FDTA)</i></li> <li>- <i>Gray Level Difference Statistics (GLDS)</i></li> <li>- <i>Statistical Feature Matrix (SFM)</i></li> <li>- <i>Fourier Power Spectrum (FPS)</i></li> <li>- <i>Higher Order Spectra (HOS)</i></li> <li>- <i>Local Binary Pattern (LPB)</i></li> <li>- <i>Histogram</i></li> <li>- <i>Zernikes' Moments</i></li> <li>- <i>Hu's Moments</i></li> <li>- <i>Threshold Adjacency Matrix (TAS)</i></li> <li>- <i>Histogram of Oriented Gradients (HOG)</i></li> </ul>	
	<b>Scikit-image</b>	<p><b>Módulo feature</b></p> <ul style="list-style-type: none"> <li>- <i>Blob dog</i></li> <li>- <i>Corner Fast</i></li> <li>- <i>Corner Foerstner</i></li> <li>- <i>Corner Harris</i></li> <li>- <i>Corner Kitchen Rosenfeld</i></li> <li>- <i>Corner Moravec</i></li> <li>- <i>Corner Peaks</i></li> <li>- <i>Corner Shi Tomasi</i></li> <li>- <i>Daisy</i></li> <li>- <i>Hessian Matrix</i></li> <li>- <i>Hessian Matrix Eigvals</i></li> <li>- <i>Multiscale Basic Features</i></li> <li>- <i>Peak Local Max</i></li> <li>- <i>Shape Index</i></li> <li>- <i>Structure Tensor</i></li> <li>- <i>Structure Tensor Eigenvalues</i></li> </ul>	<p><b>Módulo measure</b></p> <ul style="list-style-type: none"> <li>- <i>Label</i></li> <li>- <i>Region Props</i> <ul style="list-style-type: none"> <li>○ <i>num_pixels</i></li> <li>○ <i>area_bbox</i></li> <li>○ <i>area_convex</i></li> <li>○ <i>axis_major_lenght</i></li> <li>○ <i>bbox</i></li> <li>○ <i>centroid</i></li> <li>○ <i>centroid_weighted</i></li> <li>○ <i>equivalent_diameter_area</i></li> <li>○ <i>euler_number</i></li> <li>○ <i>extent</i></li> <li>○ <i>moments_normalized</i></li> <li>○ <i>solidity</i></li> </ul> </li> </ul> <p><i>Profile Linesvm</i></p>

## B. Anexo B

Tabela B.1 - Número de imagens sem nóculo e com nóculo, tamanho do passo e rácio (imagens sem nóculo/imagem com nóculo), para um tamanho da janela de 17x17 píxeis.

Número de imagens		Tamanho do passo (píxeis)	Rácio
Sem nóculo	Com nóculo		
2297	2380	3	0,97
885	939	5	0,94
487	524	7	0,93
307	341	9	0,90
224	259	11	0,86
173	198	13	0,87
128	156	15	0,82

Tabela B.2 - Número de imagens sem nóculo e com nóculo, tamanho da janela e rácio imagens sem nóculo/imagem com nóculo), para um tamanho do passo de 9 píxeis.

Número de imagens		Tamanho da janela (píxeis)	Rácio
Sem nóculo	Com nóculo		
457	550	11x11	0,83
400	472	13x13	0,85
352	398	15x15	0,88
307	341	17x17	0,90
280	293	19x19	0,96
229	252	21x21	0,91
214	195	23x23	1,10
187	155	25x25	1,21
163	123	27x27	1,33
143	90	29x29	1,59
68	125	31x31	0,54
120	47	33x33	2,55
107	32	35x35	3,34
107	23	37x37	4,65

## **C. Anexo C**

# Ultrasound versus Elastography in the Study of Thyroid Nodules

Tiago Rocha  
uc2018263670@student.uc.pt

Jaime Santos  
jaime@deec.uc.pt

Alexandra André  
alexandra.andre@estescoimbra.pt

José Silvestre Silva  
jose.silva@academiamilitar.pt

Department of Physics, University of Coimbra, Portugal

University of Coimbra, CEMMPRE, ARISE, Department of Electrical and Computer Engineering, Portugal

