

---

## Avaliação da Caracterização de Lesões em Mamografia com Recurso a Sistemas CAD

### *Mammographic CAD Systems Evaluation*

Ana C. Perre<sup>1</sup>, Luís C. Freire<sup>2</sup>

1 Técnica de Radiologia no Centro Hospitalar do Oeste, Portugal. Escola Superior de Saúde Dr. Lopes Dias – Instituto Politécnico de Castelo Branco, *PhD Student in* Faculdade Ciências Saúde – Universidade da Beira Interior

2 *PhD*, Escola Superior de Tecnologia da Saúde de Lisboa – Instituto Politécnico de Lisboa, Lisboa, Portugal

---

#### Resumo:

Os sistemas CAD auxiliam a deteção e diferenciação de lesões benignas e malignas, aumentando a performance no diagnóstico do cancro da mama. Uma vez que as lesões da mama estão fortemente correlacionadas com a forma e contorno, neste estudo, foram aplicados dois métodos diferentes para classificação de lesões em imagens de mamografia. O primeiro consiste em medidas quantitativas baseadas na dimensão fractal, calculadas através da aplicação do método “*box-counting*”, diretamente em imagens de lesões segmentadas antes e após a aplicação de um algoritmo de dilatação/erosão. O segundo método baseou-se na aplicação de Redes Neurais de Convolução (CNN), as quais têm demonstrado um elevado grau de sucesso, na deteção e classificação de patologias em diferentes modalidades de imagem médica, incluindo a mamografia. De forma a ultrapassar a limitação do reduzido número de amostras disponíveis nas bases de dados de mamografia, foi aplicado o método de “*transfer learning*”, no qual três modelos CNN pré-treinados num grande conjunto de dados foram ajustados de forma a permitir a classificação de lesões em imagens de mamografia antes e após a aplicação de um processo de normalização. Também foi avaliada a performance de uma SVM com a utilização de características extraídas das CNN isoladamente ou combinadas com “*handcrafted features*”. Os resultados obtidos são encorajadores e demonstram que tanto o uso da medida de dimensão fractal como das CNN pode ajudar a melhorar o diagnóstico automático de lesões em mamografia, o que se traduziu em valores de AUC até 81,3%.

#### Palavras Chave:

Mamografia, CAD, Dimensão Fractal, CNN, Transfer Learning, Deep Learning.

#### Abstract

Computer-Aided Systems can improve the performance of breast cancer diagnosis, helping the differentiation and classification of breast benign and malignant lesions. Breast lesions are strongly correlated with their shape, therefore, in this work, firstly, were used quantitative measures based on fractal dimension, which can help characterizing the smoothness or the roughness of mammographic lesion shape, being calculated through the box-counting method, directly from manually segmented lesions, and after applying a region growing/erosion algorithm. Then, Convolutional neural networks (CNN), which recently have been successfully used in the medical field to detect and classify pathologies in different imaging modalities, were applied in mammographic images. Since the mammographic databases have a restricted number of samples, one way to solve this problem is using a transfer learning approach, in which a CNN, previously pre-trained with a large amount of labelled non-medical data, is subsequently finetuned using a smaller dataset of mammographic images. In this study were applied three different pre-trained networks and were made an evaluation if their performance to classify lesions in mammograms is affected by the use, or not, of normalized images. Was also evaluated the performance of a support vector machine fed with features extracted from the CNN and the combined use of handcrafted features to complement the CNN-extracted features. The obtained results are encouraging and reveal that both use of fractal dimension and CNN can help to improve computer-aided diagnostic of mammographic lesions, with AUC values around 81.3%.

#### Keywords:

Mammography, CAD, Fractal Dimension, CNN, Transfer Learning, Deep Learning.

**Introdução:**

A mamografia é o exame de primeira linha no diagnóstico de lesões da mama e a sua interpretação é realizada por médicos radiologistas altamente treinados. No entanto, segundo Kohli e Saurabh (2017) [1], apenas são detetados 84% dos cânceros de mama, sendo que os restantes 16% sugerem que neste procedimento podem existir limitações técnicas e humanas, nomeadamente: 1) a incerteza relativamente ao diagnóstico e classificação da patologia, devido, por exemplo a uma variante anatómica; 2) a fadiga ocular, devido à análise de uma grande quantidade de imagens; 3) questões ambientais, como a iluminação, que pode afetar a visualização das estruturas que compõem a imagem da mama; 4) a falta de qualidade das imagens, que podem ser prejudicadas, por exemplo, por ruído ou por mau posicionamento da paciente, 5) a falta de estudos comparativos e, em alguns casos, 6) a falta de experiência do médico radiologista [1–3].

