

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Previsão da evolução de doença oncológica a partir da análise de imagens de PET scan

Miguel José de Melo Tavares

VERSÃO DE TRABALHO



Mestrado Integrado em Engenharia Informática e Computação

Orientador: Professor Rui Camacho

Co-orientador: Professor Pedro Henriques Abreu

27 de Janeiro de 2017

Previsão da evolução de doença oncológica a partir da análise de imagens de PET scan

Miguel José de Melo Tavares

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Professora Cristina Ribeiro

Arguente: Professora Inês Dutra

Vogal: Professor Rui Camacho

27 de Janeiro de 2017

Resumo

A resposta de um paciente oncológico a um tratamento pode ser aferida pela avaliação de imagens de PET/CT *scans*. É de notar que esta análise não é válida para todas as doenças oncológicas. Esta técnica de imagiologia médica resulta da combinação de duas técnicas: PET (*Positron Emission Tomography*), responsável por oferecer uma imagem relativa à atividade metabólica, através da injeção de isótopos radioativos; CT (*Computed Tomography*) providencia uma imagem relativa à morfologia do paciente, recorrendo a raios-x.

O objetivo deste trabalho consiste no desenvolvimento de uma abordagem que permita obter modelos capazes de prever o resultado da evolução de doenças oncológicas após tratamento. Neste trabalho foram utilizados 63 conjuntos de imagens: 34 de tumores neuroendócrinos e 29 de linfomas de Hodgkin. Os conjuntos de imagens foram fornecidos pelo Instituto Português de Oncologia do Porto.

As imagens foram reduzidas a regiões de interesse, onde as lesões estavam situadas, tendo sido depois extraídas características, das imagens PET/CT obtidas antes do tratamento, que formaram o conjunto de dados processados por técnicas de *Machine Learning*, juntamente com a informação clínica do paciente e o resultado após o tratamento. Foram testadas várias técnicas de normalização do conjunto de dados, juntamente com cinco métodos de aprendizagem supervisionada de diferentes tipos, cujos parâmetros foram alvo de otimização. Para além disso, foram utilizados dois tipos de métodos de seleção de atributos e técnicas capazes de lidar com a assimetria na distribuição das instâncias de cada classe de resultados.

Os resultados obtidos incluem previsões com *f1-score* - a métrica utilizada para a avaliação dos modelos - de 0.3745, 0.7135, 0.4143 e 0.7705 para cada um dos resultados previstos de tratamento.

O reduzido número de instâncias combinado com o desequilíbrio entra a sua distribuição pelas diferentes classes de resultados talvez tenham levado a que as performances obtidas não fossem as melhores. Assim sendo, para a continuação do trabalho no futuro, um dos passos a tomar seria a aquisição de dados de novos pacientes oncológicos.

Abstract

An oncology patient response to a treatment can be assessed by the evaluation of PET/CT scans. One should notice this analysis is not valid to every oncology disease. This medical imaging technique is the combination of two other: PET (Positron Emission Tomography), that offers an image related to the metabolic activity, through the injection of radioactive isotopes; CT (Computed Tomography) provides an image related to the patient's morphology, with the help of x-rays.

The goal of this work consists in the development of an approach to obtain models capable of predicting the outcome of the evolution of oncology diseases after its treatment. In this work were used 63 sets of PET scans: 34 from neuroendocrine tumors and 29 from Hodgkin lymphomas. This sets were provided by the Institute of Oncology of Porto.

The images were reduced to regions of interest, where the lesions were located, having been extracted features from PET/CT scans obtained before the treatment took place, creating the data set to be processed by Machine Learning techniques, alongside patients clinical data and treatment outcome. Several normalization techniques have been tested, with five supervised learning methods of different types, whose parameters were tuned. Two feature selection types were also used and techniques to deal with the class imbalance problem.

The results obtained include predictions with f1-scores - the evaluation metric used to assess models performance - of 0.3745, 0.7135, 0.4143 and 0.7705 for each of the treatment outcomes.

The reduced number of instances combined with data distribution imbalance between the different outcome classes has maybe led to under-performance. Thus, to keep developing this work in the future, one step to take would be gathering more data from new oncology patients.

Agradecimentos

Neste espaço, gostaria de agradecer a todos aqueles que, de alguma forma, contribuíram positivamente para o meu percurso académico.

Aos meus colegas de curso, por todos os momentos de camaradagem e de trabalho, que me permitiram obter muito mais do que um diploma.

Aos meus professores, que me passaram o conhecimento necessário para ultrapassar todas as etapas do meu percurso.

Aos meus orientadores, Professor Rui Camacho e Professor Pedro Abreu, pela disponibilidade e pelo aconselhamento que me permitiram desenvolver este trabalho.

A Mariana Nogueira, pela ajuda providenciada numa fase inicial do projeto.

Aos meus amigos, em especial ao Chico, Marcelo, Pedro e Rui, por todos os momentos bem passados.

À minha família, pela paciência e pelo apoio incondicional, em todas as etapas da minha vida.

Aos meus pais, irmãos e avós em especial, pelo carinho e incentivo diário.

À minha tia Assunção e aos meus primos Ana Sofia, João Pedro e Maria João em especial, por me terem acolhido e aturado durante 4 anos no Porto.

Ao meu tio Luís e primas Laurane e Sónia em especial, pela ajuda oferecida no meu semestre de mobilidade Erasmus+.

