



# **Avaliação da Caracterização de Lesões em Mamografia com Recurso a Sistemas CAD (Diagnóstico Assistido por Computador)**

## **Sistemas CAD em Mamografia**

Ana Catarina Filipe Perre

Trabalho de natureza profissional apresentado à Escola Superior de Saúde Dr. Lopes Dias do Instituto Politécnico de Castelo Branco para cumprimento dos requisitos necessários à obtenção do Título de Especialista na área das Tecnologias de Diagnóstico e Terapêutica – Radiologia, de acordo com o Decreto-Lei n.º 206/2009 de 31 de agosto.

**Junho de 2019**







## Resumo

Os sistemas *Computer-Aided Detection/Diagnosis* (CAD) auxiliam os médicos na detecção e diferenciação de lesões benignas e malignas, aumentando a performance no diagnóstico do cancro da mama. As lesões da mama estão fortemente correlacionadas com a forma do contorno: lesões benignas apresentam contornos regulares, enquanto as lesões malignas tendem a apresentar contornos irregulares. Desta forma, neste estudo, em primeiro lugar, foram utilizadas medidas quantitativas baseadas na dimensão fractal (DF), as quais podem ajudar na caracterização dos contornos regulares ou irregulares das lesões encontradas nas imagens de mamografia, sendo calculadas através da aplicação do método “*box-counting*”, diretamente em imagens de lesões segmentadas e após a aplicação de um algoritmo de dilatação/erosão. Seguidamente, foi testada a aplicação de Redes Neurais de Convolução (CNN), as quais têm sido recentemente utilizadas, com elevado grau de sucesso, na detecção e classificação de patologias em diferentes modalidades de imagem médica, incluindo a mamografia. Uma vez que as bases de dados de mamografia disponíveis apresentam um número reduzido de amostras, existe uma forma de ultrapassar esta limitação através do uso do método de “*transfer learning*”, no qual uma rede previamente treinada num grande conjunto de imagens comuns, é posteriormente ajustada para ser aplicada a uma base de dados com imagens de mamografia. Foram utilizados três modelos CNN pré-treinados, tendo sido avaliada a sua performance na classificação de lesões em imagens antes e após a aplicação de um processo de normalização. Também foi avaliada a performance de uma *Support Vector Machine* (SVM) com a utilização de características extraídas das CNN isoladamente ou combinadas com “*handcrafted features*”. Os resultados obtidos são encorajadores e podem ajudar a melhorar o diagnóstico automático de lesões em mamografia.

## Palavras chave

Mamografia, Imagem de Mamografia, Cancro de Mama, *Diagnóstico assistido com computador (CAD)*, Dimensão Fractal, *Redes Neurais de Convolução (CNN)*, *Transfer Learning*, *Support Vector Machine (SVM)*, *Deep Learning*



## **Abstract**

Computer-Aided Systems can assist differentiation and classification of breast benign and malignant lesions enhanced the performance of breast cancer diagnosis. Breast lesions are strongly correlated with their shape: benign lesions present regular shape, although malignant lesions tend to present irregular shape. Therefore, in this work, firstly, were used quantitative measures based on fractal dimension (FD), 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 (CNNs) application were tested, which have recently been successfully used in the medical field to detect and classify pathologies in different imaging modalities, including in mammography. 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. Were applied three different pre-trained networks and evaluated 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 (SVM) 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 can help to improve computer-aided diagnostic of mammographic lesions.

## **Keywords**

*Mammography, Mammographic image, Breast Cancer, Computed-Aided Diagnosis (CAD), Fractal Dimension, Convolutional Neural Network (CNN), Transfer Learning, Support Vector Machine (SVM), Deep Learning*



# Índice geral

1. Introdução.....	1
2. Enquadramento Teórico .....	3
2.1. Mamografia .....	3
2.1.1. Sistema de Classificação BIRADS.....	4
2.1.2. Bases de Dados em Mamografia .....	7
2.2. <i>Computer Aided Detection/Diagnosis</i> .....	13
2.2.1. Dimensão Fractal.....	15
2.2.2. Redes Neurais de Convolução (CNN) .....	17
2.2.3. Processamento de Imagem.....	21
3. Metodologia.....	23
3.1. Dimensão Fractal .....	23
3.2. Redes Neurais de Convolução.....	26
4. Resultados e Discussão .....	29
4.1. Dimensão Fractal .....	29
4.4.1. DF versus Escala BIRADS .....	29
4.4.2. DF0 e DF1 versus Tipo de Lesão.....	31
4.2. Redes Neurais de Convolução.....	34
5. Conclusão .....	38
6. Perspetivas Futuras.....	39
7. Bibliografia .....	40



## Índice de figuras

<b>Figura 1</b> – Imagens de mamografia com a respetiva descrição das estruturas visualizadas.....	3
<b>Figura 2</b> – Imagens de mamografia (incidências básicas de rotina – CCD, CCL, OMLD e OMLE). Fonte: <a href="http://newsroom.gehealthcare.com/wp-content/uploads/2017/10/CLINICAL-IMAGE-Senographe-Pristina-0128_econtrast4.png">http://newsroom.gehealthcare.com/wp-content/uploads/2017/10/CLINICAL-IMAGE-Senographe-Pristina-0128_econtrast4.png</a> .....	4
<b>Figura 3</b> – Exemplos de imagens mamográficas com diferentes densidades radiológicas, incidências CCD e OMLD, da esquerda para a direita, tipo I, II, III e IV. Fonte: Própria. ....	5
<b>Figura 4</b> – Grau de malignidade dos nódulos mamários relativamente à sua forma e contorno. ....	5
<b>Figura 5</b> – Exemplos de calcificações mamárias que podem ou não ser suspeitas de malignidade.....	6
<b>Figura 6</b> – Exemplos de imagens de mamografia extraídas da base de dados MIAS – Incidência OMLD e OMLE, ambas consideradas normais. ....	8
<b>Figura 7</b> - Exemplos de imagens de mamografia extraídas da base de dados DDSM – Incidência CCD e OMLE, ambas consideradas normais. ....	9
<b>Figura 8</b> – Exemplos de imagens de mamografia extraídas da base de dados BCDR-FM - Incidências CCD e MLOE, ambas com lesões malignas com segmentação da área da lesão. ....	10
<b>Figura 9</b> – Exemplos de imagens de mamografia extraídas da base de dados BCDR-DM – Incidências CCE e OMLE, ambas benignas.....	11
<b>Figura 10</b> – Exemplos de imagens de mamografia extraídas da base de dados INBreast -Incidências CCD e OMLD, ambas benignas.....	12
<b>Figura 11</b> - Ilustração Sistemática de um sistema CAD-Dx. ....	14
<b>Figura 12</b> - Similaridade entre a) tecido mamário e b) fractais gerados sinteticamente (36).....	15
<b>Figura 13</b> - Koch Snowflake fractal com comprimento infinito e área zero (39). ....	15
<b>Figura 14</b> - O Koch Snowflake coberto por quadrados com cada vez menor dimensão. Quanto menor a área de cobertura de cada quadrado, mais detalhes serão considerados (39).....	16

<b>Figura 15</b> - Exemplo de uma Curva log-log adquirida através do método <i>Box Counting</i> (39). A dimensão fractal calculada é, neste exemplo, 1,2612.....	16
<b>Figura 16</b> - Gráfico com as diferentes modalidades que utilizam sistemas CAD (40). .....	17
<b>Figura 17</b> - Gráfico das diferentes MLTs utilizadas nos sistemas CAD para estudo do cancro da mama (40). .....	18
<b>Figura 18</b> - Representação da estrutura de uma rede neuronal (8).....	19
<b>Figura 19</b> - Conceito de “ <i>Transfer Learning</i> ”. No “ <i>source model</i> ” o modelo é treinado com uma quantidade suficiente de dados e labels de elevada qualidade e o método de <i>transfer learning</i> permite que, após ajustes, o “ <i>target model</i> ” seja aplicado na área de estudo pretendida com menor quantidade de dados, o que é benéfico principalmente na área da imagem médica (33). .....	20
<b>Figura 20</b> - Representação esquemática da arquitetura da rede VGG16: 13 “ <i>convolutional layers</i> ”, 3 “ <i>full-connettion layers</i> ”, 5 “ <i>pooling layers</i> ” e 1 “ <i>softmax layer</i> ” (33). .....	20
<b>Figura 21</b> - 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).....	24
<b>Figura 22</b> - Exemplo da aplicação do algoritmo de dilatação/erosão a uma lesão considerada benigna (A) e a uma lesão considerada maligna (B); Imagens da esquerda: Lesão inicial; Imagens centrais: Lesão após dilatação; Imagens da direita: Lesão após erosão aplicada às lesões das imagens centrais. ....	25
<b>Figura 23</b> - Comparação dos valores médios e desvios padrão da <b>DF2</b> consoante a escala <b>BIRADS</b> . .....	31
<b>Figura 24</b> - Exemplo do gráfico com a curva ROC do melhor resultado obtido na aplicação do modelo Caffe em imagens sem normalização (AUC = 0.814).....	35

## Índice de tabelas

<b>Tabela 1</b> - Descrição das diferentes categorias da escala BIRADS (12). .....	7
<b>Tabela 2</b> – Bases de dados de mamografia mais utilizadas na literatura. ....	7
<b>Tabela 3</b> – Distribuição das lesões consoante a escala BIRADS. ....	23
<b>Tabela 4</b> - Distribuição das lesões consoante o tipo. ....	24
<b>Tabela 5</b> – Modelos CNN pré-treinados utilizados neste trabalho de investigação (adaptado de (50)). ....	27
<b>Tabela 6</b> – Indicação dos valores do coeficiente de Pearson, p, obtidos através do teste K-W para as diferentes medidas DF tendo em conta a escala BIRADS. ....	29
<b>Tabela 7</b> - Resultados do teste M-W (valores p) relativos à comparação da <b>DF0</b> das diferentes escalas BIRADS. ....	30
<b>Tabela 8</b> – Resultados do teste M-W (valores p) relativos à comparação da <b>DF0DE</b> das diferentes escalas BIRADS. ....	30
<b>Tabela 9</b> - Resultados do teste M-W (valores p) relativos à comparação da <b>DF1DE</b> das diferentes escalas BIRADS. ....	30
<b>Tabela 10</b> - Resultados do teste M-W (valores p) relativos à comparação da <b>DF2</b> das diferentes escalas BIRADS. ....	30
<b>Tabela 11</b> - Indicação dos valores do coeficiente de Pearson, p, obtidos através do teste K-W para as diferentes medidas DF, tendo em conta o tipo de lesão. ....	31
<b>Tabela 12</b> - Resultados do teste M-W (valores p) relativos à comparação da <b>DF0</b> dos diferentes tipos de lesão. ....	32
<b>Tabela 13</b> - Resultados do teste M-W (valores p) relativos à comparação da <b>DF1</b> dos diferentes tipos de lesão. ....	32
<b>Tabela 14</b> - Comparação dos resultados do teste de M-W relativos ao tipo de lesão, obtidos através do uso combinado das medidas <b>DF0 e DF1</b> . ....	32
<b>Tabela 15</b> – Tabela resumo dos testes realizados para definição dos melhores parâmetros a aplicar na classificação de lesões em imagens de mamografia. Valores de AUC média obtidos após 5 repetições, utilizando <b>imagens com normalização</b> (apenas foram utilizados os conjuntos de treino e validação). ....	34

**Tabela 16** - Tabela resumo dos testes realizados para definição dos melhores parâmetros a aplicar na classificação de lesões em imagens de mamografia, valores de AUC média obtidos após 5 repetições, utilizando **imagens sem normalização** (apenas foram utilizados os conjuntos de treino e validação).....34

**Tabela 17** - Resultados da aplicação dos três modelos CNN pré-treinados a imagens de mamografia com e sem normalização. O treino foi realizado com a fusão dos conjuntos de treino e validação, sendo testado no conjunto de teste. Os valores que se encontram a negrito correspondem ao melhor resultado obtido para cada modelo...35

**Tabela 18** - Resultados da classificação de lesões com a recurso a uma SVM após extração das características das imagens mamográficas sem normalização com o Caffe reference model (valores médios obtidos após 5 repetições). Resultados a negrito correspondem aos melhores valores médios de AUC obtidos.....36

**Tabela 19** - A influência da idade, densidade e HCFs (Intensidade, Forma e Textura) 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 obtidos.....37

## Lista de abreviaturas, siglas e acrónimos

ACR – *American College of Radiology*  
AI – *Artificial Intelligence/Inteligência Artificial*  
ANN – *Artificial Neural Networks*  
AUC – *Area Under the receiver operating Curve*  
BCDR – *Breast Cancer Digital Repository*  
BCDR-FM – *Film Mammography-based Repository*  
BCDR-DM – *Full-Field Digital Mammography-based Repository*  
BCDR-F03 – *Film Mammography Dataset Number 3*  
BIRADS – *Breast Imaging Reporting and Data System*  
CAD – *Computer Aided Detection/Diagnosis*  
CAdE – *Computer Aided Detection*  
CAD-Dx – *Computer Aided Detection/Diagnosis*  
CADx – *Computer Aided Diagnosis*  
CBIS-DDSM – *Curated Breast Imaging subset of DDSM*  
CC – *Crânio-Caudal*  
CCD – *Crânio-Caudal Direita*  
CCE – *Crânio-Caudal Esquerda*  
CLAHE – *Contrast-Limited Adaptive Histogram Equalization*  
CNN – *Convolutional Neural Networks/Redes Neurais de Convolução*  
CNN-F – *Fast, imagenet-vgg-f model*  
CNN-M – *Medium, imagenet-vgg-m model*  
Conv# – *Convolutional Layers*  
DDSM – *Digital Database for Screening Mammography*  
DGS – *Direção Geral de Saúde*  
DF – *Dimensão Fractal*  
DF0 – *DF de Contorno*  
DF0DE – *DF de Contorno após a aplicação do algoritmo de dilatação/erosão*  
DF1 – *DF de Área*  
DF1DE – *DF de Área após a aplicação do algoritmo de dilatação/erosão*  
DF2 – *DF Final – diferença normalizada entre os valores de DF1 e DF1DE*  
DICOM – *Digital Imaging and Communications in Medicine*  
DM – *Digital Mammography*  
DoG – *Difference of Gaussians*  
DP – *Deep Learning*  
EUA – *Estados Unidos da América*  
FDA – *Food and Drug Administration*  
FFDM – *Full-Field Digital Mammography*  
FN – *Falso Negativo*  
FP – *Falso Positivo*  
Full# – *Fully Connected Layers*  
GCN – *Global Contrast Normalization*

GPU – *Graphical Processing Units*  
HCFs – *Handcrafted Features*  
IRMA – *Image Retrieval in Medical Applications*  
kgf – *Quilograma-força*  
K-NN – *K-nearest neighbors*  
K-S – *Kolmogorov-Smirnov*  
K-W – *Kruskall-Wallis*  
LCN – *Local Contrast Normalization*  
LJPEG – *Lossless Joint Photographic Experts Group*  
LRN – *Local Response Normalization*  
MIAS – *Mammographic Image Analysis Society Digital Mammogram*  
MLT – *Machine Learning Techniques*  
M-W – *Mann-Whitney*  
N – *Newton*  
OML – *Oblíqua Médio-Lateral*  
OMLD – *Oblíqua Médio-Lateral Direita*  
OMLE – *Oblíqua Médio-Lateral Esquerda*  
pad – *Spatial Padding*  
PCA – *Principal Component Normalization*  
PGN – *Portable Gray Map*  
QSE – *Quadrante Superior Externo*  
ReLU – *Rectified Linear Unit*  
ROC – *Receiver Operating Characteristic*  
ROI – *Region of Interest/Região de Interesse*  
SN – *Sensibilidade*  
SP – *Especificidade*  
SPSS – *Statistical Package for the Social Sciences*  
st – *Convolutional Stride*  
std – *Standard Deviation/Desvio Padrão*  
SVM – *Support Vector Machine*  
S-W – *Shapiro-Wilk*  
TIFF – *Tagged Image File Format*  
TL – *Transfer Learning*  
VN – *Verdadeiro Negativo*  
VP – *Verdadeiro Positivo*  
vs – *Versus*  
XML – *Extensible Markup Language*





## 1. Introdução

Segundo a Direção Geral de Saúde (DGS), *“o cancro da mama é a neoplasia mais frequente do sexo feminino. Afeta 1 em cada 9 mulheres e constitui a causa mais frequente de mortalidade na faixa etária entre os 35 aos 55 anos de idade na União Europeia”*. Atualmente, graças a várias técnicas de diagnóstico e aos programas de rastreio implementados nas sociedades ditas desenvolvidas, é possível detetar a doença em fases cada vez mais precoces e, como tal, a taxa de sucesso do tratamento é bastante elevada – *“mais de 90% das doentes com cancro podem ser curadas, se diagnosticadas num estágio precoce e adequadamente tratadas”* (1).

A mamografia é considerada 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) (2), apenas 84% dos cancros de mama são detetados pelos médicos radiologistas, sendo que os 16% que não são detetados 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 (2–4).