Para colmatar estas limitações, muitos autores recomendam a dupla visualização das imagens, por exemplo, por dois médicos especialistas, de forma a melhorar a qualidade do diagnóstico e a reduzir a quantidade de falsos negativos [4,5]. No entanto, estas recomendações apresentam limitações, tais como a falta de clínicos e o aumento de custos para os centros de diagnóstico.

Surgem assim os sistemas *Computer Aided Detection/Diagnosis* (CAD), desenvolvidos com o objetivo primordial de auxiliar na deteção e discriminação de lesões benignas e malignas, com uma elevada precisão e uma admissível quantidade de falsos-positivos [1,6]. No entanto, o benefício da utilização destes sistemas em mamografia ainda permanece uma incógnita e, apesar de serem utilizados rotineiramente, como segunda opinião, nos programas de rastreio do cancro de mama nos EUA [2], segundo a análise de Kholi e Saurabh (2017) [1] por cada verdadeiro-positivo detetado através do recurso a CAD, existem mais falsos-positivos que têm de ser ignorados pelos radiologistas. O que fez com que esta ferramenta fosse por vezes pouco valorizada, uma vez que a sua utilização trouxe dúvidas na interpretação das imagens, existindo um conseqüente aumento na taxa de repetição de exames e no número biópsias efetuadas [6].

Com a evolução tecnológica surgiram novos algoritmos que prometem revolucionar os sistemas de deteção e classificação automática iniciais. Todos os dias são publicados novos trabalhos de investigação que pretendem transferir para a área da imagem médica, os sistemas de Inteligência Artificial utilizados na análise de dados visuais aplicados a outro tipo de tarefas, os quais apresentam resultados promissores, como é o caso das técnicas de *Machine Learning* ou das *Artificial Neural Networks* [1,2,7].

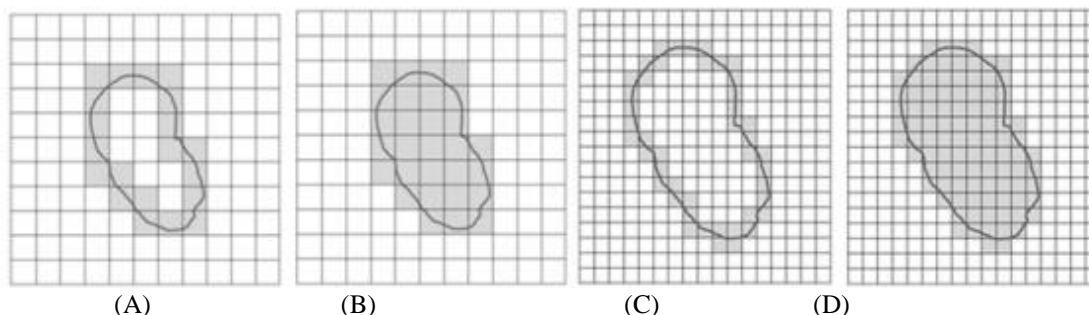
Tal como foi referido anteriormente e considerando que a investigação na área do diagnóstico médico é de extrema importância na evolução da medicina, neste trabalho pretendeu-se testar a aplicação de dois métodos diferentes na classificação de lesões em imagens de mamografia. Assim, na secção 2 são apresentados os materiais e métodos utilizados, seguindo-se a apresentação e discussão dos resultados obtidos e, por último, as conclusões do estudo.

**Materiais e Métodos:**

Numa primeira fase, pretendeu-se avaliar se as medidas baseadas na Dimensão Fractal (DF) poderiam melhorar o desempenho na classificação de lesões em mamografia de acordo com a escala BIRADS e o tipo de lesão. Para tal foram utilizadas 230 imagens de pacientes do sexo feminino (98,3%) e do sexo masculino (1,7%), com idades compreendidas entre os 24 e os 89 anos, com média de idades de 55,2 anos. As imagens foram adquiridas através do mesmo equipamento/tecnologia e digitalizadas em formato DICOM a partir de um sistema KODAC®, com dimensão de matriz de 32,7 Megabytes ( $3584 \times 4784$  pixels), com aproximadamente 20,4 pixels por milímetro, dimensão do *pixel* de  $0,05 \times 0,05$  mm<sup>2</sup> e profundidade de 16 bits por *pixel*. As imagens foram segmentadas manualmente tendo como referência os relatórios efetuados por uma médica radiologista com elevada experiência na área do diagnóstico de patologias mamárias.