Coimbra Health School, Coimbra, Portugal

CINAMIL & Portuguese Military Academy, Portugal

LIBPhys-UC & LA-Real, University of Coimbra, Portugal

## Abstract

Thyroid nodules, despite appearing as a discrete lesion, constitute a prevailing pathological occurrence within the global population. The timely detection and diagnosis can help preventing the pathology from growing, minimising more severe effects on the human body. In this study, supervised machine learning and deep learning techniques were implemented to analyse ultrasound and elastography medical images increasing and improving the effectiveness of thyroid nodule detection. The results achieved using deep learning were superior to those achieved using machine learning. Specifically, for machine learning it was obtained a F1-Score of 97.20%, for the ultrasound images and a F1-Score of 75.40% using elastography images. Deep learning methodologies reached a F1-Score of 98.85% for ultrasound images and 89.15% for elastography images.

Keywords: Thyroid, ultrasound, elastography, traditional classifiers, Deep Learning

## 1 Introduction

Ultrasound and elastography are two important medical imaging tools widely used in the assessment of thyroid nodules (figure 1). Ultrasound is a medical imaging technique that uses sound waves to produce images of the thyroid and its nodules with good resolution, providing important information such as size, shape, structure, and the presence of fluids [1]. Elastography is a more recent medical imaging technique based on highly developed software that makes it possible to determine and assess changes in the structural properties of tissues and, consequently, their stiffness, which is crucial in differentiating thyroid nodules [2]. Ultrasound and elastography are safe, non-invasive medical imaging techniques and their use in the study of thyroid nodules is essential for making an accurate and effective diagnosis [3].

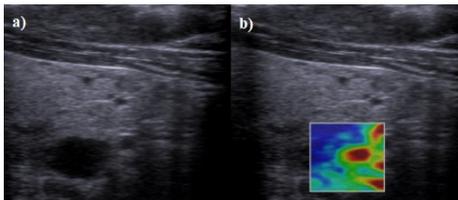


Figure 1 - Image of a thyroid nodule: a) Ultrasound; b) Elastography

## 2 Methods

### 2.1 Datasets

In total, 207 images were used, where 145 were collected directly by a health professional (76 ultrasound images and 69 elastography images with and without nodules) and 62 ultrasound images were retrieved from a public dataset, DDTI (Digital Database of Thyroid Ultrasound Images), which contains ultrasound images accompanied by a diagnosis made by radiologists and confirmed by biopsies when necessary, [4].

### 2.2 Classification and Feature Selection

After acquiring thyroid images of patients with and without pathology, the images are submitted to two classification models. In the case of machine learning, a sliding window is applied to the images to increase its number of images to make up for the shortage of images. This technique allows an image to be scrolled through a window with previously chosen dimensions from a starting point to the end of the image. It is also possible to define the step size, i.e., the number of pixels to move between each window. Then, the images are normalized and around 600 different features are extracted, which undergo a normalization process immediately before the best ones are chosen by

the algorithm. The data achieved was processed by different classifiers such Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) and K-nearest neighbors (KNN) to produce a variety of results to be compared. Cross-validation and classifier optimization mechanisms were also applied. The final aim is an implementation of classifiers to provide statistical results regarding the classification of thyroid images (with or without a nodule).

Regarding deep learning, to overcome the small dataset, the images were pre-processed, subjected to a data augmentation process and pre-trained on ImageNet (transfer learning). The classification model was implemented using a Convolutional Neural Network (CNN). The VGG16 is the neural network used in this study. This CNN has an architecture with 13 convolutional layers, 3 fully connected layers and 5 pooling layers. The first two types of layers mentioned above are associated with activation functions (ReLU), however the last layer is a SoftMax activation layer.