Para colmatar estas limitações, muitos autores recomendam a dupla visualização das imagens, por exemplo, por dois médicos especialistas, de forma aumentar a qualidade do diagnóstico e reduzir a quantidade de falsos negativos (5,6). No entanto, esta alternativa apresenta limitações, como a falta de clínicos e o facto de acarretar mais custos para os centros de diagnóstico.

Surgem assim os sistemas *Computer Aided Detection/Diagnosis* (CAD), desenvolvidos com o objetivo primordial de auxiliar os médicos na deteção e discriminação entre lesões benignas e malignas com uma elevada precisão e uma admissível quantidade de falsos-positivos (2,7). No entanto, o benefício da utilização destes sistemas em mamografia ainda permanece uma incógnita e, apesar de serem utilizados rotineiramente nos programas de rastreio de cancro de mama nos Estados Unidos da América (EUA) como segunda opinião (3), segundo a análise de Kohli e Saurabh (2017) (2) 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 (7).

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 relacionados com Inteligência Artificial (AI), que pretendem transferir para a área da imagem médica, os sistemas de análise de dados visuais aplicados noutra tipo de tarefas, como é o caso das Técnicas de *Machine Learning* (MLT) ou das *Artificial Neural Networks* (ANN), as quais apresentam resultados promissores (2,3,8).

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 tipos de algoritmos diferentes na classificação de lesões em imagens de mamografia. Sendo assim, numa primeira fase, foram aplicados algoritmos de visão computacional, tendo sido utilizada a medida da dimensão fractal (DF) para fazer uma diferenciação entre lesões benignas e malignas ou para tentar fazer uma distinção entre o tipo de lesão. Posteriormente, numa segunda fase, foram aplicados algoritmos de *Deep Learning*, particularmente Redes Neurais de Convolução ou *Convolutional Neural Networks* (CNN), na classificação automática de lesões, fazendo novamente a distinção entre lesões benignas e malignas. Nesta segunda fase, para além de ser testada a aplicação das CNN isoladamente, com a classificação a ser realizada pela *Softmax layer*, também foi testado o desempenho na combinação das CNN com uma *Support Vector Machine* (SVM), e de ambas com *handcrafted features* (HCFs) e a informação da idade e da densidade mamária das pacientes.

Desta forma, seguidamente é apresentado um enquadramento teórico que ajuda a contextualizar as técnicas e o material utilizado no processo de investigação, sendo seguidamente explicada a metodologia utilizada em ambas as fases do estudo. Posteriormente, na secção 4, são apresentados e discutidos os resultados e, nas secções seguintes encontram-se as conclusões do estudo e as sugestões para futuras aplicações.

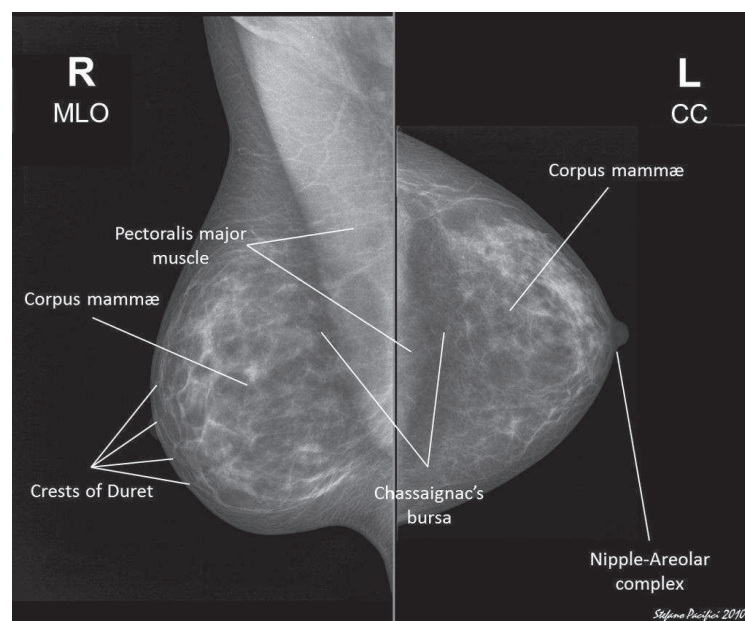
## 2. Enquadramento Teórico

### 2.1. Mamografia

A mamografia é considerada, desde há muito tempo, o exame de diagnóstico por imagem mais eficaz na deteção do cancro de mama (9). Pode ser realizada por rotina, em mulheres assintomáticas, ou para diagnóstico, sendo uma ferramenta fundamental na deteção de lesões em fases precoces, permitindo um prognóstico favorável e um aumento na taxa de sucesso dos tratamentos (10).

De acordo com a norma 051/2011 da DGS, a mamografia de rastreio está indicada, a cada dois anos, em mulheres assintomáticas, com idades compreendidas entre os 50 e os 69 anos, podendo a ecografia mamária ser utilizada como exame complementar em mulheres com densidade mamária elevada ou prótese mamária. Acima dos 69 anos, a periodicidade recomendada é de 2 a 3 anos, entre os 35 e os 50 anos apenas é realizada no caso de existir sintomatologia associada e, abaixo dos 35 anos a ecografia mamária é o exame de primeira linha (1).

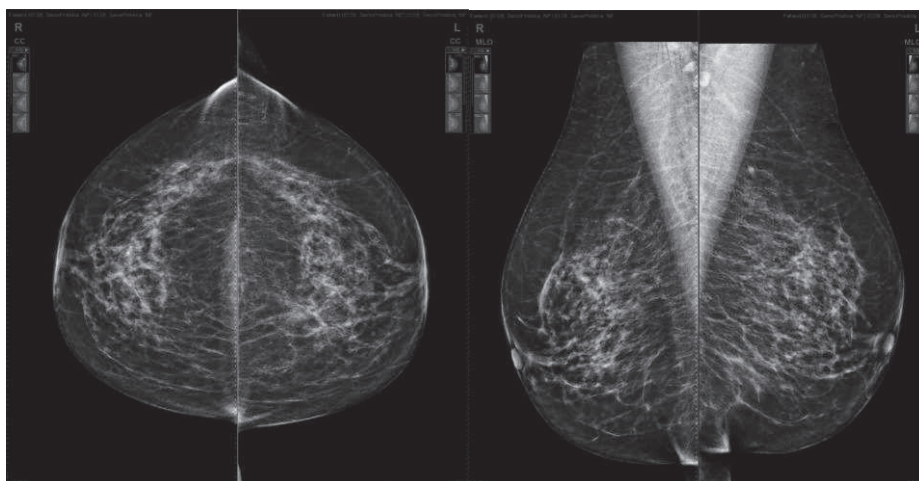
No decorrer de um exame mamográfico, através da utilização de um feixe de radiação X, obtém-se uma projeção bidimensional das estruturas tridimensionais que compõem a mama. As estruturas com menor densidade, como o tecido adiposo, irão surgir hipodensas, pois absorvem menos radiação, enquanto que as estruturas com maior densidade, como o tecido fibroglandular ou o músculo peitoral, irão surgir hiperdensas (10), resultando numa imagem semelhante às que se encontram representadas na Figura 1.



**Figura 1** - Imagens de mamografia com a respetiva descrição das estruturas visualizadas.

Fonte: [https://prod-images.static.radiopaedia.org/images/5040918/7619287b4bf54b031335c25cca3220\\_jumbo.jpeg](https://prod-images.static.radiopaedia.org/images/5040918/7619287b4bf54b031335c25cca3220_jumbo.jpeg)

As incidências básicas realizadas num exame de mamografia são a incidência crânio-caudal (CC) e a incidência oblíqua médio-lateral (OML), sendo realizadas bilateralmente para comparação (crânio-caudal direita e esquerda – CCD e CCE; oblíqua médio-lateral direita e esquerda – OMLD e OMLE) (1,11) – Figura 2.



**Figura 2** - Imagens de mamografia (incidências básicas de rotina - CCD, CCL, OMLD e OMLE).

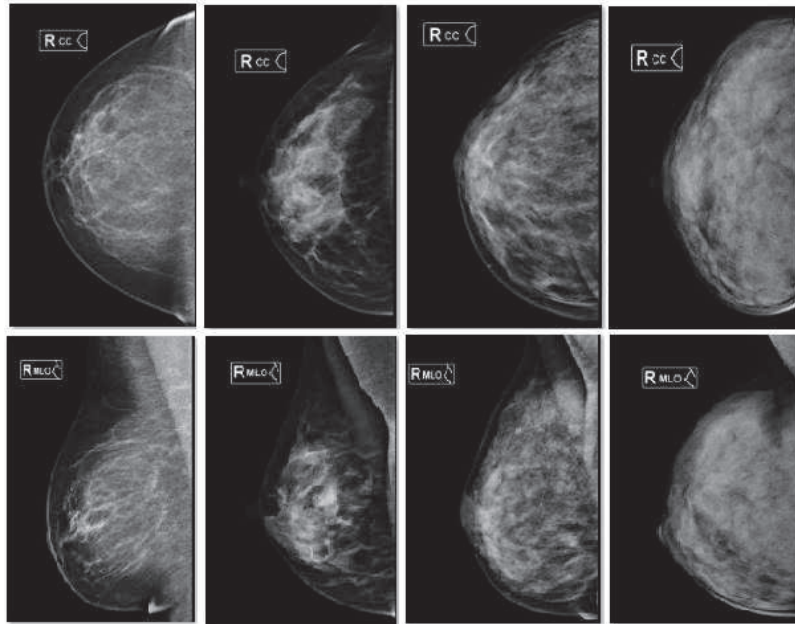
Fonte: [http://newsroom.gehealthcare.com/wp-content/uploads/2017/10/CLINICAL-IMAGE-Senographe-Pristina-0128\\_econtrast4.png](http://newsroom.gehealthcare.com/wp-content/uploads/2017/10/CLINICAL-IMAGE-Senographe-Pristina-0128_econtrast4.png)

Os equipamentos de mamografia possuem um dispositivo de compressão, controlado pelo técnico executante, utilizado para comprimir a mama através da aplicação de uma força entre 11 e 20 kgf (quilograma-força), ou seja, entre 108 e 196 N (Newton). A compressão mamária tem uma dupla função: em primeiro lugar, a diminuição da espessura da mama e, em segundo, a colocação das estruturas mamárias o mais próximo possível do detetor. Estes dois fatores melhoram a qualidade e a resolução espacial por reduzirem a dispersão e a ampliação das estruturas mamárias (10).

### 2.1.1. Sistema de Classificação BIRADS

O sistema de classificação BIRADS (*Breast Imaging Reporting and Data System*) foi implementado pelo *American College of Radiology* (ACR), nos EUA, desde 1995. Atualmente é utilizado em vários países com programas de rastreio de forma a uniformizar a terminologia dos relatórios, facilitar a avaliação das imagens e a recomendação das ações a serem tomadas após um exame de imagem da mama (12).

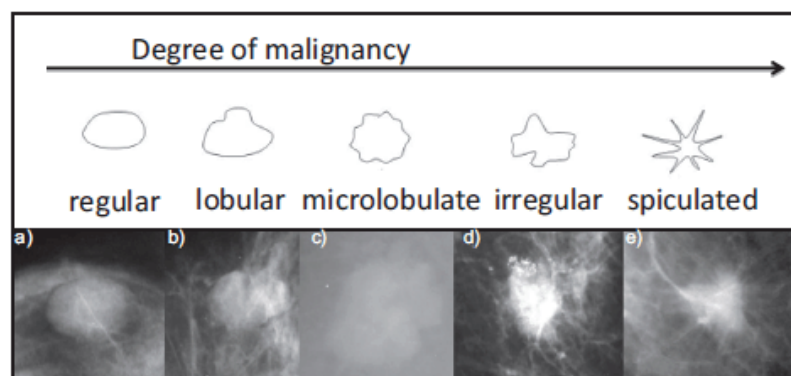
O relatório deve descrever, inicialmente, a densidade radiológica das mamas, tendo em conta que uma densidade elevada do parênquima mamário reduz a sensibilidade do método, mas não altera a categoria. A densidade mamária é classificada em 4 categorias: Tipo I – Adiposa (menos de 10% de tecido denso); Tipo II – Fibroglandular (10-49% de tecido denso); Tipo III – Heterogeneamente Densa (49-90% de tecido denso) e Tipo IV – Densa e Homogénea (mais de 90% de tecido denso) (12-14). A Figura 3 apresenta exemplos de imagens mamográficas com as diferentes categorias de densidade radiológica.



**Figura 3** - Exemplos de imagens mamográficas com diferentes densidades radiológicas, incidências CCD e OMLD, da esquerda para a direita, tipo I, II, III e IV. *Fonte: Própria.*

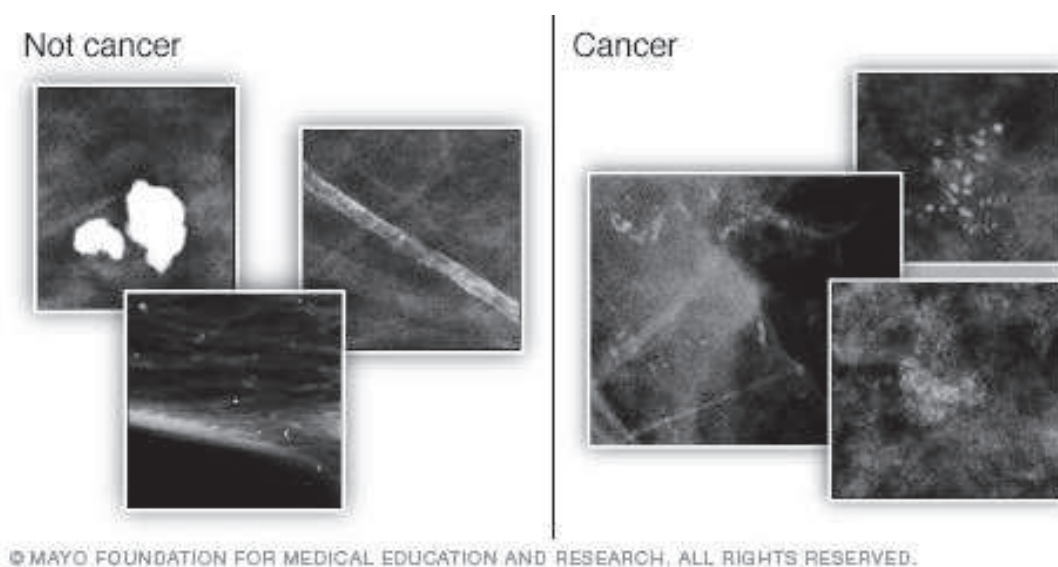
Seguidamente o relatório deve incluir a descrição dos achados imagiológicos presentes na mama, tais como nódulos, calcificações ou distorções da arquitetura do tecido mamário, assim como a sua localização e os sinais associados, como por exemplo a presença de nódulos linfáticos na axila, lesões cutâneas ou retração mamilar (12).

Os nódulos são considerados lesões ocupando espaço visualizadas em duas incidências e a sua descrição deve incluir informação sobre as suas dimensões, forma (redonda, oval, lobulada ou irregular), contornos (circunscrita, microlobulada, obscurecida, indistinta ou espiculada) e densidade (hiperdensa, isodensa ou hipodensa) (12). Uma lesão irregular ou com contornos espiculados sugere malignidade enquanto que uma lesão redonda ou oval e com contornos circunscritos tem maior probabilidade de ser benigna (15). A Figura 4 permite perceber o grau de malignidade das lesões nodulares em relação à sua forma e contorno.



**Figura 4** - Grau de malignidade dos nódulos mamários relativamente à sua forma e contorno.  
 Fonte: [https://www.researchgate.net/profile/Marwa\\_Hmida/publication/321494909/figure/fig3/AS:568031001747456@1512440552208/Classification-of-breast-masses-according-to-their-contours.png](https://www.researchgate.net/profile/Marwa_Hmida/publication/321494909/figure/fig3/AS:568031001747456@1512440552208/Classification-of-breast-masses-according-to-their-contours.png)

As calcificações são normalmente pequenos depósitos de cálcio que se acumulam nos tecidos e surgem na mamografia como uma região. Podem classificar-se em três tipos: tipicamente benignas (calcificações da pele, vasculares, grosseiras, redondas) que não necessitam de seguimento; intermédias (amorfas ou indistintas na aparência) que necessitam frequentemente de seguimento especial; e com alta probabilidade de malignidade (pleomórficas com 0,2 a 0,3 mm de dimensão, irregulares e mais densas que as amorfas, podem ainda ser finas, lineares e ramificadas quando se situam nos pequenos ductos terminais). A distribuição das calcificações de modo difuso ou em grupo pode alterar o seu grau de suspeição e uma distribuição agrupada ou linear é preocupante, mesmo se as características forem do tipo intermédio. Quando a distribuição segue uma distribuição segmentar, regional ou difusa, a suspeição de malignidade diminui (12,15). A Figura 5 apresenta exemplos de calcificações que podem ou não ser consideradas suspeitas de malignidade.