Posteriormente, foi aplicado um método *box-counting*, o qual consiste em dividir a imagem em quadrados progressivamente mais pequenos e contar o número destes que contêm pelo menos um *pixel* pertencente ao contorno ou à área da lesão, permitindo assim calcular dois valores de dimensão fractal, respetivamente *DF de contorno* (DF0) e *DF de área* (DF1). As medidas de DF são calculadas através da Equação 1, na qual  $N(s)$  é o número de quadrados contabilizados e  $1/s$  é o fator de magnificação para cada partição (de 2 a 64 pixels). O processo de contagem dos quadrados é repetido para os diferentes tamanhos de quadrados ( $1/s$ ), tal como é exemplificado na Figura 1.

$$D = \lim_{s \rightarrow 0} \frac{\log(N(s))}{\log(1/s)} \quad (1)$$



**Figura 1** – Ilustração esquemática da diferença na contabilização dos quadrados pertencentes ao contorno (A e C) e à área da lesão (B e D) para dois tipos de partição (A-B, C-D).

Seguidamente, foram obtidas duas medidas, iguais às anteriores, mas calculadas após a aplicação de um algoritmo de dilatação/erosão – *DF de contorno após a aplicação do algoritmo de dilatação/erosão (DFODE)* e *DF de área após a aplicação do algoritmo de dilatação/erosão (DFIDE)*; este tipo de algoritmos é utilizado para, através da aplicação de sucessivas interações – baseadas em regras simples entre *pixels* –, obter estimativas da forma da lesão caso esta fosse maior ou menor. Neste trabalho foram utilizadas 20 iterações tendo em conta os 8 vizinhos mais próximos de cada *pixel*. No caso de uma lesão com uma forma regular, os valores de DF inicial e final deveriam ser similares, caso contrário seriam esperados valores significativamente diferentes.

A última medida avaliada baseia-se na diferença normalizada entre as duas medidas DF de área antes e após a aplicação do algoritmo de dilatação/erosão (*DF2*), de acordo com a Equação 2. O tratamento e análise dos dados foram realizados com recurso ao programa estatístico SPSS versão 19.

$$DF2 = \frac{DF1DE - DF1}{DF1} \quad (2)$$

Na segunda fase do estudo foram aplicadas CNN através do método de *transfer learning*, o qual se baseia na utilização de redes neuronais pré-treinadas em diferentes tarefas, as quais são posteriormente ajustadas ao problema da classificação de lesões em mamografia [8]. Esta abordagem é justificada pela necessidade de ultrapassar o problema da grande quantidade de dados que a aplicação de uma CNN requer para que a rede seja adequadamente treinada. Esta é uma grande limitação na área da imagem médica, uma vez que é difícil obter grandes quantidades de dados, o que se deve não só à dificuldade em aceder às imagens, atualmente ainda mais agravada pelas leis da proteção de dados, como também ao enorme trabalho e dispêndio de tempo necessário para catalogar as imagens com a respetiva classificação.

Desta forma, as três redes neuronais pré-treinadas utilizadas neste trabalho foram o modelo *CNN-F (Fast)*, o modelo *CNN-M (Medium)* [9], e o modelo *Caffe* [9,10]. O treino destas três redes encontrava-se previamente realizado com recurso às imagens do *ImageNet ILSVRC Challenge Data*, tendo posteriormente sido ajustado de forma a que as mesmas pudessem ser aplicadas na classificação de imagens mamográficas de lesões benignas e malignas. Para isso, adaptou-se o software *MatConvNet* [11] que se encontra disponível para o *Matlab*.

A base de dados utilizada neste estudo, para ajuste das três redes neuronais pré-treinadas, foi a *Breast Cancer Digital Repository (BCDR)*, especificamente a *BCDR-FM* [12], ficheiro *BCDR-F03*.