The metrics used to evaluate the results obtained by both above approaches were precision, accuracy, F1-Score and AUC (Area Under the ROC Curve).

## 3 Results and Discussion

A fixed number of features (20) was used to evaluate the performance of the different classifiers. Figure 2 illustrates the influence of step size and window size on the performance of the different metrics for the different classifiers.

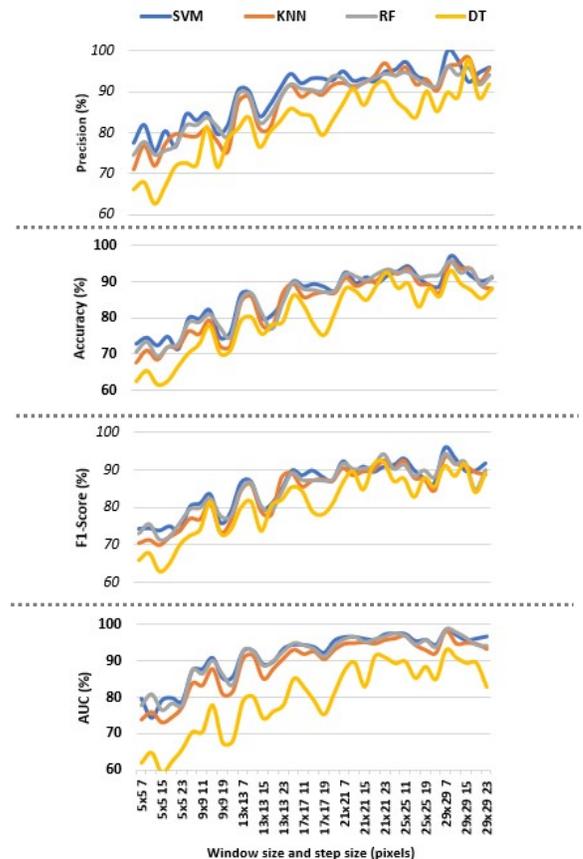


Figure 2 – Performance metric values by window size and step size for the different classifiers

It is observed that SVM classifier shows a better classification result overall, although in the AUC metric has a slight increase in performance on the part of the RF classifier. Therefore, the SVM classifier was used to acquire the subsequent results due to its good overall performance compared to the others.

The SVM was further tested to produce the maximum performance. The best values were achieved using a window of 25 x 25 (pixels) for a step size of 9. In the case of feature normalization, the StandardScaler and MinMaxScaler functions were studied and compared. The comparison included the incorporation of SelectKBest and Recursive Feature Elimination methods for feature selection, where the pair StandardScaler and Recursive Feature Elimination produced the best overall performance for the different metrics.

The performance of the classification model is closely linked to the number of features selected. Figure 3 shows the results of the classifier's performance in relation to variations in the number of features. In this case, only the F1-Score will be presented as a metric, to have a more detailed analysis of the ideal number of features.



Figure 3 – SVM performance over the number of features

It is noticed a sudden increase in SVM performance as the number of selected features is 10. Then there is a less drastic rise with subsequent stabilization of performance. The SVM classifier performs best when the number of features is near 40. Finally, once all the variables had been explored and defined, the performance of the SVM was compared (figure 4) for the two groups of images (ultrasound and elastography).

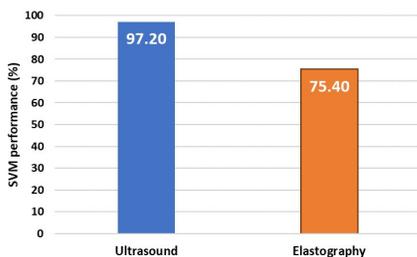


Figure 4 – SVM performance for ultrasound and elastography images

The images acquired through ultrasound produce a much better performance than the elastography. We also implemented a Deep Learning technique. After pre-processing the images, different parameters and configurations were tested to produce the best classification performance. In this context, a high number of epochs was selected to calculate an initial batch size. The analysis showed that a batch size of 16 yielded the best results. Using this established batch size, the ideal number of epochs needed was investigated, as shown in figure 5.