**Figura 5** - Exemplos de calcificações mamárias que podem ou não ser suspeitas de malignidade.

Fonte: [https://www.mayoclinic.org/-/media/kcms/gbs/patient-consumer/images/2013/11/15/17/39/ds00328\\_-ds00983\\_-my00303\\_im01869\\_ans7\\_breastcalcificationsth.jpg.jpg](https://www.mayoclinic.org/-/media/kcms/gbs/patient-consumer/images/2013/11/15/17/39/ds00328_-ds00983_-my00303_im01869_ans7_breastcalcificationsth.jpg.jpg)

As distorções da arquitetura correspondem à retração do tecido glandular com a produção de espículas radiadas e finas. Pode ou não ser associada a uma massa e é, frequentemente, um achado de malignidade. A assimetria glandular, ou presença de tecido glandular, numa parte da mama sem correspondência na mama contralateral é frequentemente um achado normal (15).

No final do relatório deve-se incluir a respetiva categoria BIRADS para cada mama, assim como a recomendação de periodicidade dos exames (12,13). Na Tabela 1 encontra-se a descrição das diferentes categorias da escala BIRADS.

**Tabela 1** - Descrição das diferentes categorias da escala BIRADS (12).

Escala BIRADS	Interpretação	Risco de Malignidade	Recomendação
BIRADS 0	Inconclusivo Avaliação incompleta		Rever estudos anteriores e/ou complementar com métodos adicionais
BIRADS 1	Normal Sem alterações	0,05%	Continuação do controlo de rotina (2 em 2 anos)
BIRADS 2	Alterações Benignas	0,05%	Continuação do controlo de rotina (6 a 12 meses)
BIRADS 3	Alterações provavelmente benignas	Até 2%	Repetição do exame após 6 meses ou possível biópsia
BIRADS 4	Alterações suspeitas	>20%	Biópsia, preferencialmente aspirativa
BIRADS 5	Provavelmente maligno Altamente suspeito de malignidade	>75%	Biópsia e tratamento, se necessário
BIRADS 6	Já diagnosticada como maligna, não submetida a cirurgia/tratamento	100%	

### 2.1.2. Bases de Dados em Mamografia

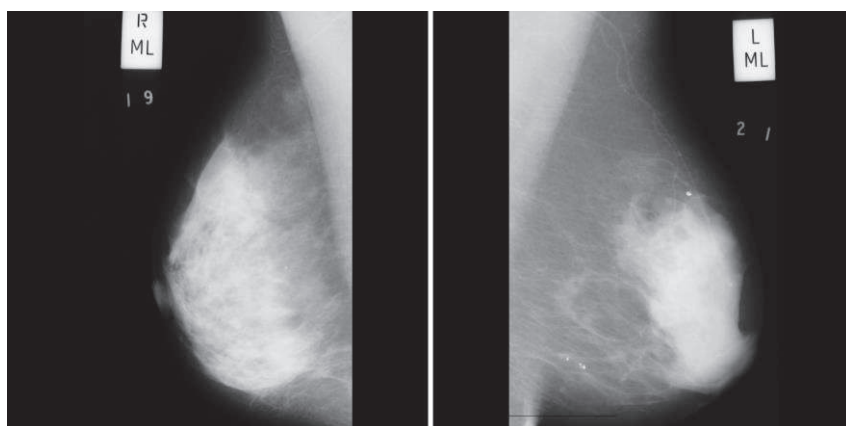
A utilização de imagens de mamografia provenientes de bases de dados privadas dificulta a correta avaliação da performance dos algoritmos CAD e sua comparação com outros métodos, uma vez que a sua avaliação é dependente da base de dados utilizada. Desta forma, existem várias bases de dados de mamografia disponíveis, com acesso público ou restrito, que podem ser utilizadas no desenvolvimento de estratégias que ajudem a melhorar a deteção e o diagnóstico do cancro da mama (16). A Tabela 2 resume as bases de dados de mamografia disponíveis.

**Tabela 2** - Bases de dados de mamografia mais utilizadas na literatura.

Características	MIAS	DDSM	IRMA	BancoWeb	BCDR	INBreast	CBIS_DDSM
Ano	1994	1999	2008	2010	2012	2012	2017
Origem	UK	USA	Alemanha	Brasil	Portugal	Portugal	USA
Acesso	Sim	Sim	(?)	Sim	Sim	Sim	Sim
Suportada	Não	Não	(?)	Sim	Sim	Sim	Sim
n.º casos	161	2620	--	320	1125/1042	117	1597
n.º Imagens	322	10.480	10.509	1400	3703/3612	410	3061
Aquisição	Screen Film	Screen Film	Screen Film	Screen Film	Screen Film/FFDM	FFDM	Screen Film
Incidências	MLO	CC/MLO	CC/MLO	CC/MLO	CC/MLO	CC/MLO	CC/MLO
Tipo de ficheiro	PGM	LJPEG	Vários	TIFF	TIFF	DICOM	DICOM
Resolução	8 bits/pixel	8/16 bits/pixel	Várias	12 bits/pixel	8/14 bits/pixel	14 bits/pixel	16 bits/pixels
Bi-Rads	Não	Sim	Sim	Sim	Sim	Sim	Sim
Densidade	Sim	Sim (ACR)	Sim (ACR)	Sim (ACR)	Sim (ACR)	Sim (ACR)	Sim (ACR)
Historial clínico	Não	Idade	Sim	Sim	Sim	Sim	Não

### ***Mammographic Image Analysis Society Digital Mammogram Database***

A *Mammographic Image Analysis Society Digital Mammogram Database* (MIAS)<sup>1</sup> datada de 1994, é composta por 161 casos, 322 imagens de mamografia obtidas através de um sistema ecrã-película e digitalizadas (incidência OML), num ficheiro PGM (*Portable Gray Map*) com uma resolução de 8 bits/pixel. Esta base de dados inclui imagens consideradas normais e imagens com lesões benignas e malignas, assim como a informação da densidade mamária (mas não de acordo com os standards do ACR). Não possui informação sobre a classificação BIRADS das imagens ou historial clínico. As anotações disponibilizadas consistem no centro e raio em torno da região de interesse, o que implica uma segmentação manual das lesões (16–18). A Figura 6 mostra duas imagens extraídas desta base de dados.



**Figura 6** - Exemplos de imagens de mamografia extraídas da base de dados MIAS - Incidência OMLD e OMLE, ambas consideradas normais.

### ***Digital Database for Screening Mammography***

A *Digital Database for Screening Mammography* (DDSM)<sup>2</sup> é a base de dados de mamografia mais utilizada pelos investigadores em todo o mundo e foi desenvolvida na Universidade de South Florida, nos EUA (1999). É constituída por 2620 casos clínicos, que resultaram em 10,480 imagens ecrã-película digitalizadas em formato LJPEG (*Lossless Joint Photographic Experts Group*), com uma resolução de 8 ou 16 bits/pixels. Inclui incidências CC e MLO de cada mama (ver exemplo Figura 7), com informação da classificação BIRADS e densidade mamária de acordo com os standards do ACR (16,19–23).

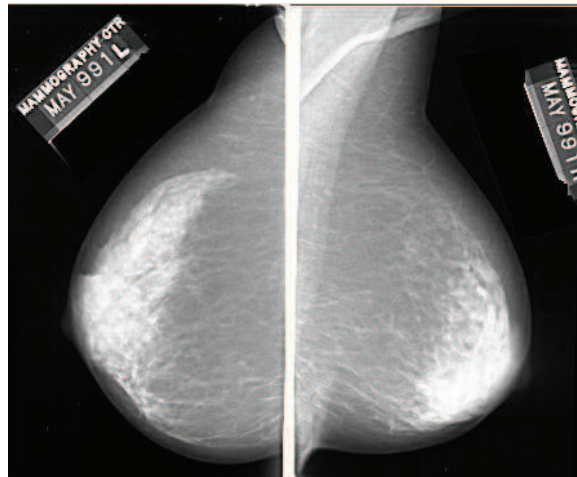
Recentemente, Lee *et al.* (2017) publicaram uma nova versão da DDSM, a *Curated Breast Imaging Subset of DDSM* (CBIS-DDSM)<sup>3</sup>, que inclui 3061 imagens de mamografia provenientes de 1597 casos, os quais foram selecionados e reavaliados por um especialista. Os autores fornecem novas segmentações das regiões de interesse,

<sup>1</sup> <http://peipa.essex.ac.uk/info/mias.html>

<sup>2</sup> <http://marathon.csee.usf.edu/Mammography/Database.html>

<sup>3</sup> <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>

utilizando o método Otsu, e “*bounding boxes*” das imagens convertidas para o formato DICOM (*Digital Imaging and Communications in Medicine*), assim como a sua classificação diagnóstica. Para além disso, sugerem um agrupamento de imagens em conjuntos de treino e teste com 1231 e 361 imagens, respetivamente (24,25).



**Figura 7** - Exemplos de imagens de mamografia extraídas da base de dados DDSM - Incidência CCD e OMLE, ambas consideradas normais.

### ***Image Retrieval in Medical Applications***

Tendo como base o projeto alemão *Image Retrieval in Medical Applications* (IRMA), Oliveira *et al.* (2008) desenvolveram uma base de dados através da fusão de quatro bases de dados diferentes: MIAS, DDSM, imagens do *Lawrence Livermore National Laboratory* e imagens do *Rheinisch-Westfälische Technische Hochschule*, o que resultou num conjunto de 10,509 imagens (com diferentes resoluções). Através do código IRMA, os autores disponibilizam o tipo de tecido mamário de acordo com os standards do ACR, a classificação BIRADS e descrição das lesões (26,27).

### ***BancoWeb LAPIMO Database***

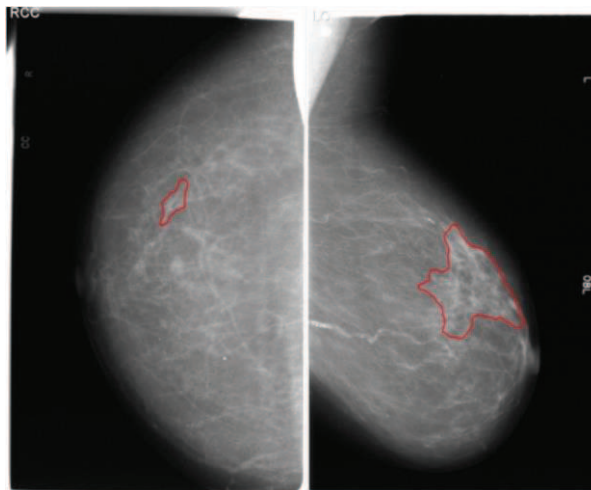
A *BancoWeb LAPIMO Database*<sup>4</sup> é uma base de dados de mamografia desenvolvida no Brasil e disponível após registo. Composta por 1430 imagens CC, OML e magnificações, de 320 casos normais, com lesões benignas e malignas. Os autores disponibilizam a informação relativa à classificação BIRADS das imagens, no entanto, os dados referentes aos contornos e textura das lesões não estão disponíveis para a totalidade das imagens (16).

<sup>4</sup><http://lapimo.sel.eesc.usp.br/bancoweb/>

## Breast Cancer Digital Repository

A *Breast Cancer Digital Repository* (BCDR)<sup>5</sup> é uma base de dados publica criada a partir de arquivos médicos disponibilizados pela Faculdade de Medicina do Centro Hospitalar de S. João – Universidade do Porto, a qual é subdividida em dois repositórios diferentes: a *Film Mammography-based Repository* (BCDR-FM) e a *Full Field Digital Mammography-based Repository* (BCDR-DM)(28).

A BCDR-FM inclui 1125 estudos com 3703 imagens CC e OML de 1010 casos clínicos. As imagens são maioritariamente de pacientes do sexo feminino (998), dos 20 aos 90 anos. A base de dados também inclui 1044 lesões identificadas e clinicamente descritas, 1517 segmentações realizadas manualmente e as classificações BIRADS propostas por radiologistas experientes. O conjunto de dados BCDR-F03 - *Film Mammography Dataset Number 3* é o mais utilizado, sendo constituído por 736 imagens de mamografia digitalizadas (426 lesões benignas e 310 lesões malignas) de 344 pacientes, com uma resolução de 720x1168 pixels e uma profundidade de imagem de 8 bits por pixel no formato TIFF (*Tagged Image File Format*). Os autores também disponibilizam os dados clínicos e descritores das imagens. A Figura 8 mostra dois exemplos de imagens da base de dados BCDR-FM.



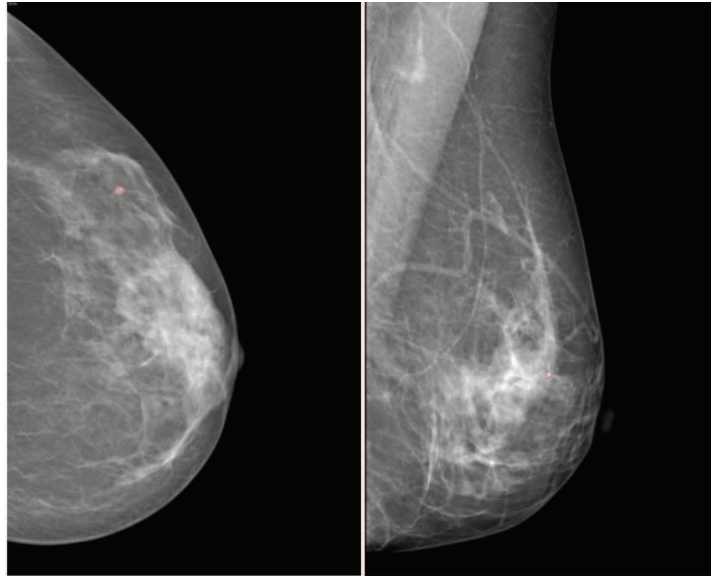
**Figura 8** - Exemplos de imagens de mamografia extraídas da base de dados BCDR-FM - Incidências CCD e MLOE, ambas com lesões malignas com segmentação da área da lesão.

A BDCR-DM inclui 612 imagens (incidências CC e OML) de 724 pacientes (1024 estudos), maioritariamente mulheres (723) com idades compreendidas entre os 27 e os 92 anos. Contém 452 lesões clinicamente descritas, 818 segmentações realizadas manualmente e a correspondente classificação BIRADS proposta por radiologistas especializados. As imagens encontram-se no formato TIFF, com uma matriz de 3328x4048 ou 2560x3328 pixels, dependendo do tamanho da mama e da placa de compressão utilizada na aquisição, a profundidade da imagem é de 14 bits por pixel.

---

<sup>5</sup><http://bcdr.inegi.up.pt>

Na Figura 9 são apresentados dois exemplos de imagens extraídas da base de dados BCDR-DM.



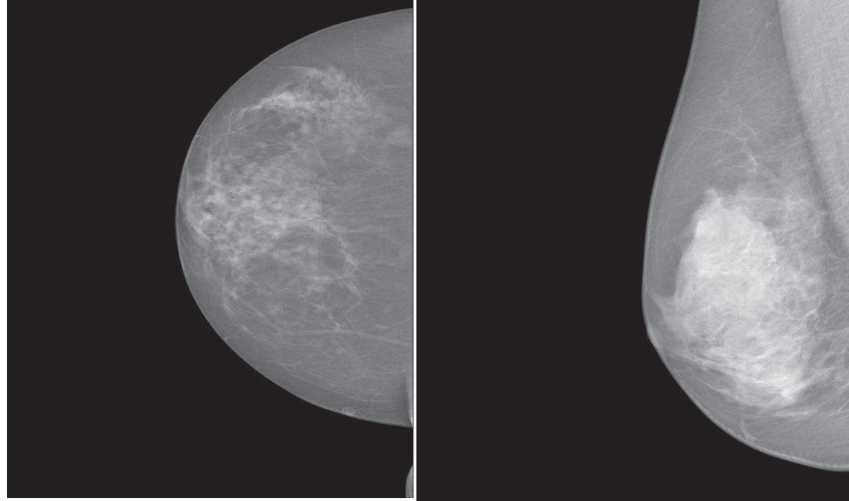
**Figura 9** - Exemplos de imagens de mamografia extraídas da base de dados BCDR-DM - Incidências CCE e OMLE, ambas benignas.

### ***INBreast Database***

A base de dados *INBreast* é uma base de dados de mamografia criada pela Unidade de Mama do Centro Hospitalar de S. João, Porto, a qual inclui um total de 117 casos, com um total de 410 imagens *Full-Field Digital Mammography* (FFDM) obtidas a partir de casos de rastreio, diagnóstico e *follow-up*. A aquisição das imagens foi realizada entre abril de 2008 e julho de 2010, num equipamento *MammoNovation Siemens FFDM*, com um detetor sólido de selénio amorfo, com um tamanho de pixel de 70  $\mu\text{m}$  e uma resolução de contraste de 14-bit e uma matriz com dimensão 3328 $\times$ 4084 ou 2560 $\times$ 3328 pixels, dependendo da placa de compressão utilizada na aquisição. As imagens estão disponíveis no formato DICOM. As anotações foram realizadas por dois radiologistas especializados na área do estudo da mama e guardadas em formato XML (*Extensible Markup Language*), incluindo assimetrias, calcificações, agrupamentos de microcalcificações, massas, distorções, regiões espiculadas e músculo peitoral (apenas nas incidências OML).