A etapa de pré-processamento das imagens neste trabalho é similar à estratégia utilizada pelos autores supramencionados, nomeadamente: 1) seleção de uma ROI de  $150 \times 150$  *pixels* utilizando a informação do *bounding box* das regiões segmentadas, sendo sempre preservadas as suas proporções, mesmo quando o tamanho das lesões é superior ao tamanho do ROI utilizado. Quando a lesão se encontra próxima dos bordos da imagem, optou-se por mover o quadrado do corte, alterando as coordenadas da imagem e incluindo o tecido mamário envolvente, em vez do *zero-padding* na porção da ROI que se encontrava fora da imagem; 2) *data augmentation* utilizando uma combinação de operações de rotação ( $90^\circ$ ,  $180^\circ$  e  $270^\circ$ ) e de reflexão horizontal.

Também a normalização das imagens foi similar à utilizada por Arevalo *et al.* (2016) [12], a qual consiste numa operação realizada de forma global (através da subtração da média das intensidades da imagem em cada *pixel*) em combinação com uma operação local (uma filtragem passa-alto realizada na vizinhança de cada voxel).

As imagens foram divididas em três grupos: 50% para treino, 10% para validação e 40% para teste, com uma dimensão de entrada de  $224 \times 224$  *pixels*. Os parâmetros explorados foram a dimensão da *batch*, os valores da taxa de aprendizagem, os modelos pré-treinados e o uso - ou não - de imagens normalizadas.

Após o ajuste dos três modelos pré-treinados e utilizando os conjuntos de treino e validação (com 2800 e 560 imagens, respetivamente) com e sem normalização, foram selecionados os melhores parâmetros para aplicar ao conjunto de teste (2240 imagens).

Posteriormente, foi selecionada a rede que obteve o melhor desempenho e foram extraídos os valores (*activations*) registados em uma - ou em todas - das últimas camadas, da 16 à 19, os quais foram utilizados para treinar uma SVM. Desta forma, pretendeu-se verificar se o desempenho de classificação poderia ser melhorado através da utilização de outro tipo de classificador. As *activations* foram utilizadas isoladamente, em combinações duas-a-duas ou três-a-três, ou todas em conjunto. Por fim, foram utilizadas também características diversas, de intensidade, forma e textura [12], as quais são habitualmente designadas por *hand crafted features* (HCF). Estas foram fornecidas pelos autores da base de dados e destinaram-se a verificar novamente se a sua utilização conduziria a melhorias nos resultados da classificação. Para além disso, foi ainda avaliada a inclusão de informação sobre a idade e densidade mamária das pacientes.

## Resultados:

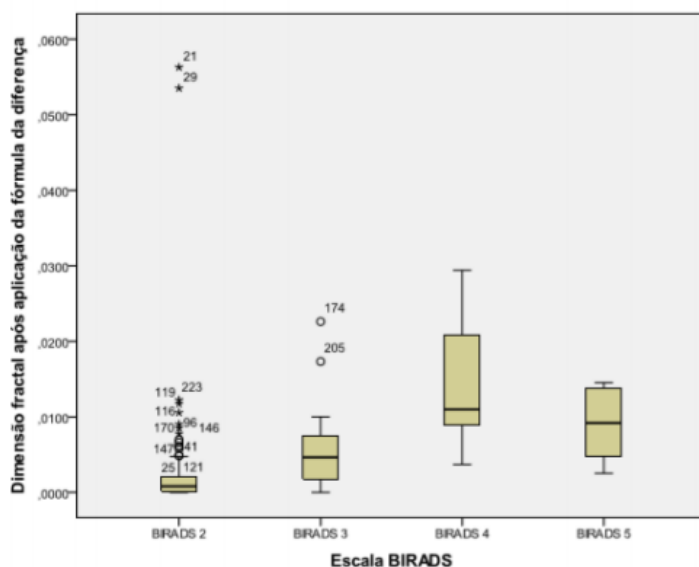
### 1.1. Dimensão Fractal

Após a verificação de que, em pelo menos um dos tipos de lesão, os valores de DF não seguem uma distribuição normal, foi aplicado o teste não-paramétrico de *Kruskall-Wallis* (K-W) para comparação de *k*-grupos independentes, o qual permitiu verificar que apenas para os valores de DF1 não existem diferenças significativas em função da escala BIRADS das lesões (valor  $p=0.081$ ).