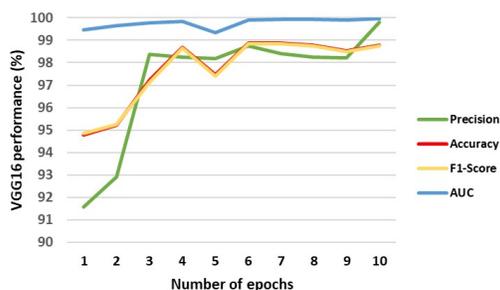


Figure 5 – VGG16 performance for different number of epochs

Examining the graph (figure 5), it becomes evident that the AUC metric always gives results that are too good, so it was promptly discarded. The performance of the accuracy metric increases, reaching its maximum at 10 epochs. The accuracy and F1-Score curves show a very similar behaviour, only varying in the order of tenths, with maximum performance peaks at 4, 6 and 8. Given the limitations of computational resources, 4 epochs were used in the following steps. As with machine learning, the metric used in the next tests was the F1-Score. A comparison between various optimisers was carried out. The results showed that the Adaptive Moment Estimation (Adam) optimiser performed better. Finally, the use of VGG16 in classifying thyroid nodules demonstrated superior performances, when complemented with a Multi-Head Attention method (figure 6).

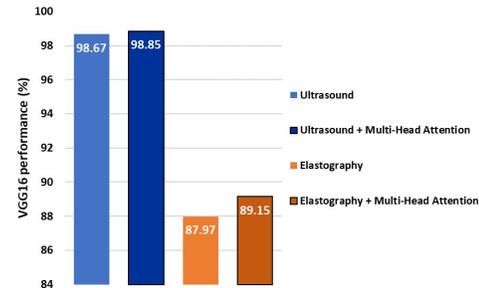


Figure 6 – VGG16 performance comparison

### 4 Conclusion

This work presents a comparative study of different classifiers as machine learning and deep learning, applied to the classification of ultrasound and elastography images of thyroid. The primary objective was to assess the performance of classification within this pathological context. The algorithms capacity to discriminate between images showing nodules and those devoid of such features was analysed. It was observed that machine learning exhibited a higher proficiency in classifying ultrasound-acquired images (97.20%) in contrast to those obtained through elastography (75.40%), with Support Vector Machine (SVM) emerging as the most suitable classifier. On the other hand, the use of deep learning, employing a Convolutional Neural Network (CNN) architecture, achieved a F1-Score of 98.85% for ultrasound images and 89.15% for elastography images. Machine Learning and Deep Learning stand as two pivotal techniques making easy the resolution of difficult problems characterized by substantial complexity and extensive data volume. The traditional classifier, while less intricate, needs limited data but requires manual programming for feature extraction. In contrast, Deep Learning operates on the premise of automated feature extraction processes. This distinction in feature acquisition processes potentially underlies the observed performance discrepancies in relation to elastography images. It is possible that the essential features required for accurate classification of elastography images using SVM might not be effectively extracted through the actual mechanism. To improve the results for elastography images a larger database should be used.

### Acknowledgements

This research is sponsored by national funds through FCT – Fundação para a Ciência e a Tecnologia, under the project UIDB/00285/2020 and LA/P/0112/2020.

### References

- [1] C. Rumack and D. Levine, *Diagnostic ultrasound*, 5th ed. Philadelphia, USA: Elsevier, 2017.
- [2] Barr Richard G., *Elastography: A Practical Approach*, 1st ed. New York, USA: Thieme, 2017.
- [3] S. Tamhane and H. Gharib, "Thyroid nodule update on diagnosis and management," *Clin Diabetes Endocrinol*, vol. 2, no. 1, 2016, doi: 10.1186/s40842-016-0035-7.
- [4] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, "An open access thyroid ultrasound image database," in *10th International Symposium on Medical Information Processing and Analysis*, 2015. doi: 10.1117/12.2073532.