Os autores fornecem informação detalhada sobre os contornos do músculo peitoral e de lesões como assimetrias, calcificações e massas, e ainda elipses onde se encontram incluídos os agrupamentos de microcalcificações e todas as áreas espiculadas localizadas em regiões mais densas da mama. Também são disponibilizadas informações acerca da idade dos pacientes, data de aquisição das imagens, história familiar, classificações ACR da densidade mamária e escala BIRADS e, finalmente, a classificação Benigno/Maligno das lesões.

Este subconjunto contém um total de 116 massas, localizadas maioritariamente no quadrante superior externo (QSE), e um total de 6880 calcificações, em 299 imagens (16,22,29–31). A Figura 10 apresenta dois exemplos de imagens da base de dados *INBreast*.



**Figura 10** - Exemplos de imagens de mamografia extraídas da base de dados *INBreast* - Incidências CCD e OMLD, ambas benignas.

### **Digital Mammography DREAM Challenge**

A *Digital Mammography (DM) DREAM Challenge*<sup>6</sup> foi criada para permitir a submissão de novos algoritmos de *machine learning*, *deep learning*, entre outros, aplicando-os a imagens de mamografia, com o objetivo de melhorar a classificação de imagens digitais e melhorar o diagnóstico do cancro de mama. A base de dados é composta por 640,000 imagens digitais de mamografia não identificadas de cerca de 86,000 pacientes do *Breast Cancer Surveillance Consortium* e da *Icahn School of Medicine* no Mount Sinai. Os participantes fazem o *upload* dos algoritmos numa plataforma *cloud* segura e o *download* das imagens de mamografia não está disponível. Deste modo, a composição dos conjuntos de treino e teste é desconhecida, e durante estes procedimentos os participantes não conseguem fazer alterações no seu programa, o que torna possível a comparação da performance dos diferentes métodos nas mesmas condições, tanto em relação às imagens utilizadas como em relação à avaliação da performance da classificação (32).

---

<sup>6</sup> <https://www.synapse.org/!Synapse:syn4224222>

## 2.2. Computer Aided Detection/Diagnosis

Os sistemas CAD foram desenvolvidos de forma a reduzir as limitações técnicas e humanas que surgem na interpretação de imagens médicas, alertando os radiologistas para regiões suspeitas da mamografia e dignas de uma investigação mais profunda (4).

Os primeiros estudos relacionados com a análise e detecção automática de lesões remontam ao ano de 1967, quando Winsberg *et al.* criaram um dispositivo capaz de detetar variações na densidade das imagens de mamografia. No entanto, apesar dos resultados alcançados com as novas aplicações computacionais, estas descobertas não tiveram grande sucesso uma vez que as imagens ainda eram, na sua maioria, adquiridas através de um sistema ecrã-película, o poder dos algoritmos computacionais era insuficiente e ainda não existiam técnicas de processamento de imagem (3).

O ano de 1998 marca a transição da fase de investigação para a prática industrial, com a aprovação do primeiro sistema CAD comercial pela *Food and Drug Administration* (FDA), o *ImageChecker™*, criado para auxiliar os médicos radiologistas na avaliação das imagens mamográficas, em vez de os substituir (2,3). Atualmente, os sistemas CAD são utilizados rotineiramente em hospitais, clínicas e centros de diagnóstico nos EUA (3).

Existem dois tipos de aplicações dos sistemas CAD, nomeadamente, os *Computer-Aided Detection* (CADe) que são aplicados na deteção, localização ou segmentação de regiões suspeitas da imagem na sua totalidade, baseando-se na localização de padrões anormais através da análise da imagem pelo computador, e os *Computer-Aided Diagnosis* (CADx) aplicados como auxílio ao diagnóstico, realizado através da quantificação de características da imagem e da sua classificação em padrões normais ou anormais, associando-os à benignidade ou malignidade, podendo ser usados para obter informações adicionais ou para suportar a classificação final, a qual é da total responsabilidade do médico radiologista (2,33). A fusão destes dois tipos de aplicações deu origem aos sistemas CAD-Dx, que possibilitam o auxílio na deteção e no diagnóstico das lesões da mama (Figura 11).

Uma imagem de mamografia *standard* de 12 *bits* contém  $2^{12}$  (4096) tons de cinzento. No entanto, a perceção do olho humano está limitada a cerca de  $2^4$  (16) tons. Os algoritmos computacionais têm a capacidade de detetar alterações mínimas na escala de cinzentos e extrair o máximo de informação de uma imagem e, desta forma, detetam lesões impercetíveis ao olho humano (2).

Os sistemas CAD recorrem a técnicas de **Visão Computacional e Inteligência Artificial**. A Visão Computacional baseia-se na utilização de um sistema computacional para extração da informação de uma imagem digital, podendo ser dividida em duas etapas, nomeadamente, o processamento de imagem para realce e segmentação de lesões, através da subdivisão da imagem em regiões com informação semelhante, com base nas descontinuidades dos níveis de cinzento (contornos, linhas, fronteiras ou similaridade), e a quantificação da informação contida na imagem, tal como o tamanho,

o contraste e a forma dos objetos que constituem a imagem. Após a extração da informação, as diferentes características quantificadas são utilizadas no processo de classificação das estruturas presentes na imagem, através de utilização de técnicas de Inteligência Artificial, como por exemplo redes neuronais (2,3,8). A AI envolve a utilização de computadores para o processamento de dados, com o objetivo, por exemplo, de distinguir padrões normais e anormais a partir das características extraídas da imagem (3).

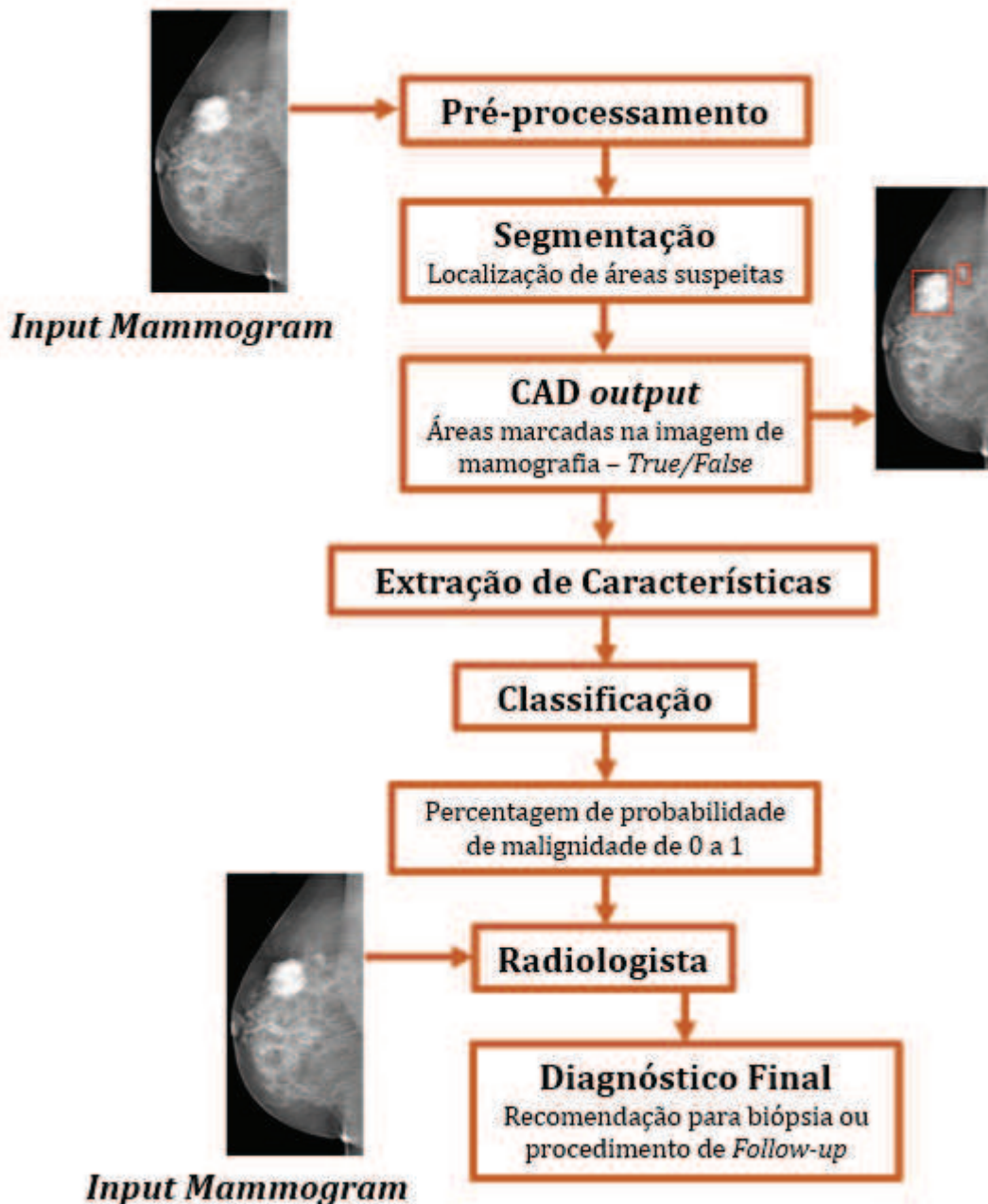


Figura 11 - Ilustração Sistemática de um sistema CAD-Dx.

### 2.2.1. Dimensão Fractal

A DF é considerada uma técnica de visão computacional. Esta medida pode ser utilizada para quantificar a irregularidade da forma dos contornos das lesões que surgem nas imagens de mamografia, podendo, assim, ser uma boa ferramenta de auxílio na descrição, e posterior classificação dessas lesões (34).

Tal como já foi referido anteriormente, uma das características mais importantes na categorização de uma mamografia é a forma do contorno da lesão; um contorno regular está associado a casos benignos, enquanto contornos irregulares estão associados a casos de malignidade (34,35), sem descurar claro as outras características a ter em conta na classificação da imagem. Desta forma, a DF pode ser utilizada para caracterização de lesões em mamografia, tendo sido sugerida pela semelhança entre o tecido mamário e a imagem fractal gerada sinteticamente (Figura 12)(36).

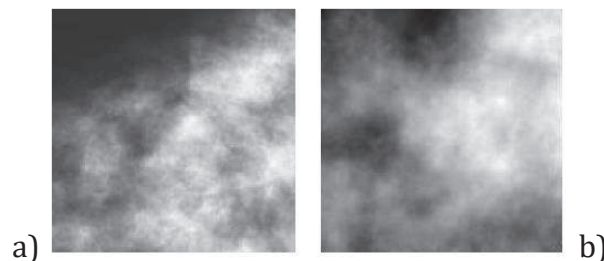


Figura 12 - Similaridade entre a) tecido mamário e b) fractais gerados sinteticamente (36).

### Análise fractal

Os tumores malignos possuem um certo grau de aleatoriedade associada ao seu crescimento, sendo, tipicamente, irregulares e de forma complexa. Desta forma, a análise fractal pode fornecer uma medida mais eficaz do que a geometria Euclidiana convencional a respeito da complexidade dos seus padrões (37,38). Os padrões complexos e irregulares dos tumores malignos podem ser caracterizados pela DF sendo expectável que apresentem altos valores de DF comparativamente com os contornos mais suaves dos tumores benignos (38).

A dimensão da similaridade própria é baseada nas propriedades de homogeneidade própria dos objetos fractais, sendo o objeto representado por cópias de si próprio em outras escalas. Na Figura 13 é apresentado o *Koch Snowflake* um fractal conhecido, obtido pela repetição de um determinado procedimento infinitamente.

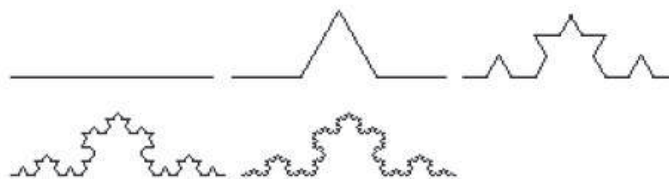


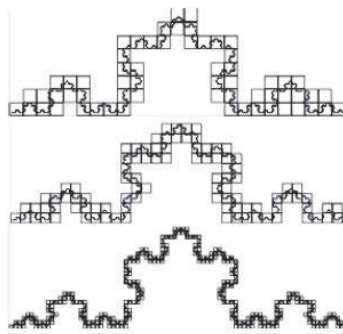
Figura 13 - *Koch Snowflake* fractal com comprimento infinito e área zero (39).

O matemático germânico Felix Hausdorff sugeriu a equação 1 para avaliar a dimensão fractal (35).

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

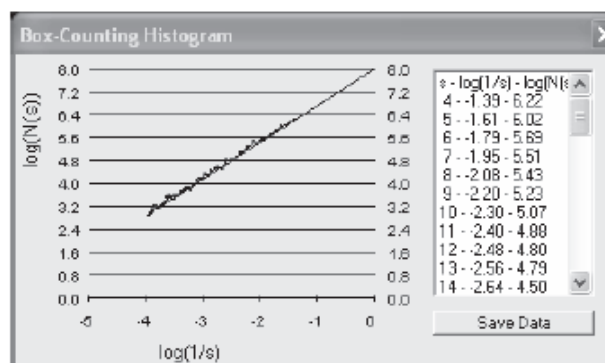
em que  $N(s)$  é o número de peças com similaridade própria no padrão ou objeto e  $1/s$  é o fator de magnificação. Posteriormente a DF pode ser calculada através do valor da declive da reta (*D-slope*) da alteração do tamanho dos objetos ( $\log N(s)$ ) versus a alteração na escala de medição ( $\log(1/s)$ ) (34,35).

Existem vários métodos computacionais para estimar a DF, sendo um dos mais conhecidos o *box counting method*, que consiste na partição do espaço padrão ou da imagem em quadrados de igual tamanho, e posterior contagem do número de quadrados que contêm pelo menos um *pixel* pertencente ao contorno. O processo é repetido pela partição do espaço da imagem em quadrados cada vez mais pequenos (Figura 14) (35,38).



**Figura 14** - O Koch Snowflake coberto por quadrados com cada vez menor dimensão. Quanto menor a área de cobertura de cada quadrado, mais detalhes serão considerados (39).

Como foi dito anteriormente, a DF é obtida através do declive da reta que inclui os valores do logaritmo do número de quadrados contabilizados ( $N(s)$ ) versus o logaritmo do índice de magnificação ( $s$ ) para cada fase de partição (Figura 15) (35).



**Figura 15** - Exemplo de uma Curva log-log adquirida através do método *Box Counting* (39). A dimensão fractal calculada é, neste exemplo, 1,2612.

Segundo Crisan *et al.* (2007) (39), a dimensão fractal mede a complexidade de um objeto e o seu valor aumenta com o aumento da irregularidade da forma do mesmo, o que é bastante útil na caracterização de lesões em mamografia.

### 2.2.2. Redes Neurais de Convolução (CNN)

Existem várias técnicas de *Machine Learning* (MLT) aplicadas na deteção e diagnóstico do cancro de mama, tanto em imagens de mamografia como em imagem de ultrassonografia, Ressonância Magnética, imagens histológicas e de infravermelhos (40). A Figura 16 mostra um gráfico com as diferentes modalidades que utilizam sistemas CAD, sendo a mamografia a que possui mais aplicações nesta área.