Assim, foi aplicado às restantes medidas, o teste de *Mann-Whitney* (M-W) para comparações múltiplas, de forma a comparar a DF das diferentes escalas BIRADS duas-aduas. Os resultados demonstram que a utilização combinada das diferentes medidas DF permite, pelo menos uma vez, a diferenciação das diferentes categorias da escala BIRADS entre si. A medida que apresenta melhores resultados nesta diferenciação é a DF2, que apenas é incapaz de fazer a distinção entre as escalas BIRADS 4 e 5, o que pode dever-se à reduzida dimensão da amostra ou à semelhança entre este tipo de lesões, pois ambas apresentam lesões tipicamente irregulares.

A diferença da DF de área antes e após a aplicação do algoritmo de dilatação/erosão (DF2) tem tendência a aumentar com a escala BIRADS, como se pode verificar através do gráfico apresentado na

Figura 2, o qual indica que, com o algoritmo de dilatação/erosão, a forma da lesão sofre uma variação superior no caso das lesões mais irregulares, uma vez que a diferença entre ambos os valores de DF de área é superior neste tipo de lesões consideradas malignas. Assim, podemos afirmar que a diferença entre a DF de área antes e após a aplicação do algoritmo de dilatação/erosão pode ser utilizada na diferenciação de lesões benignas regulares de lesões malignas irregulares, apresentado resultados mais satisfatórios que a DF1 ou DF1DE testadas individualmente.



**Figura 2** - Comparação dos valores médios e desvios padrão da DF2 consoante a escala BIRADS.

Posteriormente, através da aplicação do teste K-W foi possível verificar que os valores de DF0 e DF1 diferem significativamente consoante o tipo de lesão, pelo que foi aplicado o teste de M-W para comparação das medidas DF0 e DF1 dos diferentes tipos de lesão, o qual demonstrou que não é possível utilizar estas medidas, por si só, na diferenciação de todos os tipos de lesão entre si, no entanto, a medida que apresenta melhores resultados é a DF0. No entanto, como se pode verificar na Tabela 2, o uso combinado destas medidas reduz o número de tipos de lesão que permanecem indistinguíveis, os quais são indicados pelas áreas mais escuras contendo o símbolo (=)|(=). Os resultados obtidos não apresentam diferenças significativas no que diz respeito à diferenciação de áreas de maior densificação de outras lesões, tais como áreas de encapsulamento esboçando nodularidade, densificações irregulares, nódulos com contornos espiculados e nódulos, o que se pode dever à reduzida dimensão da amostra de alguns destes tipos de lesões (AMD  $n=5$ ; DI  $n=15$  e DCE  $n=5$ ).

**Tabela 2** - Comparação dos resultados do teste de M-W relativos ao tipo de lesão, obtidos através do uso combinado das medidas DF0 e DF1.

DF0/DF1	M	Enc.	C	AMD	DI	NCE
BN	(≠) (≠)	(=) (≠)	(≠) (=)	(=) (=)	(≠) (=)	(≠) (=)
M		(≠) (≠)	(≠) (≠)	(≠) (≠)	(≠) (≠)	(≠) (≠)
Enc.			(≠) (≠)	(=) (=)	(≠) (=)	(≠) (=)
C				(=) (≠)	(≠) (≠)	(=) (≠)
AMD					(=) (=)	(=) (=)
DI						(=) (=)

Abreviaturas: AMD – Área de maior densificação; C – Calcificações; DI – Densificação Irregular; Enc. – Encapsulamento esboçando nodularidade; M – Microcalcificações; N – Nódulos; NE – Nódulos com contornos espiculados

## 1.2. CNN

Relativamente à aplicação de modelos CNN, como referido na secção anterior, em primeiro lugar foram realizados testes com os conjuntos de treino e validação, de forma a definir os melhores parâmetros a aplicar. Nesta fase, os resultados demonstraram que, para as imagens *com normalização*, o melhor valor médio de AUC após 5 repetições, foi obtido utilizando o modelo *Caffe* (AUC média =  $0.775 \pm 0.014$ ); no caso das imagens *sem normalização*, o melhor valor médio de AUC foi obtido pelo modelo CNN-M (AUC média =  $0.785 \pm 0.003$ ). Após a determinação da melhor combinação de parâmetros para cada modelo, foram obtidos novos resultados com o conjunto de teste, os quais se encontram apresentados na Tabela 3.

**Tabela 3** - Resultados da aplicação dos três modelos CNN pré-treinados a imagens de mamografia com e sem normalização (valores a negrito correspondem ao melhor resultado obtido para cada modelo).

Rede	Dimensão da batch	Épocas	Taxa de Aprendizagem	Norm.	AUC média	Desvio Padrão	Tempo aprox. (min)
CNN-F	256	50	5e-3	Sim	0.721	0.001	28.31
CNN-F	256	50	1e-2	Não	<b>0.767</b>	0.002	25.81
CNN-M	64	50	5e-4	Sim	0.733	0.002	76.84
CNN-M	64	50	5e-4	Não	<b>0.776</b>	0.006	76.25
Caffe	256	50	5e-2	Sim	0.584	0.004	27.28
Caffe	256	50	5e-3	Não	<b>0.813</b>	0.001	25.24

O melhor desempenho foi obtido através do modelo *Caffe* nas imagens em que *não* foi aplicado o processo de normalização, com um valor médio de AUC de  $0.813 \pm 0.001$ , ultrapassando o valor de 0.79 alcançado por Arevalo *et al.* (2016) [12], o qual foi obtido com o uso combinado do modelo *DeCAF*, de imagens normalizadas e da substituição do classificador *Softmax* por uma SVM (uma vez que consideram que esta apresenta um melhor desempenho na classificação do que o *Softmax* que é parte integrante destes modelos pré-treinados).

Relativamente ao tempo de processamento das redes, o modelo *Caffe* foi o mais rápido, com 22.66 minutos para o treino e 2.58 minutos para o teste, totalizado 25.24 minutos; no geral, conseguiu-se verificar que o tempo necessário para a aprendizagem e posterior classificação aumenta quando são usadas as imagens após normalização.

Tal como aconteceu com o conjunto de validação, as melhores performances foram obtidas quando foram utilizadas as imagens sem normalização, com valores médios de AUC de 0.776 para a CNN-M e 0.767 para a CNN-F, valores esses que são similares aos alcançados na fase de ajuste. No geral, os valores médios de AUC obtidos com as imagens após normalização são mais baixos comparativamente com os que foram alcançados com o conjunto de validação, sendo significativamente mais baixo com o modelo *Caffe*, (AUC média=0.584; anteriormente 0.775).

A utilização das *activations* provenientes das últimas camadas e de um classificador SVM foi feita tendo por base a rede que apresentou os melhores resultados nos testes anteriores – o modelo *Caffe*. Os resultados são apresentados na Tabela 4.

**Tabela 4** - Resultados da classificação de lesões com a recurso a uma SVM (valores médios obtidos após 5 repetições). Resultados a negrito correspondem aos melhores valores médios de AUC.

<i>Camada de Ativação</i>	<i>AUC média</i>
<b>16</b>	<b>0.773</b>
<b>17</b>	0.749
<b>18</b>	0.766
<b>19</b>	0.732
<b>16&amp;17</b>	0.767
<b>16&amp;18</b>	<b>0.772</b>
<b>16&amp;19</b>	0.760
<b>17&amp;18</b>	0.760
<b>17&amp;19</b>	0.744
<b>18&amp;19</b>	0.755
<b>16&amp;17&amp;18</b>	0.769
<b>16&amp;17&amp;19</b>	0.761
<b>16&amp;18&amp;19</b>	0.762
<b>17&amp;18&amp;19</b>	0.754
<b>16&amp;17&amp;18&amp;19</b>	0.763

No geral, os valores médios de AUC são mais baixos comparativamente com os resultados obtidos apenas com a CNN e o melhor desempenho foi alcançado quando foram utilizadas as características extraídas da camada 16 (AUC média=0.773), e na combinação das características extraídas das camadas 16 e 18 (AUC média=0.772).

Os resultados referentes à combinação das *activations* extraídas da camada 16 com as HCF e a informação da idade e densidade mamária das pacientes são apresentados na Tabela 5.

**Tabela 5** - A influência da idade, densidade e HCF em combinação com as características extraídas da camada 16 na performance da classificação com uma SVM (valores médios após 5 repetições). Resultados a negrito correspondem aos melhores valores médios de AUC.