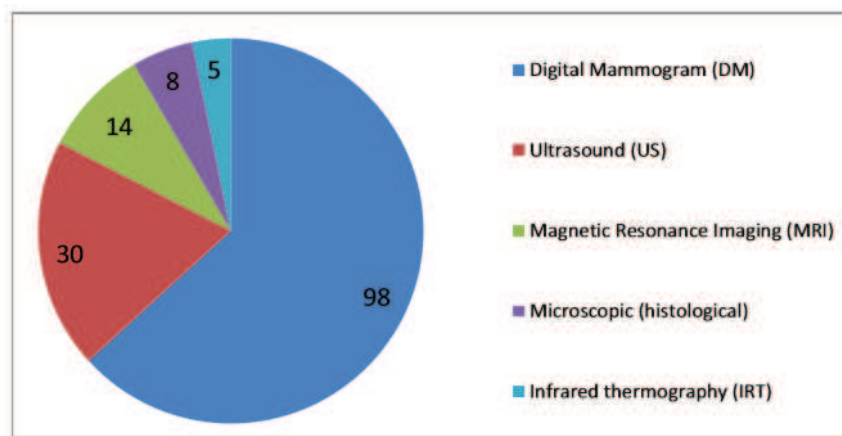
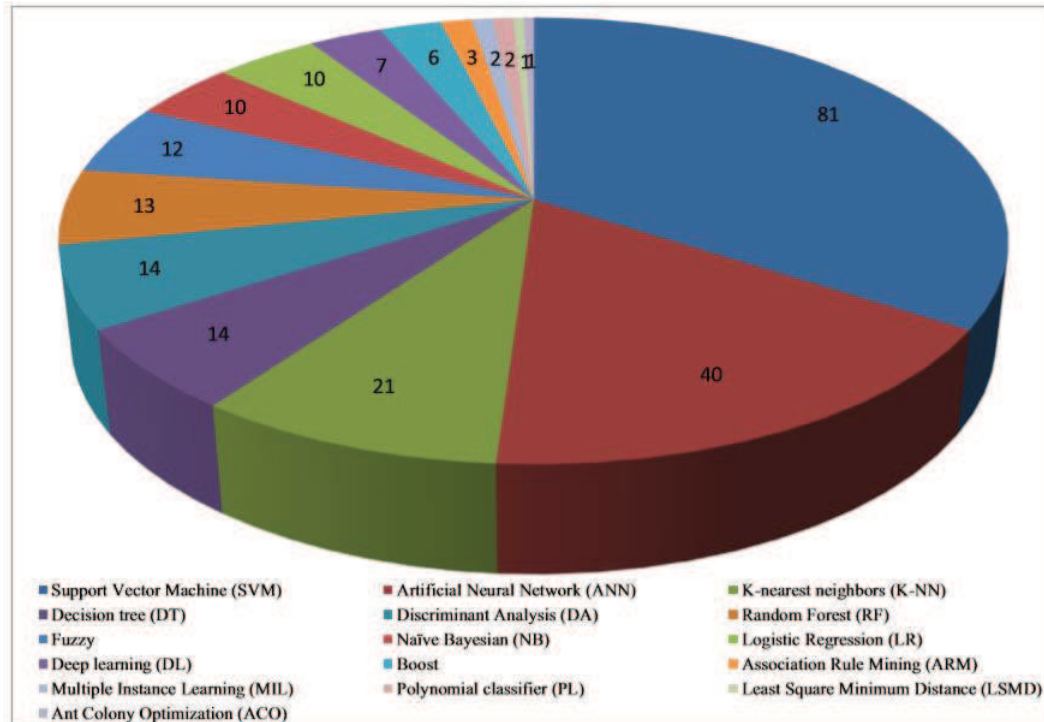


Figura 16 - Gráfico com as diferentes modalidades que utilizam sistemas CAD (40).

A aplicação dos diferentes algoritmos de classificação envolve frequentemente a divisão das bases de dados em dois ou três subconjuntos, denominados “treino”, “validação” e “teste”. A percentagem relativa entre estes subconjuntos varia de 50%, 10% e 40% (28) para 80%, 10% e 10% (41) ou 60%, 20% e 20% (20). Assim, o algoritmo é desenvolvido através da utilização do conjunto de treino, sendo posteriormente aplicado ao conjunto de teste, o que garante que os dados a ser alvo de classificação não são conhecidos. O treino na rede pode ser supervisionado, quando os dados de treino e as respetivas *labels* (maligno/benigno) são disponibilizados ao algoritmo no decorrer do processo, ou não-supervisionado, quando os dados de treino não possuem *labels* conhecidas, podendo ser fornecida alguma informação mas não necessariamente para todos os exemplos utilizados no treino (8).

As MLT são utilizadas no processo de classificação do tecido mamário com base nas características extraídas da imagem e podem ser usadas isoladamente ou em conjunto de forma a melhorar a performance dessa mesma classificação (40). A Figura 17 mostra o gráfico com os diferentes tipos de MLT utilizadas nos sistemas CAD aplicados às imagens no estudo do cancro de mama. Como se pode verificar as mais utilizadas são as SVM, as ANN e as *K-nearest neighbors* (K-NN).



**Figura 17** - Gráfico das diferentes MLTs utilizadas nos sistemas CAD para estudo do cancro da mama (40).

As CNN são um exemplo de algoritmos de DL utilizados não só em *Computer Vision* para extração e seleção de características, como também na classificação de lesões num procedimento de aprendizagem supervisionado (33,42). Foram introduzidas por Fukushima e mais tarde aperfeiçoadas por LeCun *et al.*, sendo consideradas como o algoritmo de DL mais bem sucedido no reconhecimento de padrões em imagens (28). Comparativamente com as MLT consideradas clássicas, como é o caso das SVM por exemplo, o processo de treino das redes neuronais é mais demorado uma vez que estes algoritmos possuem milhões de parâmetros e, como tal, é necessário possuir hardware com elevada capacidade computacional (8).

As redes neuronais, como o próprio nome indica, funcionam de forma semelhante às células neuronais que constituem o córtex visual primário do cérebro humano e que aprendem as características das imagens visualizadas. Estas redes são constituídas por camadas ou “*layers*” consecutivas: a “*input layer*”, na qual é fornecida, por exemplo, uma imagem de mamografia alvo de classificação, as “*hidden layers*”, cada uma com grupos de neurónios que apresentam funções semelhantes, e a “*output layer*”, na qual consta a predição do algoritmo relativamente à imagem fornecida (por exemplo a classificação benigno/maligno). A Figura 18 demonstra a estrutura básica de uma rede neuronal (8).

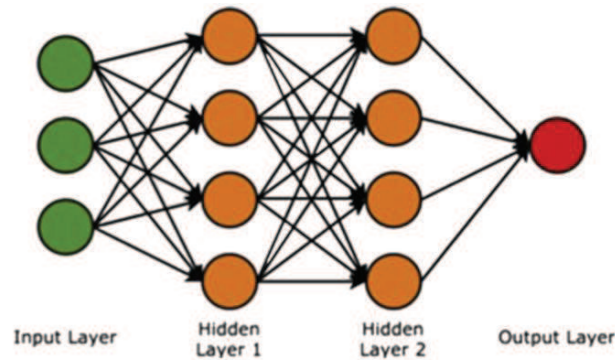


Figura 18 - Representação da estrutura de uma rede neuronal (8).

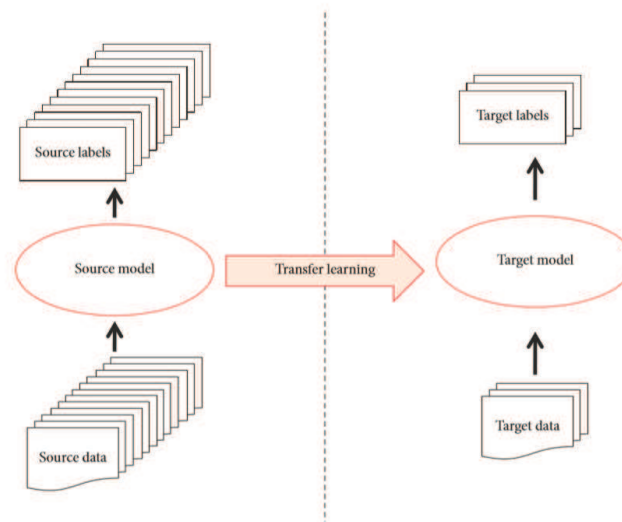
Genericamente, as “*hidden layers*” que constituem a arquitetura de uma CNN compreendem as “*convolutional layers*”, “*full-connection*” ou “*fully connected layers*”; e as “*pooling layers*” (8).

As camadas de convolução, ou *kernels*, atuam como um filtro na matriz de pixels da imagem fornecida para extrair as características correlacionadas espacialmente gerando um “*feature map*” ou mapa de características. Posteriormente, uma “*non-linear activation function*” (por exemplo uma “*rectified linear unit*” – ReLU) converte para zero todos os valores negativos desse “*feature map*”, o qual é seguidamente transferido para a “*pooling layer*”, a qual permite a redução da sua resolução. Por último, os dados resultantes são transferidos para a “*fully connected layer*” para classificação. Podem existir milhares de “*hidden layers*” o que torna mais complexa a extração das características (8). Por último, existe ainda a “*Softmax layer*”, a qual gera um vetor que representa a distribuição de probabilidades de aquela imagem fornecida como *input* pertencer a uma das possíveis classes (p. ex. benigno vs maligno), basicamente, é a camada que fornece a informação relativa à decisão final da classificação.

As CNN podem ser treinadas a partir do zero, o que requer não só uma grande quantidade de dados, como também uma grande capacidade tecnológica e de memória, o que apenas seria exequível com as mais recentes *Graphical Processing Units* (GPU), que permitiriam a redução do tempo necessário para o treino da rede (43). Infelizmente, para além dos elevados custos inerentes ao *hardware*, no campo da medicina e mais especificamente no campo da mamografia, a quantidade de imagens é também uma grande limitação, uma vez que a maioria dos dados não são disponibilizados gratuitamente, muitos deles não se encontram corretamente classificados, ou as bases de dados públicas possuem poucos dados comparativamente com o que seria necessário para treinar uma rede desde o início (8,43).

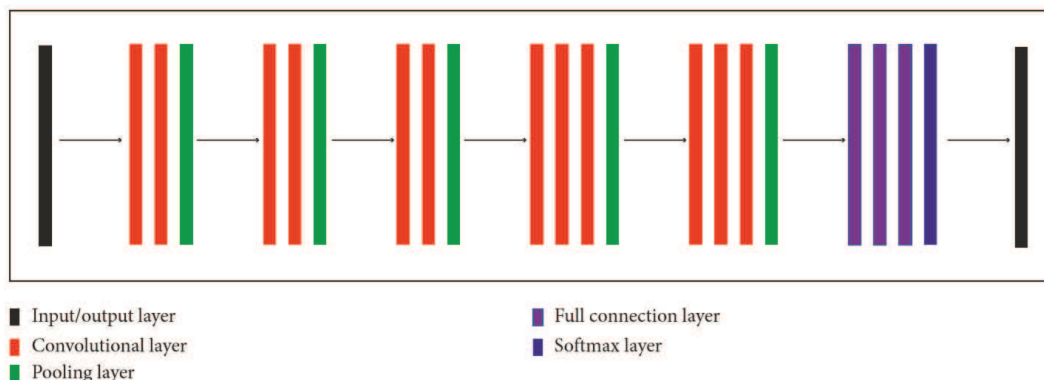
De forma a contornar estas limitações, muitos investigadores optaram por utilizar um método designado por “*transfer learning*” (TL). Neste método são utilizados modelos de redes pré-treinadas num determinado problema, as quais são ajustadas e aplicadas numa área diferente, como por exemplo na classificação de imagens de mamografia (33,43).

A Figura 19 apresenta um diagrama que ajuda a perceber melhor o conceito de “*transfer learning*”.



**Figura 19** - Conceito de “*Transfer Learning*”. No “*source model*” o modelo é treinado com uma quantidade suficiente de dados e labels de elevada qualidade e o método de *transfer learning* permite que, após ajustes, o “*target model*” seja aplicado na área de estudo pretendida com menor quantidade de dados, o que é benéfico principalmente na área da imagem médica (33).

Existem vários modelos de redes pré-treinadas (p. ex. *VGG*, *LeNet*, *AlexNet*, *GoogLeNet*, *ResNet* e *YOLO*) que utilizam grandes quantidades de dados, como por exemplo a base de dados *ImageNet*. Esta base de dados disponibiliza mais de 1.2 milhões de imagens naturais categorizadas, imagens essas que não pertencem ao campo da medicina (33,43). Na Figura 20 encontra-se representado um esquema relativo à arquitetura do modelo pré-treinado VGG16. Desta forma, os modelos pré-treinados podem ser aplicados na extração de características das imagens e os parâmetros da rede podem ser ajustados (“*fine-tuned*”) aos dados que se pretende classificar, o que permite reduzir drasticamente o tempo de treino da rede, passando o treino a demorar algumas horas em vez de vários dias ou meses (33).



**Figura 20** - Representação esquemática da arquitetura da rede VGG16: 13 “*convolutional layers*”, 3 “*full-connection layers*”, 5 “*pooling layers*” e 1 “*softmax layer*” (33).

Como já foi referido, as CNN possuem uma dupla função, podendo ser utilizadas tanto na extração e seleção de características, como também na classificação de lesões. Deste modo, existe também a possibilidade de utilizar este tipo de algoritmos na extração das características das imagens em vários níveis de representação (*low-, mid- e high-level features*) e, posteriormente, utilizar uma ou mais MLT convencionais no processo de classificação (SVM por exemplo) em vez do *Softmax*. Neste processo, as características extraídas pela rede neuronal podem ser utilizadas isoladamente ou em conjunto com HCFs (fornecidas por exemplo pelos autores das bases de dados) ou com a informação clínica constante no processo da paciente. Este tipo de estratégias pode melhorar a performance da classificação (33).

Na classificação de imagens de mamografia com recurso a CNN, podem ainda ser utilizados outros métodos para ultrapassar a limitação da quantidade de dados, como por exemplo "*data augmentation*", ou para melhorar a performance da classificação, através, por exemplo, da aplicação de algoritmos de normalização (8,33,44). Estes e outros métodos são abordados seguidamente na secção sobre processamento de imagem.

### 2.2.3. Processamento de Imagem

O processamento das imagens é uma etapa fundamental na deteção e classificação de lesões. A extração de características da lesão e o melhoramento das características das imagens pode ajudar a melhorar a performance do algoritmo de classificação (28). Normalmente compreende as etapas que serão abordadas seguidamente.

#### Cropping

O *cropping* ou seleção de uma região de interesse (ROI) que inclua a lesão ajuda a reduzir o tempo necessário para treinar a rede (18), podendo ser obtido manualmente ou através de uma deteção automática (28). Geralmente os autores das bases de dados fornecem a informação das coordenadas da lesão para que os investigadores consigam cortar a imagem com pretendem e alterar o seu tamanho se acharem necessário.

Existem várias abordagens: 1) Arevalo *et al.* (2016) (28) utilizaram um ROI de 150x150 pixels tendo em conta a delimitação da região segmentada; 2) Hepsag *et al.* (2017) (18) cortaram as imagens tendo em conta as coordenadas da lesão e o raio do círculo que circunda a localização da mesma, alterando as suas dimensões para 50x50 pixels. 3) Gardezi *et al.* (2017) (26) e Zhang *et al.* (2017) (45) com base nas coordenadas das lesões, utilizaram ROI com 224x224 e 227x227 pixels, respetivamente, de forma cumprir as dimensões de input nas redes pré-treinadas VGG e AlexNet.

### Data Augmentation

As bases de dados de mamografia geralmente contém um número reduzido de amostras, uma vez que a sua construção requer um grande esforço dos profissionais e das instituições envolvidos (17,21). A *data augmentation* produz exemplos adicionais através da aplicação de transformações às imagens originais (ou *crops*), com o intuito de reduzir problemas de *overfitting*, podendo incluir rotações, translações, *flips* horizontais e verticais, *zooms* e filtros digitais (18,28).

A técnica de *data augmentation* mais utilizada é a combinação de *flips* e rotações a 0, 90, 18 e 270 graus, o qual normalmente produz 7 novas amostras preservando as suas *labels* (18,21,28,29,43). Kooi *et al.* (2017) (46) testaram o efeito da *data augmentation* e concluíram que o seu uso permite aumentar a performance da classificação de 87,5% to 92,9% (AUC). Domingues *et al.* (2018) (47) sugeriram o uso de um co-registo em torno de todas as imagens para gerar variabilidade, o que aumenta a performance de uma forma substancial (70.4% para 83.2% com o método por eles proposto - *Difference of Gaussians (DoG) method*).

### Normalização de Imagem

A melhoria da performance da classificação requer que as imagens apresentem o contraste ideal (48), o que é justificado pelo facto das imagens serem obtidas com diferentes condições de exposição e afetadas por ruído e alguns artefactos (44,48). Desta forma, muitos investigadores aplicam diferentes métodos de normalização no decorrer da etapa de processamento das imagens de mamografia (19).

Existem várias abordagens que podem incluir, por exemplo, uma equalização de histograma seguida da aplicação de um filtro de mediana (48) ou uma normalização da imagem seguida da utilização de uma *Principal Component Analysis (PCA) image whitening* (19). Outros autores propuseram o uso de uma normalização de contraste global e local (GCN e LCN) (28), assim como também se pode optar pela utilização de um algoritmo de CLAHE (*Contrast-Limited Adaptive Histogram Equalization*) (27), isoladamente ou em conjunto com um *false color enhancement* ao longo do espectro de cores da imagem (RGB), o que resultou numa melhoria substancial na performance do algoritmo de classificação (86% to 92%) (49).

### 3. Metodologia

#### 3.1. Dimensão Fractal

Primeiramente pretendeu-se avaliar se a utilização da medida de DF tradicional, a qual foi designada “DF de contorno”, conjuntamente com as quatro medidas de DF propostas, melhoravam o desempenho na classificação de lesões em mamografia de acordo com a escala BIRADS e o tipo de lesão.

A primeira medida proposta foi designada por “DF de área”, uma vez que é calculada através da contagem do número de quadrados que contêm pelo menos um pixel pertencente à lesão. Seguidamente, foram obtidas duas medidas, iguais às anteriores, mas calculadas após a aplicação de um algoritmo de dilatação/erosão: no caso de uma lesão com uma forma regular, os valores de DF inicial e final deveriam ser similares, caso contrário seriam significativamente diferentes. A última medida baseia-se na diferença normalizada entre as duas medidas DF de área antes e após a aplicação do algoritmo de dilatação/erosão.