	<i>Isoladamente</i>	<i>Idade</i>	<i>Densidade</i>	<i>Idade&amp;Densidade</i>
<b>Idade</b>	0.766	--	--	--
<b>Densidade</b>	0.773	--	--	--
<b>Idade&amp;Densidade</b>	0.764	--	--	--
<b>Intensidade (I)</b>	0.774	0.766	0.772	0.764
<b>Forma (F)</b>	<b>0.796</b>	0.788	<b>0.796</b>	0.788
<b>Textura (T)</b>	0.782	0.775	0.780	0.773
<b>I&amp;F</b>	<b>0.796</b>	0.789	<b>0.796</b>	0.788
<b>I&amp;T</b>	0.779	0.772	0.778	0.770
<b>F&amp;T</b>	<b>0.801</b>	0.794	<b>0.800</b>	0.793
<b>I&amp;F&amp;T</b>	<b>0.799</b>	0.794	0.799	0.792

De notar que, de uma forma geral, os valores médios de AUC são similares ou superiores aos anteriores. A melhor performance foi alcançada com a combinação das características da forma e textura (AUC média=0.801), forma, textura e densidade (AUC média=0.800), intensidade, forma e textura (AUC média=0.799), e intensidade, forma, textura e densidade (AUC média=0.799). As características de forma e textura parecem ser as HCF que apresentam uma maior influência no processo de classificação (AUC média=0.796 e 0.782, respetivamente), confirmando que formas irregulares e texturas heterogéneas estão altamente relacionadas com a malignidade das lesões da mama [13,14]. O uso isolado da idade, densidade e intensidade não parece ter grande influência na performance da classificação, uma vez que não existe um aumento nos valores médios da AUC, e quando combinados com outras características, por vezes pioram os resultados da mesma.

### Conclusão:

O uso combinado das diferentes medidas DF pode ser uma forma viável de classificar as lesões mamárias segundo a sua escala BIRADS ou diferenciá-las segundo o tipo de lesão. As melhores performances na distinção de lesões foram obtidas através da utilização da medida DF2, proposta neste estudo, e DF0, já implementada por outros investigadores como Crisan *et al.* (2007) [15] e Nguyen *et al.* (2005) [16], a qual aumenta com a escala BIRADS e sofre alterações consideráveis consoante o tipo de lesão. Esta diferenciação de lesões é ainda mais evidente quando se combinam as diferentes medidas DF utilizadas neste estudo.

Paralelamente, na aplicação das CNN na classificação das lesões, quando as imagens foram fornecidas à rede sem qualquer tipo de normalização, o modelo *Caffe* atingiu a melhor performance, com uma AUC média de 0.813, seguido pela CNN-M e pela CNN-F, com 0.776 e 0.767 respetivamente. Quando foram utilizadas imagens com normalização, com o conjunto de teste, os resultados baixaram substancialmente comparativamente com os resultados previamente obtidos com o conjunto de validação, principalmente no caso do modelo *Caffe*, o qual obteve uma AUC média de 0.584. Os resultados obtidos demonstram que os métodos de normalização propostos por Arevalo *et al.* (2016) [12] reduziram a performance da classificação das redes utilizadas neste estudo. Perre *et al.* (2018) [17] consideram que, os métodos de normalização escolhidos podem alterar a performance da classificação dependendo do tipo de rede aplicado.

O facto de os ROI utilizados conterem o padrão do tecido mamário (em vez de ser aplicado um *zero-padding*) e, em alguns casos, o facto de a lesão não se encontrar centrada, pode ter sido uma vantagem no processo de aprendizagem da CNN sem fatores de confundimento.

Após a extração das diferentes características a partir modelos CNN pré-treinados, na classificação das lesões com recurso a uma SVM a melhor performance foi obtida com as características extraídas da camada 16, com uma AUC média de 0.773. No entanto, os resultados desta classificação foram mais baixos comparativamente com os resultados obtidos com o modelo pré-treinado *Caffe* isoladamente, o que não é concordante com a opinião de Arevalo *et al.* (2016) [12], uma vez que estes autores consideram que a SVM é mais eficaz no processo de classificação comparativamente com o *Softmax* que é o classificador geralmente utilizado nas CNN. A combinação entre as HCF e as características extraídas pela rede aumentou a performance de classificação da SVM. O uso combinado de toda a informação permitiu atingir um valor médio de AUC de 0.799. No entanto, a forma e textura das lesões foram as

características que permitiram atingir o melhor resultado no processo de classificação, com uma AUC média de 0.801. A informação da densidade mamária das pacientes demonstrou uma grande influência na classificação quando combinada com as características de forma e textura por exemplo, com uma AUC média de 0.800. As características de intensidade das lesões e a idade das pacientes não demonstraram grande influência na performance da classificação.