Neste estudo foram utilizadas 230 imagens de lesões mamográficas de pacientes do sexo feminino (98,3%) e do sexo masculino (1,7%), com idades compreendidas entre os 24 e os 89 anos, sendo a média de idades de 55,2 anos. As imagens foram obtidas a partir de um equipamento KODAC®, em formato DICOM, com dimensão de matriz de 32,7 Megabytes (3584x4784 pixels), com aproximadamente 20,4 pixels por milímetro, *pixel size* de 0,05x0,05 mm<sup>2</sup> e profundidade de 16 bits por pixel (*unsigned*).

Os critérios de exclusão das imagens foram: classificação BIRADS 1; lesão incompleta na imagem selecionada; dificuldade em definir a localização da lesão; e outros tipos de lesões, tais como distorção da arquitetura do tecido mamário, cicatrizes cirúrgicas ou inflamação/infeção mamária. A distribuição relativa à escala BIRADS e ao tipo de lesão está presente nas tabelas 3 e 4.

**Tabela 3** - Distribuição das lesões consoante a escala BIRADS.

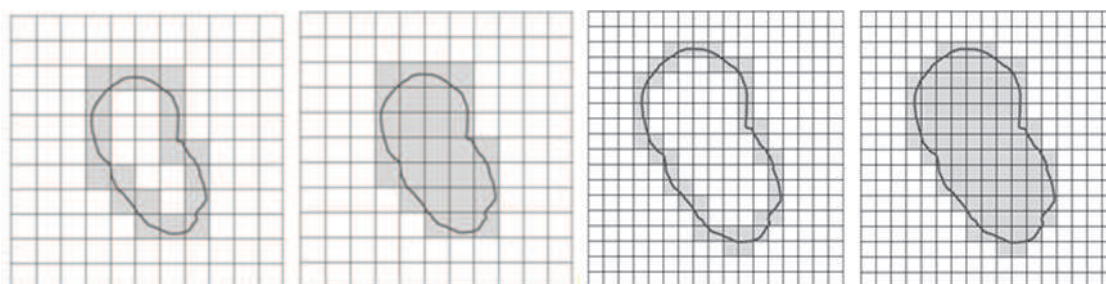
Escala BIRADS	2	3	4	5
Número de Imagens ( <i>n</i> )	181	26	6	17

Tabela 4 - Distribuição das lesões consoante o tipo.

Tipo de Lesão	Número de imagens (n)
Nódulos	133
Microcalcificações	16
Encapsulamento esboçando nodularidade	44
Calcificações	13
Áreas de maior densificação	5
Áreas de densificação irregular	15
Nódulos com contornos espiculados	5

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 foram calculadas as medidas DF através da equação 1 e após aplicação do método *box-counting*, onde  $N(s)$  é o número de quadrados contabilizados e  $1/s$  é o fator de magnificação para cada partição (de 2 a 64 pixels).

A diferença entre a DF de contorno e de área está nos quadrados que são contabilizados, no caso da DF de contorno, apenas são contabilizados os quadrados que contêm pelo menos um pixel pertencente ao contorno da lesão, enquanto na DF de área são contabilizados todos os quadrados que possuem pelo menos um pixel pertencente à lesão, processo repetido para os diferentes tamanhos de quadrados ( $1/s$ ), como é exemplificado na Figura 21.



**Figura 21** - 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).

Relativamente ao algoritmo de dilatação/erosão, a operação de dilatação foi aplicada para 20 iterações consecutivas, seguidas de erosão com mesmo número de iterações, ambas as operações foram realizadas tendo em conta os 8 vizinhos mais próximos de cada pixel (Figura 22).



**Figura 22** - Exemplo da aplicação do algoritmo de dilatação/erosão a uma lesão considerada benigna (A) e a uma lesão considerada maligna (B); Imagens da esquerda: Lesão inicial; Imagens centrais: Lesão após dilatação; Imagens da direita: Lesão após erosão aplicada às lesões das imagens centrais.

Nesta fase do estudo, foram obtidas cinco medidas de DF diferentes: DF de contorno (DF0), DF de área (DF1), DF de contorno após aplicação do algoritmo de dilatação/erosão (DF0DE) e DF de área após aplicação do algoritmo de dilatação/erosão (DF1DE). Finalmente foi calculada a medida DF final (DF2) que diz respeito à diferença normalizada entre os valores de DF1 e a DF1DE, de acordo com a equação 5.

$$DF2 = \frac{DF1DE - DF1}{DF1} \quad (5)$$

O tratamento e análise dos dados foram realizados com recurso ao programa estatístico *Statistical Package for the Social Sciences* - SPSS versão 19.

## 3.2. Redes Neurais de Convolução

Como já foi referido na secção anterior, a aplicação de CNN requer uma grande quantidade de dados para que a rede seja adequadamente treinada. No entanto, na área da imagem médica é complicado conseguir grandes volumes de dados, o que se deve não só à dificuldade em aceder às imagens, atualmente ainda mais dificultada pelas recentes leis da proteção de dados, como também ao enorme trabalho e dispêndio de tempo que é necessário para catalogar as imagens com a respetiva classificação (*label*).

Desta forma, este estudo será baseado no método de *transfer learning*, pelo que serão utilizadas redes neurais pré-treinadas em diferentes tarefas, as quais serão ajustadas (*fine-tuned*) ao problema da classificação de lesões em mamografia. Os três modelos que serão aplicados foram previamente utilizados na classificação de imagens no *ImageNet ILSVRC Challenge Data*: CNN-F (*Fast, imagenet-vgg-f*) model, CNN-M (*Medium, imagenet-vgg-m*) model (50), e o *Caffe reference model* (51). Estes modelos pré-treinados foram ajustados de forma a serem aplicados na classificação de lesões benignas e malignas em imagens de mamografia. Para isso, adaptou-se o software *MatConvNet* (52) que se encontra disponível para o Matlab (especificações do sistema: Matlab R2015a and Intel i7-3820 CPU @ 3.60GHz with 32GB RAM).

A Tabela 5 apresenta as diferenças entre os três modelos pré-treinados utilizados no estudo. Nas colunas das *convolutional layers*, denominadas através da sigla 'Conv#', o conjunto '*num×size×size*' indica o número de filtros de convolução e o seu *receptive field size*. As indicações 'st.' e 'pad.' representam o *convolution stride* e o *spatial padding*, enquanto que o LRN diz respeito à *Local Response Normalization* com ou sem o *max-pooling down-sampling factor*. Nas colunas das *fully connected layers*, com a sigla 'Full#', o número indica a sua dimensionalidade, e, para além disso, a 'Full6' e a 'Full7' são regularizadas através da aplicação de *dropout*. A última camada compreende o *softmax classifier* (50).

A arquitetura do modelo CNN-F consiste em 8 camadas de aprendizagem (5 *convolutional layers* e 3 *fully connected layers*), e o processamento mais rápido é garantido pelo *stride* de 4-pixel na primeira camada de convolução. Por outro lado, a arquitetura CNN-M, na primeira camada de convolução, apresenta um *stride* mais baixo e um *receptive field* mais pequeno, e, na segunda camada de convolução, possui um *stride* de 2-pixels o que faz com que o tempo de processamento seja razoável (50). O *Caffe reference model*, tal como os modelos mencionados anteriormente, apresenta um conjunto completo de camadas e o treino da rede utiliza um *standard stochastic gradient descent algorithm*, sendo bastante utilizado na classificação de imagens (51).

**Tabela 5** – Modelos CNN pré-treinados utilizados neste trabalho de investigação (adaptado de (50)).

Archit.	Conv1	Conv2	Conv3	Conv4	Conv5	Full6	Full7	Full8
CNN-F	6411 × 11 st.4, pad.0 LRN, × 2pool	256 × 5 × 5 st.1, pad.2 LRN, × 2pool	256 × 3 × 3 st.1, pad.1	256 × 3 × 3 st.1, pad.1	256 × 3 × 3 st.1, pad.1 × 2pool	4096 dropout	4096 dropout	1000 soft- max
CNN-M	96 × 7 × 7 st.2, pad.0 LRN, × 2pool	256 × 5 × 5 st.2, pad.1 LRN, × 2pool	512 × 3 × 3 st.1, pad.1	512 × 3 × 3 st.1, pad.1	512 × 3 × 3 st.1,pad.1 × 2pool	4096 dropout	4096 dropout	1000 softmax
Caffe	96 × 11 × 11 st.4,pad.0 LRN, × 2pool	256 × 5 × 5 st.1, pad.2 LRN, × 2pool	384 × 3 × 3 st.1, pad.1	384 × 3 × 3 st.1, pad.1	256 × 3 × 3 st.1,pad.1 × 2pool	4096 dropout	4096 dropout	1000 softmax

A base de dados utilizada no estudo foi a BCDR-FM, já mencionada na secção anterior, através do download do ficheiro BCDR-F03. Apesar de existirem bases de dados com imagens digitais, optou-se pela utilização desta base de dados com imagens digitalizadas de forma a permitir a comparação com o trabalho de Arevalo *et al.* (2016) (28); para além disso, as imagens digitais têm uma resolução superior (14 bits per pixel), o que seria mais complicado do ponto de vista computacional.

Desta forma, a etapa de pré-processamento das imagens neste trabalho é similar à estratégia utilizada pelos autores supramencionados, nomeadamente: 1) *cropping* um ROI de 150×150 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. No entanto, 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 do *crop* que se encontrava fora da imagem; 2) *data augmentation* utilizando uma combinação de *flipping* e rotações a 90, 180 e 270 graus.

Também a normalização das imagens foi similar à utilizada por Arevalo *et al.* (2016)(28), a qual consiste numa GCN, obtida através da subtração da média das intensidades da imagem a cada pixel, em combinação com uma LCN.

As imagens foram divididas em três grupos: 50% para treino, 10% para validação e 40% para teste. O *input size* das imagens para os diferentes modelos foi de 224×224 pixels; e os parâmetros explorados incluíram: três *fully connected layers*, 50 épocas, uma *fc8 initially randomised layer*, cinco valores de *learning rate* (1e-2, 1e-3, 1e-4, 5e-2, 5e-3 e 5e-4), os três modelos pré-treinados (*vgg-f*, *vgg-m* e *caffe*) e o uso ou não de imagens normalizadas.

Após o ajuste dos três modelos pré-treinados utilizando os conjuntos de dados e validação (os quais compreendiam 2800 e 560 imagens, respetivamente) com e sem normalização, foram selecionados os melhores parâmetros para aplicar ao conjunto de teste (com 2240 imagens); subsequentemente, o conjunto de treino passou a ter 3360, que resultaram da fusão dos anteriores conjuntos de treino e validação. Posteriormente, foi selecionada a rede que obteve a melhor performance e foram extraídas as características (*activations*) a partir de uma ou de todas as últimas camadas, da 16 à 19. Seguidamente, as *activations* foram utilizadas para treinar uma SVM de forma a verificar se a performance de classificação poderia melhorar através

da utilização de outro tipo de classificador. As *activations* foram utilizadas separadamente, em combinações duas-a-duas, ou todas em conjunto.

Por fim, foram utilizadas também as HCFs fornecidas pelos autores da base de dados, uma vez mais para verificar se existiam melhorias nos resultados da classificação. Estas HCFs incluem: *Intensity features* ou características de intensidade (média, mediana, máximo, mínimo, desvio padrão, assimetria e curtose), *Shape features* ou características de forma (área, perímetro circularidade, alongação, y\_centro\_massa, x\_centro\_massa e forma) e *Texture features* ou características de textura (contraste, correlação e entropia). Para além disso, foi incluída também a informação da idade das pacientes e a respetiva densidade mamária. Esta informação foi utilizada em combinação ou separadamente de forma a perceber qual apresenta maior influência na performance da classificação.

A avaliação das CNN pode incluir diferentes métodos, tais como o cálculo da média dos resultados de 5 ou mais iterações (43) ou através da utilização de 5 ou mais *cross validation folds* (27,30,46,53).

Tendo em conta as medidas utilizadas, a maioria dos autores avalia a performance das suas redes utilizando a *Accuracy* ou Precisão (Acc) (18,26,27) ou apenas o valor da *Area Under the receiver operating Curve* (AUC) (28,43,45), o que pode ser um fator limitante quando se pretende fazer a comparação entre estudos. Portanto, o ideal é utilizar uma combinação de ambas as medidas (19,54).

A *Accuracy* considera todos os casos possíveis: Verdadeiros e Falsos Negativos (VN e FN) e os Verdadeiros e Falsos Positivos (VP e FP), sendo calculada utilizando a equação (2). A análise *Receiver Operating Characteristic* (ROC) calcula a área abaixo da curva - *Area Under the receiver operating Curve* (AUC), a qual varia entre 0.50 a 1.0, e relaciona a sensibilidade (SN) – Equação 3) – no eixo das ordenadas e a especificidade (SP) – Equação (4) – no eixo das abcissas. A performance da classificação e a precisão aumentam com o aumento da área abaixo da curva (27,42,53)

$$Acc = \frac{VP+VN}{VP+FP+VN+FN} \quad (2)$$

$$SN = \frac{TP}{TP+FN} \quad (3)$$

$$SP = 1 - \left( \frac{FP}{FP+TN} \right) \quad (4)$$

## 4. Resultados e Discussão

### 4.1. Dimensão Fractal

Em primeiro lugar, verificou-se se as cinco medidas de DF propostas podem auxiliar na classificação de lesões de acordo com a escala BIRADS, em segundo lugar, se a DF0 e DF1 conseguem auxiliar a distinção de acordo com o tipo de lesão. Em ambos os testes começou-se por verificar a normalidade dos dados através dos testes de *Kolmogorov-Smirnov* (K-S, para  $n \geq 50$ ) e de *Shapiro-Wilk* (S-W, para  $n < 50$ ), através dos quais se verificou que, em pelo menos uma das categorias BIRADS e em pelo menos um dos tipos de lesão, os valores de DF não seguem uma distribuição normal e, como tal, foi aplicado o teste não paramétricos de *Kruskall-Wallis* (K-W) para comparação de k-grupos independentes e o teste de *Mann-Whitney* (M-W) para comparações múltiplas.

#### 4.4.1. DF versus Escala BIRADS

A Tabela 6 demonstra os valores do coeficiente de *Pearson*,  $p$ , obtidos através da aplicação do teste *K-W* para verificar se existe relação entre as diferentes medidas DF e a escala BIRADS.

**Tabela 6** - Indicação dos valores do coeficiente de *Pearson*,  $p$ , obtidos através do teste *K-W* para as diferentes medidas DF tendo em conta a escala BIRADS.

Medida DF	Valor $p$ (Teste K-W)
DF0	0.000
DF1	0.081
DF0DE	0.000
DF1DE	0.006
DF2	0.000

Verifica-se que apenas para os valores de DF1 não existem divergências significativas consoante a escala BIRADS. Assim, foi aplicado o teste de *M-W* às restantes medidas, de forma a comparar a DF das diferentes escalas BIRADS duas-a-duas. Os resultados são apresentados nas Tabelas 7 a 10. Os símbolos ( $\neq$ ) e ( $=$ ) indicam as escalas BIRADS que são, ou não diferenciadas pela medida DF em estudo, respetivamente.

**Tabela 7** - Resultados do teste *M-W* (valores p) relativos à comparação da **DF0** das diferentes escalas BIRADS.

DF0	BIRADS 3	BIRADS 4	BIRADS 5
BIRADS 2	0.000 (≠)	0.001 (≠)	0.000 (≠)
BIRADS 3		0.067 (=)	0.008 (≠)
BIRADS 4			0.753 (=)

**Tabela 8** - Resultados do teste *M-W* (valores p) relativos à comparação da **DF0DE** das diferentes escalas BIRADS.

DF0DE	BIRADS 3	BIRADS 4	BIRADS 5
BIRADS 2	0.003 (≠)	0.002 (≠)	0.000 (≠)
BIRADS 3		0.020 (≠)	0.087 (=)
BIRADS 4			0.172 (=)

**Tabela 9** - Resultados do teste *M-W* (valores p) relativos à comparação da **DF1DE** das diferentes escalas BIRADS.

DF1DE	BIRADS 3	BIRADS 4	BIRADS 5
BIRADS 2	0.070 (=)	0.597 (=)	0.002 (≠)
BIRADS 3		0.562 (=)	0.021 (≠)
BIRADS 4			0.042 (≠)

**Tabela 10** - Resultados do teste *M-W* (valores p) relativos à comparação da **DF2** das diferentes escalas BIRADS.

DF2	BIRADS 3	BIRADS 4	BIRADS 5
BIRADS 2	0.000 (≠)	0.000 (≠)	0.000 (≠)
BIRADS 3		0.009 (≠)	0.016 (≠)
BIRADS 4			0.327 (=)

Como se pode verificar através das tabelas acima apresentadas, 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 BIRAS 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 a após a aplicação do algoritmo de dilatação/erosão tem tendência a aumentar com a escala BIRADS, como se pode verificar através do gráfico apresentado na Figura 23, o que 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 também ser utilizada na diferenciação de lesões benignas regulares de lesões malignas irregulares, apresentado resultados mais satisfatórios que a DF de área ou DF de área após a aplicação do algoritmo de dilatação/erosão testadas individualmente.