Apesar de existirem imensos estudos realizados tanto na área da mamografia como também em outras áreas da imagem médica, vários temas continuam por explorar e os avanços tecnológicos irão com certeza permitir alcançar resultados encorajadores no diagnóstico assistido por computadores, que permitam melhorar a deteção e classificação precoce de neoplasias, aumentando assim a taxa de sobrevivência de doentes oncológicos em todo o mundo.

### Referências:

1. A. Kohli, S. Jha, Why CAD Failed in Mammography, *J. Am. Coll. Radiol.* 15 (2018) 535–537. doi:10.1016/j.jacr.2017.12.029.
2. M.I. Fazal, M.E. Patel, J. Tye, Y. Gupta, The past, present and future role of artificial intelligence in imaging, *Eur. J. Radiol.* 105 (2018) 246–250. doi:10.1016/j.ejrad.2018.06.020.
3. C. Rodrigues, *Sistemas CAD em Patologia Mamária*, Universidade do Porto, 2008.
4. G. Giannakopoulou, G.M. Spyrou, A. Antarakis, I. Andreadis, D. Koulocheri, F. Zagouri, A. Nonni, G.M. Filippakis, K.S. Nikita, P.A. Ligomenides, G.C. Zografos, Downgrading BIRADS 3 to BIRADS 2 category using a computer-aided microcalcification analysis and risk assessment system for early breast cancer, *Comput. Biol. Med.* 40 (2010) 853–859. doi:10.1016/j.combiomed.2010.09.005.
5. P.C. Brennan, A. Ganesan, M.P. Eckstein, E.U. Ekpo, K. Tapia, C. Mello-Thoms, S. Lewis, M.Z. Juni, Benefits of Independent Double Reading in Digital Mammography: A Theoretical Evaluation of All Possible Pairing Methodologies, *Acad. Radiol.* (2018). doi:https://doi.org/10.1016/j.acra.2018.06.017.
6. J.D. Keen, J.M. Keen, J.E. Keen, Utilization of Computer-Aided Detection for Digital Screening Mammography in the United States, 2008 to 2016, *J. Am. Coll. Radiol.* 15 (2018) 44–48. doi:10.1016/j.jacr.2017.08.033.
7. E.P. V Le, Y. Wang, Y. Huang, S. Hickman, F.J. Gilbert, Artificial intelligence in breast imaging, *Clin. Radiol.* 74 (2019) 357–366. doi:10.1016/j.crad.2019.02.006.
8. L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, Y. Xie, A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis., *Comput. Math. Methods Med.* (2019) 1–16. doi:10.1155/2019/6509357.
9. K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the Devil in the Details: Delving Deep into Convolutional Nets, (2014). <http://arxiv.org/abs/1405.3531>.
10. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional Architecture for Fast Feature Embedding, (2014). <http://arxiv.org/abs/1408.5093>.
11. A. Vedaldi, K. Lenc, MatConvNet - Convolutional Neural Networks for MATLAB, (2014). <http://arxiv.org/abs/1412.4564>.
12. J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A. Guevara Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, *Comput. Methods Programs Biomed.* 127 (2016) 248–257. doi:http://dx.doi.org/10.1016/j.cmpb.2015.12.014.
13. J.M. Pisco, *Imagiologia Básica – Texto e Atlas*, 1ª Edição, Lisboa, 2003.
14. D. Yi, R.L. Sawyer, D. Cohn III, J. Dunnmon, C. Lam, X. Xiao, D. Rubin, Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors, 2017. <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsarx&AN=1705.06362>.

15. D.A. Crisan, R. Dobrescu, P. Planinsi, Mammographic Lesions Discrimination Based on Fractal Dimension as an Indicator, in: 2007 14th Int. Work. Syst. Signals Image Process. 6th EURASIP Conf. Focus. Speech Image Process. Multimed. Commun. Serv., IEEE, 2007: pp. 74–77. doi:10.1109/IWSSIP.2007.4381156.
16. T.M. Nguyen, R.M. Rangayyan, Shape Analysis of Breast Masses in Mammograms via the Fractal Dimension, in: 2005 IEEE Eng. Med. Biol. 27th Annu. Conf., IEEE, 2005: pp. 3210–3213. doi:10.1109/IEMBS.2005.1617159.
17. A.C. Perre, L.A. Alexandre, L.C. Freire, Exploring different normalization and classification approaches for mammography analysis with CNNs, 2018. doi:10.3233/978-1-61499-929-4-315.

Recebido / Received: 17/06/2021

Aceite / Accept: 19/07/2021