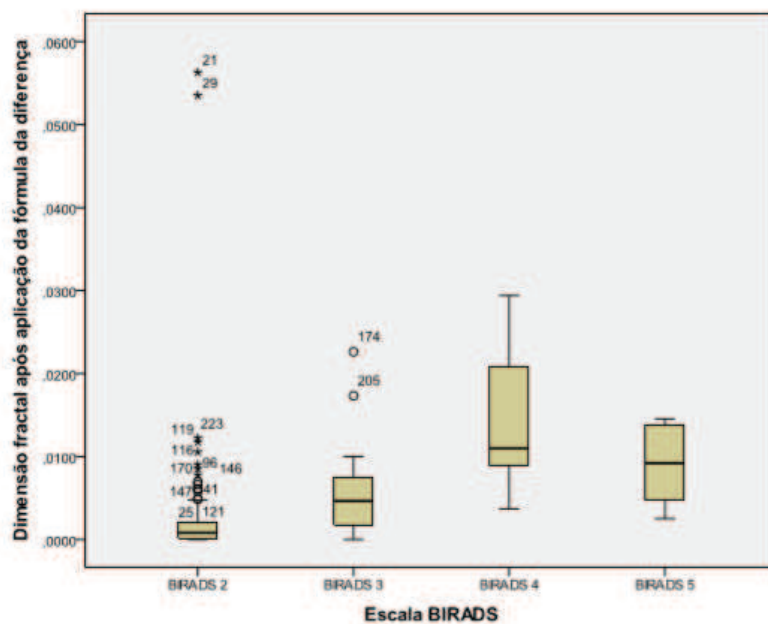


Figura 23 - Comparação dos valores médios e desvios padrão da DF2 consoante a escala BIRADS.

#### 4.4.2. DF0 e DF1 versus Tipo de Lesão

Na Tabela 11 são apresentados os valores do coeficiente de *Pearson*,  $p$ , obtidos através da aplicação do teste *K-W* para verificar se existe relação entre as DF0 e DF1 e o tipo de lesão. Sendo possível verificar que os valores de DF0 e DF1 diferem significativamente consoante o tipo de lesão.

Tabela 11 - Indicação dos valores do coeficiente de *Pearson*,  $p$ , obtidos através do teste *K-W* para as diferentes medidas DF, tendo em conta o tipo de lesão.

Medida DF	Valor $p$ (Teste <i>K-W</i> )
DF0	0.000
DF1	0.000

Desta forma, foi aplicado o teste de *M-W* para comparar as medidas DF0 e DF1 dos diferentes tipos de lesão. As Tabelas 12 e 13 apresentam os resultados deste teste, os símbolos ( $\neq$ ) e ( $=$ ) indicam os tipos de lesão que são, ou não diferenciados pela medida DF em estudo, respetivamente.

**Tabela 12** - Resultados do teste M-W (valores p) relativos à comparação da **DF0** dos diferentes tipos de lesão.

DF0	M	Enc.	C	AMD	DI	NCE
N	0.000 ( $\neq$ )	0.751 ( $=$ )	0.037 ( $\neq$ )	0.125 ( $=$ )	0.000 ( $\neq$ )	0.002 ( $\neq$ )
M		0.000 ( $\neq$ )	0.000 ( $\neq$ )	0.005 ( $\neq$ )	0.000 ( $\neq$ )	0.001 ( $\neq$ )
Enc.			0.108 ( $=$ )	0.101 ( $=$ )	0.000 ( $\neq$ )	0.002 ( $\neq$ )
C				0.571 ( $=$ )	0.008 ( $\neq$ )	0.085 ( $=$ )
AMD					0.072 ( $=$ )	0.142 ( $=$ )
DI						0.965 ( $=$ )

Abreviaturas: (AMD – Áreas de maior densificação; C – Calcificações; DI – Densificação irregular; Enc. – Encapsulamento esboçando nodularidade; M – Microcalcificações; N – nódulos; NE – Nódulos com contornos espiculados)

**Tabela 13** - Resultados do teste *M-W* (valores p) relativos à comparação da **DF1** dos diferentes tipos de lesão.

DF1	M	Enc.	C	AMD	DI	NCE
N	0.000 ( $\neq$ )	0.000 ( $\neq$ )	0.140 ( $=$ )	0.136 ( $=$ )	0.114 ( $=$ )	0.058 ( $=$ )
M		0.000 ( $\neq$ )	0.000 ( $\neq$ )	0.000 ( $\neq$ )	0.000 ( $\neq$ )	0.001 ( $\neq$ )
Enc.			0.000 ( $\neq$ )	0.794 ( $=$ )	0.079 ( $=$ )	0.597 ( $=$ )
C				0.031 ( $\neq$ )	0.008 ( $\neq$ )	0.012 ( $\neq$ )
AMD					0.230 ( $=$ )	0.221 ( $=$ )
DI						0.570 ( $=$ )

Abreviaturas: (AMD – Áreas de maior densificação; C – Calcificações; DI – Densificação irregular; Enc. – Encapsulamento esboçando nodularidade; M – Microcalcificações; N – nódulos; NE – Nódulos com contornos espiculados)

Como se pode verificar através das tabelas apresentadas, 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. A Tabela 14 demonstra que o uso combinado destas medidas reduz o número de tipos de lesão que permanecem indistinguíveis, os quais são indicados pelas áreas a cinzento escuro contendo o símbolo ( $=$ )|( $=$ ).

**Tabela 14** - Comparação dos resultados do teste de M-W relativos ao tipo de lesão, obtidos através do uso combinado das medidas **DF0 e DF1**.

DF0/1	M	Enc.	C	AMD	DI	NCE
BN	( $\neq$ ) ( $\neq$ )	( $=$ ) ( $\neq$ )	( $\neq$ ) ( $=$ )	( $=$ ) ( $=$ )	( $\neq$ ) ( $=$ )	( $\neq$ ) ( $=$ )
M		( $\neq$ ) ( $\neq$ )	( $\neq$ ) ( $\neq$ )	( $\neq$ ) ( $\neq$ )	( $\neq$ ) ( $\neq$ )	( $\neq$ ) ( $\nu$ )
Enc.			( $=$ ) ( $\neq$ )	( $=$ ) ( $=$ )	( $\neq$ ) ( $=$ )	( $\neq$ ) ( $=$ )
C				( $=$ ) ( $\neq$ )	( $\neq$ ) ( $\neq$ )	( $=$ ) ( $\neq$ )
AMD					( $=$ ) ( $=$ )	( $=$ ) ( $=$ )
DI						( $=$ ) ( $=$ )

Abreviaturas: As áreas a cinzento escuro e com os símbolos ( $=$ )|( $=$ ) representam as lesões que são indistinguíveis por ambas as medidas; o símbolo ( $\neq$ )|( $=$ ) representa as lesões que são distinguidas pela medida DF0 e não pela DF1 (AMD – Áreas de maior densificação; C – Calcificações; DI – Densificação irregular; Enc. – Encapsulamento esboçando nodularidade; M – Microcalcificações; N – nódulos; NE – Nódulos com contornos espiculados).

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  $n=5$ ).

## 4.2. Redes Neurais de Convolução

Relativamente à aplicação de CNNs, primeiramente são apresentados, nas Tabelas 15 e 16, os resultados dos testes realizados para definição dos melhores parâmetros a aplicar na classificação de lesões benignas e malignas em imagens de mamografia.

**Tabela 15** - Tabela resumo dos testes realizados para definição dos melhores parâmetros a aplicar na classificação de lesões em imagens de mamografia. Valores de AUC média obtidos após 5 repetições, utilizando **imagens com normalização** (apenas foram utilizados os conjuntos de treino e validação).

Network	Batch size	Learning rate	Top1err train mean	Top1err validacion mean	AUC mean	AUC std
CNN-F	256	1e-2	0.208	0.267	0.752	0.002
CNN-F	256	1e-3	0.302	0.360	0.720	0.006
CNN-F	256	1e-4	0.371	0.500	0.638	0.007
CNN-F	256	5e-2	0.371	0.500	0.646	0.037
CNN-F	256	5e-3	0.231	0.252	<b>0.752</b>	0.000
CNN-F	256	5e-4	0.356	0.479	0.678	0.000
CNN-M	64	1e-2	0.112	0.304	0.734	0.008
CNN-M	64	1e-3	0.186	0.286	0.732	0.001
CNN-M	64	1e-4	0.321	0.381	0.738	0.008
CNN-M	64	5e-2	0.332	0.359	0.739	0.040
CNN-M	64	5e-3	0.133	0.290	0.739	0.003
CNN-M	64	5e-4	0.212	0.263	<b>0.743</b>	0.005
Caffe	256	1e-2	0.205	0.251	0.758	0.001
Caffe	256	1e-3	0.298	0.342	0.716	0.004
Caffe	256	1e-4	0.371	0.500	0.687	0.014
Caffe	256	5e-2	0.329	0.334	<b>0.775</b>	0.014
Caffe	256	5e-3	0.224	0.251	0.753	0.000
Caffe	256	5e-4	0.337	0.458	0.696	0.036

**Tabela 16** - Tabela resumo dos testes realizados para definição dos melhores parâmetros a aplicar na classificação de lesões em imagens de mamografia, valores de AUC média obtidos após 5 repetições, utilizando **imagens sem normalização** (apenas foram utilizados os conjuntos de treino e validação).

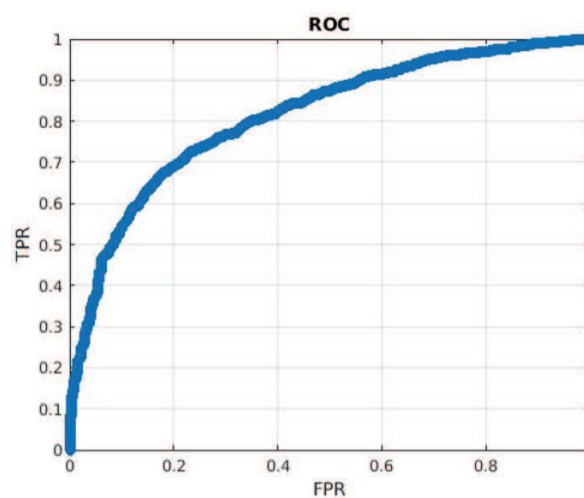
Network	Batch size	Learning rate	Top1err train mean	Top1err validacion mean	AUC mean	AUC std
CNN-F	256	1e-2	0.203	0.326	<b>0.763</b>	0.004
CNN-F	256	1e-3	0.254	0.361	0.748	0.018
CNN-F	256	1e-4	0.371	0.500	0.686	0.013
CNN-F	256	5e-2	0.370	0.483	0.745	0.005
CNN-F	256	5e-3	0.206	0.310	0.762	0.006
CNN-F	256	5e-4	0.336	0.420	0.696	0.013
CNN-M	64	1e-2	0.095	0.265	0.757	0.019
CNN-M	64	1e-3	0.150	0.265	0.781	0.005
CNN-M	64	1e-4	0.232	0.353	0.765	0.003
CNN-M	64	5e-2	0.203	0.334	0.742	0.040
CNN-M	64	5e-3	0.107	0.270	0.760	0.009
CNN-M	64	5e-4	0.170	0.270	<b>0.785</b>	0.003
Caffe	256	1e-2	0.177	0.323	0.767	0.003
Caffe	256	1e-3	0.230	0.376	0.765	0.004
Caffe	256	1e-4	0.371	0.500	0.680	0.021
Caffe	256	5e-2	0.345	0.425	0.740	0.013
Caffe	256	5e-3	0.190	0.325	<b>0.769</b>	0.002
Caffe	256	5e-4	0.281	0.381	0.756	0.003

Com os conjuntos de treino e validação, o melhor valor médio de AUC foi obtido utilizando o *Caffe reference model* (AUC média = 0.775, std = 0.014), seguido pelo CNN-F (AUC média = 0.752, std < 0.001) e, por último, o CNN-M (AUC média = 0.743, std = 0.005). estes resultados são algo surpreendentes uma vez que a CNN-M é um modelo mais robusto que a CNN-F (na CNN-M os filtros são mais complexos). Relativamente aos conjuntos de treino e validação criados após a normalização das imagens, o melhor valor médio de AUC foi obtido pelo modelo CNN-M (AUC média = 0.785, std = 0.003), seguido pelo *Caffe reference model* (AUC média = 0.769, std = 0.002) e pelo modelo CNN-F (AUC média = 0.763, std = 0.004). O modelo CNN-M apresentou uma melhoria significativa na performance de classificação antes e após a normalização das imagens (de um valor médio de AUC de 0.743 para 0.785, respetivamente).

A partir do momento que foi determinada a melhor combinação de parâmetros para cada modelo, foram obtidos novos resultados com o conjunto de teste (e o novo conjunto de treino que derivou da fusão dos conjuntos de treino e validação usados anteriormente), os quais se encontram apresentados na Tabela 17. A Figura 24 mostra o gráfico com a curva ROC do teste que obteve o melhor valor de AUC com a aplicação do modelo *Caffe* em imagens sem normalização.

**Tabela 17** – Resultados da aplicação dos três modelos CNN pré-treinados a imagens de mamografia com e sem normalização. O treino foi realizado com a fusão dos conjuntos de treino e validação, sendo testado no conjunto de teste. Os valores que se encontram a negrito correspondem ao melhor resultado obtido para cada modelo.

Network	Batch size	Epochs	Learning rate	Norm	Top1err train mean	Top1err test mean	AUC mean	AUC std	Time approx. (min.)
CNN-F	256	50	5e-3	Yes	0.196	0.378	0.721	0.001	28.31
CNN-F	256	50	1e-2	No	0.171	0.296	<b>0.767</b>	0.002	25.81
CNN-M	64	50	5e-4	Yes	0.173	0.355	0.733	0.002	76.84
CNN-M	64	50	5e-4	No	0.146	0.288	<b>0.776</b>	0.006	76.25
Caffe	256	50	5e-2	Yes	0.324	0.488	0.584	0.004	27.28
Caffe	256	50	5e-3	No	0.190	0.251	<b>0.813</b>	0.001	25.24



**Figura 24** – Exemplo do gráfico com a curva ROC do melhor resultado obtido na aplicação do modelo *Caffe* em imagens sem normalização (AUC = 0.814).

O melhor desempenho foi obtido pelo *Caffe reference model* nas imagens em que não foi aplicado o processo de normalização, com um valor médio de AUC de 0.813 (std = 0.001), ultrapassando o valor de 0.79 obtido por Arevalo *et al.* (2016) (28), o qual foi obtido com o uso combinado do *DeCAF model*, de imagens normalizadas e da SVM como classificador em vez da *Softmax layer* (uma vez que eles consideram que a SVM apresenta um melhor desempenho na classificação do que *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 usados os *crops* das imagens após normalização.

Tal como aconteceu com o conjunto de validação, as melhores performances foram obtidas quando foram utilizadas as imagens sem normalização, com os valores médios de AUC de 0.776 com a CNN-M e 0.767 com a CNN-F, valores esses que são similares aos alcançados na fase de ajuste ou *fine-tuning*. 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, principalmente quando foi aplicado o modelo *Caffe*, o qual foi significativamente mais baixo (AUC média = 0.584; anteriormente tinha sido 0.775).

Após a obtenção dos valores de AUC médios apenas com recurso a CNNs, foi selecionada a rede que apresentou os melhores resultados, o *Caffe reference model*, e foram extraídas as características ou *activations* das últimas camadas, nomeadamente da camada 16 à 19. Seguidamente, estas *activations* foram fornecidas a uma SVM de forma a verificar os resultados da classificação sofriam melhorias. Para além disso, foi testada a combinação das diferentes características extraídas (duas-a-duas, três-a-três, ou todas em conjunto). Os resultados são apresentados na Tabela 18.

**Tabela 18** – Resultados da classificação de lesões com a recurso a uma SVM após extração das características das imagens mamográficas sem normalização com o *Caffe reference model* (valores médios obtidos após 5 repetições). Resultados a negrito correspondem aos melhores valores médios de AUC obtidos.

Act. layer	AUC	Accuracy	Sensibility	Specificity
16	<b>0.773</b>	0.702	0.538	<b>0.844</b>
17	0.749	0.685	0.557	0.794
18	0.766	0.689	0.523	0.839
19	0.732	0.678	0.549	0.781
16&17	0.767	0.702	0.546	0.835
16&18	<b>0.772</b>	0.700	0.529	0.846
16&19	0.760	0.695	0.542	0.817
17&18	0.760	0.692	0.535	0.823
17&19	0.744	0.679	0.551	0.786
18&19	0.755	0.687	0.527	0.822
16&17&18	0.769	0.700	0.541	0.836
16&17&19	0.761	0.695	0.545	0.817
16&18&19	0.762	0.694	0.539	0.823
17&18&19	0.754	0.689	0.530	0.823
16&17&18&19	0.763	0.693	0.540	0.827

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; *best run* AUC = 0.784), e na combinação das características extraídas das camadas 16 e 18 (AUC média = 0.772; *best run* AUC = 0.786). A melhor precisão ou *accuracy* foi obtida com a utilização das características extraídas da camada 16 ou da combinação destas com as características extraídas da camada 17; a sensibilidade é bastante reduzida para todas as combinações testadas, o melhor resultado foi alcançado ao se utilizarem as características extraídas da camada 17 (0.557); quanto à especificidade, o melhor resultado foi obtido ao utilizar a combinação das características extraídas das camadas 16 e 18 (0.846).

Como os melhores resultados de AUC foram alcançados com a utilização das características da camada 16, foram realizados testes com a combinação destas com as HCFs e a informação da idade e densidade mamária das pacientes, cujos resultados se encontram na Tabela 19.

**Tabela 19** - A influência da idade, densidade e HCFs (Intensidade, Forma e Textura) 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 obtidos.

	Standalone	Age	Density	Age&density
Age	0.766	–	–	–
Density	0.773	–	–	–
Age&density	0.764	–	–	–
Intensity (I)	0.774	0.766	0.772	0.764
Shape (S)	<b>0.796</b>	0.788	<b>0.796</b>	0.788
Texture (T)	0.782	0.775	0.780	0.773
I&S	<b>0.796</b>	0.789	<b>0.796</b>	0.788
I&T	0.779	0.772	0.778	0.770
S&T	<b>0.801</b>	0.794	<b>0.800</b>	0.793
I&S&T	<b>0.799</b>	0.794	<b>0.799</b>	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 HCFs que apresentam uma maior influência no processo de classificação (AUC média = 0.796 e 0.782, respetivamente), o que é concordante com o que já foi mencionado na secção anterior, formas irregulares e texturas heterogêneas estão altamente relacionadas com a malignidade das lesões da mama (15,41). 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 classificação.

## 5. 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. Deste modo, os melhores resultados foram obtidos através da utilização das medidas DF2 e DF0, as quais demonstraram a melhor performance na distinção de lesões. Esta diferenciação de lesões é ainda mais evidente quando se combinam as diferentes medidas DF utilizadas neste estudo. A medida tradicional de DF0, já implementada por outros investigadores como Crisan *et al.* (2007) (39) e Nguyen *et al.* (2005) (35), aumenta com a escala BIRADS e sofre alterações consideráveis consoante o tipo de lesão. A utilização desta medida obteve melhores resultados do que uma das medidas por nós proposta DF1, com a qual não foi possível fazer a distinção das lesões pela escala BIRADS, no entanto, quando combinadas, podem melhorar a capacidade de distinguir as lesões pelo seu tipo.

Nesta fase do estudo existiram algumas limitações, nomeadamente: a segmentação manual das lesões que pode ter afetado os resultados; a dimensão da amostra pois, apesar de termos 181 casos de lesões com classificação BIRADS 2, nas restantes categorias a dimensão era significativamente mais reduzida (BIRADS 3  $n=26$ , BIRADS 4  $n=6$  e BIRADS 5  $n=17$ ) assim como para o tipo de lesão; a versão 19 do programa de estatística SPSS que ainda não tem disponíveis os testes de comparações múltiplas não-paramétricos.

Paralelamente, na aplicação das CNN para classificação de lesões em imagens de mamografia, 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. Nos casos em que foram utilizadas imagens após a normalização, com o conjunto de teste os resultados baixaram substancialmente quando comparados 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) (28) reduzem a performance da classificação das redes utilizadas neste estudo. Perre *et al.* (2018) (44) consideram que, dependendo do tipo de modelo CNN pré-treinado utilizado, os métodos de normalização escolhidos podem alterar a performance da classificação.

O facto de os *crops* 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 que possam confundir a rede.

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)(28), 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 das HCFs às 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 influencia 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.

## 6. Perspetivas Futuras

Como estudos futuros propõe-se, em primeiro lugar, uma análise mais exaustiva da DF de área e da DF de contorno, concomitantemente com a utilização de algoritmos de dilatação/erosão, utilizando diferentes números de iterações, assim como a medida da DF2, as quais podem ser utilizadas, por exemplo, como informação complementar a fornecer a uma MLT, como a SVM que foi utilizada neste estudo, de forma a melhorar a performance da classificação das lesões em imagens de mamografia.

Em segundo lugar, podem ser utilizados outros tipos de classificadores, aos quais seriam fornecidas as características extraídas da CNN como input, como por exemplo *Random Forest*, *Logistic Regression* ou *Decision Trees*. Para além da classificação binária, poderia ser acrescentada uma nova *label*, denominada 'normal', a qual permitiria classificar tecido mamário normal e aumentar assim a quantidade de dados fornecidos à rede para treino, uma vez que seria possível incluir um maior número de *crops* das imagens originais com e sem lesões.

## 7. Bibliografia

1. Direção Geral de Saúde. Norma da DGS 051/2011 [Internet]. 2011 [cited 2019 Aug 7]. p. 1–17. Available from: <https://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas/norma-n-0512011-de-27122011-jpg.aspx>
2. Kohli A, Jha S. Why CAD Failed in Mammography. *J Am Coll Radiol* [Internet]. 2018;15(3):535–7. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000427667000009&site=eds-live>
3. Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. *Eur J Radiol* [Internet]. 2018;105:246–50. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0720048X18302250&site=eds-live>
4. Rodrigues C. *Sistemas CAD em Patologia Mamária*. Universidade do Porto; 2008.
5. Giannakopoulou G, Spyrou GM, Antaraki A, Andreadis I, Koulocheri D, Zagouri F, et al. Downgrading BIRADS 3 to BIRADS 2 category using a computer-aided microcalcification analysis and risk assessment system for early breast cancer. *Comput Biol Med* [Internet]. 2010 Nov;40(11–12):853–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0010482510001344>
6. Brennan PC, Ganesan A, Eckstein MP, Ekpo EU, Tapia K, Mello-Thoms C, et al. Benefits of Independent Double Reading in Digital Mammography: A Theoretical Evaluation of All Possible Pairing Methodologies. *Acad Radiol* [Internet]. 2018; Available from: <http://www.sciencedirect.com/science/article/pii/S1076633218303313>
7. Keen JD, Keen JM, Keen JE. Utilization of Computer-Aided Detection for Digital Screening Mammography in the United States, 2008 to 2016. *J Am Coll Radiol* [Internet]. 2018;15(1):44–8. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S1546144017310888&site=eds-live>
8. Le EP V, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. *Clin Radiol* [Internet]. 2019;74(5):357–66. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0009926019301163&site=eds-live>
9. Katzen J, Dodelzon K. A review of computer aided detection in mammography. *Clin Imaging* [Internet]. 2018;52:305–9. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000454569100049&site=eds-live>
10. National Institutes of Health. Mammography [Internet]. NATIONAL INSTITUTE OF BIOMEDICAL IMAGING AND BIOENGINEERING. 2017 [cited 2019 Aug 7]. p. 1–2. Available from: [https://www.nibib.nih.gov/sites/default/files/Mammography\\_Fact\\_Sheet\\_2017\\_0.pdf](https://www.nibib.nih.gov/sites/default/files/Mammography_Fact_Sheet_2017_0.pdf)

11. Bontrager KL. Tratado de Técnica Radiológica e Base Anatômica. 5ª Edição. Guanabara Koogan, editor. Rio de Janeiro; 2001. 576 p.
12. Balleyguier C, Ayadi S, Van Nguyen K, Dromain C, Sigal R. BIRADS™ classification in mammography. *Eur J Radiol* [Internet]. 2007 Feb 1 [cited 2019 Aug 8];61(2):192–4. Available from: <https://www.sciencedirect.com/science/article/pii/S0720048X0600475X?via%3Dihub>
13. Ehsanbaksh AR, Sailanian F, Khorashadizadeh N. Different BIRADS Categories in Screening and Diagnostic Mammograph. *Iran J Radiol* [Internet]. 2009;6(3):119–23. Available from: [https://www.researchgate.net/publication/238755846\\_Different\\_BIRADS\\_Categories\\_in\\_Screening\\_and\\_Diagnostic](https://www.researchgate.net/publication/238755846_Different_BIRADS_Categories_in_Screening_and_Diagnostic)
14. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast Tissue Composition and Susceptibility to Breast Cancer. *JNCI J Natl Cancer Inst* [Internet]. 2010 Aug 18;102(16):1224–37. Available from: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djq239>
15. Pisco JM. Imagiologia Básica – Texto e Atlas. 1ª Edição. LIDEL, editor. Lisboa; 2003.
16. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012 Feb;19(2):236–48.
17. Rodriguez JH, Fraile FJC, Conde MJR, Llorente PLG. Computer Aided Detection and Diagnosis in Medical Imaging: A Review of Clinical and Educational Applications. In: *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* [Internet]. New York, NY, USA: ACM; 2016. p. 517–24. (TEEM '16). Available from: <http://doi.acm.org/10.1145/3012430.3012567>
18. Hepsağ PU, Özel SA, Yazıcı A. Using deep learning for mammography classification. In: *2017 International Conference on Computer Science and Engineering (UBMK)*. 2017. p. 418–23.
19. Jiao Z, Gao X, Wang Y, Li J. A deep feature based framework for breast masses classification. *Neurocomputing* [Internet]. 2016 Jul 12;197:221–31. Available from: <http://www.sciencedirect.com/science/article/pii/S0925231216003611>
20. Bakkouri I, Afdel K. Breast tumor classification based on deep convolutional neural networks. In: *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. 2017. p. 1–6.
21. Zhang X, Zhang Y, Han EY, Jacobs N, Han Q, Wang X, et al. Whole mammogram image classification with convolutional neural networks. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2017. p. 700–4.
22. Chougrad H, Zouaki H, Alheyane O. Deep Convolutional Neural Networks for breast cancer screening. *Comput Methods Programs Biomed* [Internet]. 2018;157:19–30. Available from: <http://www.sciencedirect.com/science/article/pii/S0169260717301451>

23. Carneiro G, Nascimento J, Bradley AP. Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning. *IEEE Trans Med Imaging*. 2017;36(11):2355–65.
24. Agarwal R, Diaz O, Lladó X, Yap MH, Martí R. Automatic mass detection in mammograms using deep convolutional neural networks. *J Med Imaging [Internet]*. 2019 Feb;6(3):1–9. Available from: <https://doi.org/10.1117/1.JMI.6.3.031409>
25. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data [Internet]*. 2017 Dec 19;4:170177. Available from: <https://doi.org/10.1038/sdata.2017.177>
26. Gardezi SJS, Awais M, Faye I, Meriaudeau F. Mammogram classification using deep learning features. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA). 2017. p. 485–8.
27. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A. Three-Class Mammogram Classification Based on Descriptive CNN Features. *Biomed Res Int [Internet]*. 2017 Jan 15;1–11. Available from: <http://10.0.4.131/2017/3640901>
28. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Guevara Lopez MA. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed [Internet]*. 2016 Apr;127:248–57. Available from: <http://www.sciencedirect.com/science/article/pii/S0169260715300110>
29. Wichakam I, Vateekul P. Combining deep convolutional networks and SVMs for mass detection on digital mammograms. In: 2016 8th International Conference on Knowledge and Smart Technology (KST). 2016. p. 239–44.
30. Zhu W, Lou Q, Vang YS, Xie X. Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification. 2016 Dec 18 [cited 2017 Feb 1]; Available from: <http://arxiv.org/abs/1612.05968>
31. Li H, Meng X, Wang T, Tang Y, Yin Y. Breast masses in mammography classification with local contour features. 2017;16(1 OP-Publisher: BioMed Central Country of Publication: England NLM ID: 101147518 Publication Model: Electronic Cited Medium: Internet ISSN: 1475-925X (Electronic) Linking ISSN: 1475925X NLM ISO Abbreviation: Biomed Eng Online Subsets: MEDLINE):44. Available from: <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=mnh&AN=28410616>
32. Ribli D, Horvath A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. *Sci Rep [Internet]*. 2018;8. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000427457400001&site=eds-live>
33. Zou L, Yu S, Meng T, Zhang Z, Liang X, Xie Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Comput Math Methods Med [Internet]*. 2019;1–16. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=135530629&site=eds-live>
34. Dobrescu R. Diagnosis of Breast Cancer from Mammograms by Using Fractal

- Measures. *Int J Med Imaging* [Internet]. 2013;1(2):32. Available from: <http://www.sciencepublishinggroup.com/journal/paperinfo.aspx?journalid=156&doi=10.11648/j.ijmi.20130102.14>
35. Nguyen TM, Rangayyan RM. Shape Analysis of Breast Masses in Mammograms via the Fractal Dimension. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference [Internet]. IEEE; 2005. p. 3210–3. Available from: <http://ieeexplore.ieee.org/document/1617159/>
  36. Sztojnov I, Alexandra D, Popescu C, Voine V. Image Processing in Biology Based on the Fractal Analysis. In: *Image Processing* [Internet]. InTech; 2009. Available from: <http://www.intechopen.com/books/image-processing/image-processing-in-biology-based-on-the-fractal-analysis>
  37. Sankar D, Thomas T. Analysis of mammograms using fractal features. In: 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC) [Internet]. IEEE; 2009. p. 936–41. Available from: <http://ieeexplore.ieee.org/document/5393875/>
  38. Rangayyan RM, Nguyen TM. Fractal Analysis of Contours of Breast Masses in Mammograms. *J Digit Imaging* [Internet]. 2007 Sep 6;20(3):223–37. Available from: <http://link.springer.com/10.1007/s10278-006-0860-9>
  39. Crisan DA, Dobrescu R, Planinsi P. Mammographic Lesions Discrimination Based on Fractal Dimension as an Indicator. In: 2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services [Internet]. IEEE; 2007. p. 74–7. Available from: <http://ieeexplore.ieee.org/document/4381156/>
  40. Yassin NIR, Omran S, El Houby EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Comput Methods Programs Biomed* [Internet]. 2018;156:25–45. Available from: <http://www.sciencedirect.com/science/article/pii/S0169260717306405>
  41. Yi D, Sawyer RL, Cohn III D, Dunnmon J, Lam C, Xiao X, et al. Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors [Internet]. 2017. Available from: <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsarx&AN=1705.06362>
  42. Yassin NIR, Omran S, El Houby EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Comput Methods Programs Biomed* [Internet]. 2018;156:25–45. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000424764800004&site=eds-live>
  43. Perre AC, Alexandre LA, Freire LC. Lesion classification in mammograms using convolutional neural networks and transfer learning. *Comput Methods Biomech Biomed Eng Imaging Vis* [Internet]. 2018 Jul 26;1–7. Available from: <https://doi.org/10.1080/21681163.2018.1498392>
  44. Perre AC, Alexandre LA, Freire LC. The Influence of Image Normalization in

- Mammographic Classification with CNNs. In: APPIS 2018 - 1st International Conference on Applications of Intelligent Systems. 2018.
45. Zhang X, Sasaki T, Suzuki S, Takane Y, Kawasumi Y, Ishibashiz T, et al. Classification of mammographic masses by deep learning. In: 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). 2017. p. 793–6.
  46. Kooi T, B van G, Karssemeijer N, A den H. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. [Internet]. Vol. 44, Medical Physics. 2017. p. 1017–27. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=28094850&site=eds-live>
  47. Domingues I, Abreu PH, Santos J. Bi-Rads Classification of Breast Cancer: A New Pre-Processing Pipeline for Deep Models Training [Internet]. 2018 25th IEEE International Conference on Image Processing (ICIP), Image Processing (ICIP), 2018 25th IEEE International Conference on. 2018. p. 1378–82. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edsee&AN=edsee.8451510&site=eds-live>
  48. Rouhi R, Jafari M, Kasaei S, Keshavarzian P. Benign and malignant breast tumors classification based on region growing and CNN segmentation. Expert Syst Appl [Internet]. 2015 [cited 2017 May 1];42(3):990–1002. Available from: <http://www.sciencedirect.com/science/article/pii/S0957417414005594>
  49. Teare P, Fishman M, Benzaquen O, Toledano E, Elnekave E. Malignancy Detection on Mammography Using Dual Deep Convolutional Neural Networks and Genetically Discovered False Color Input Enhancement. J Digit Imaging [Internet]. 2017;30(4):499–505. Available from: <https://doi.org/10.1007/s10278-017-9993-2>
  50. Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the Devil in the Details: Delving Deep into Convolutional Nets. 2014 May 14; Available from: <http://arxiv.org/abs/1405.3531>
  51. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. 2014 Jun 20; Available from: <http://arxiv.org/abs/1408.5093>
  52. Vedaldi A, Lenc K. MatConvNet - Convolutional Neural Networks for MATLAB. 2014 Dec 15; Available from: <http://arxiv.org/abs/1412.4564>
  53. Abbas Q. DeepCAD: A Computer-Aided Diagnosis System for Mammographic Masses Using Deep Invariant Features. 2016;5(4 OP-Computers, Vol 5, Iss 4, p 28 (2016)):28. Available from: <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsdoj&AN=edsdoj.f885adf30cb14f59bfe5e694e82c01a9>
  54. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal [Internet]. 2017;37:114–28. Available from: <http://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S136184151730018X&site=eds-live>