Miguel Tavares

*“Better to remain silent and be thought a fool
than to speak out and remove all doubt.”*

Maurice Switzer

Conteúdo

1	Introdução	1
1.1	Contexto	1
1.2	Objetivos	3
1.3	Estrutura do relatório	4
2	Revisão Bibliográfica	5
2.1	Processamento de imagem	5
2.1.1	Pré-processamento	5
2.1.2	Segmentação de imagem	6
2.1.3	Extração de caraterísticas	6
2.2	Data Mining	9
2.2.1	Normalização de dados	9
2.2.2	Seleção de atributos	10
2.2.3	Redes Neurais Artificiais	12
2.2.4	k Vizinhos Mais Próximos	12
2.2.5	<i>Naive Bayes</i>	12
2.2.6	Algoritmos Genéticos	13
2.2.7	Árvore de decisão	13
2.2.8	Floresta Aleatória	14
2.2.9	Máquinas de Vetores de Suporte	14
2.2.10	Programação Lógica Indutiva	15
2.2.11	Metodologias de avaliação dos modelos e métricas de desempenho	16
2.2.12	O problema dos conjuntos de dados de distribuição assimétrica	17
2.3	Literatura Relacionada	21
3	Metodologia	27
3.1	Arquitetura da solução proposta	27
3.2	Conjunto de dados inicial	27
3.3	Processamento de imagem	29
3.3.1	Segmentação da imagem	29
3.3.2	Extração de características	29
3.4	Conjunto de dados final	29
3.5	Método Experimental	31
4	Resultados e discussão	37
4.1	Otimização de parâmetros e classificação do conjunto de dados original	37
4.2	Seleção de atributos	38
4.2.1	Análise da componente principal	38

CONTEÚDO

4.2.2	Filtragem e <i>wrapping</i>	38
4.3	Métodos <i>ensemble</i> e métodos sensíveis ao custo	40
4.3.1	AdaBoost	40
4.3.2	Bagging	41
4.3.3	MetaCost	42
4.3.4	<i>Stacking</i>	43
4.4	Programação Lógica Indutiva	43
4.4.1	Total e Parcial vs Neutra e Negativa	44
4.4.2	Total vs Restantes	45
4.4.3	Total vs Neutra e Negativa	46
4.4.4	Total vs Parcial	46
5	Conclusões e Trabalho Futuro	49
5.1	Satisfação dos Objetivos	49
5.2	Trabalho Futuro	50
	Referências	51
A	Parametrização dos classificadores	59

Lista de Figuras

1.1	Imagens de CT(A), PET(B) e PET/CT(C) de um tumor neuroendócrino, o carcinoma. É possível verificar a combinação das duas técnicas na formação da imagem da PET/CT.	2
2.1	Exemplo de cálculo da GLCM (b) para uma matriz de imagem (a), com $\theta = 0^\circ$ e $d = 1$, sendo i a intensidade do píxel de referência e j a intensidade do píxel vizinho, neste caso o píxel à direita do píxel de referência	7
2.2	Exemplo de cálculo da GLRLM (b) para uma matriz de imagem (a), com $\theta = 0^\circ$ (direção horizontal), sendo i a intensidade do píxel de referência e j o comprimento da sequência do píxel de intensidade i	8
2.3	Exemplo de cálculo da GLSZM (b) para uma matriz de imagem (a), sendo i a intensidade do píxel de referência e j a área de píxeis adjacentes consecutivos de intensidade i	8
2.4	Visualização simplificada do modo de operação do SVM. O algoritmo verifica qual a melhor colocação do hiperplano de forma a maximizar a distância entre as amostras mais próximas das duas classes.	15
2.5	Demonstração de uma representação gráfica possível do desempenho de um classificador, medido por AUC. É fácil perceber que quanto mais a curva está à esquerda, maior a AUC e maior é o número de positivos corretamente classificados.	18
2.6	Visualização gráfica do processo de geração sintética de amostras com SMOTE [LFG ⁺ 13]	19
3.1	Diagrama demonstrativo da arquitetura da solução proposta	28
3.2	Variância cumulativa das componentes calculadas através de PCA	32

LISTA DE FIGURAS

Lista de Tabelas

2.1	Matriz de confusão para classificação binária. Os valores entre parênteses são utilizados na definição das métricas seguintes.	17
2.2	Sumário da literatura revista	25
3.1	Descritores e características extraídas na caracterização das regiões de interesse . .	30
3.2	Escalas de valores testados na otimização de parâmetros efetuada	31
3.3	Parâmetros dos métodos <i>wrapper</i> de seleção de atributos	33
3.4	Matriz de custo utilizada no algoritmo MetaCost	33
3.5	Classes dos exemplos utilizados na construção de hipóteses com ILP	34
3.6	Conjunto de atributos utilizado na construção de hipóteses com ILP	34
4.1	Resultados obtidos no processo de otimização	37
4.2	Parâmetros obtidos no processo de otimização	37
4.3	Resultados após aplicação de PCA ao conjunto de atributos	38
4.4	Resultados obtidos após seleção sequencial	38
4.5	Resultados obtidos após eliminação sequencial	39
4.6	Resultados obtidos após aplicação de algoritmo genético	39
4.7	Lista de descritores mais selecionados nos métodos de seleção de atributos	40
4.8	Resultados obtidos com a aplicação do algoritmo AdaBoost sobre o conjunto de dados original	40
4.9	Resultados da aplicação de AdaBoost nos conjuntos de atributos obtidos no processo de seleção	41
4.10	Resultados obtidos através da utilização de <i>Bagging</i> com o conjunto original de dados	41
4.11	Resultados obtidos da aplicação de <i>bagging</i> nos conjuntos de atributos obtidos no processo de seleção	41
4.12	Resultados obtidos pela aplicação de MetaCost no conjunto de atributos original .	42
4.13	Resultados da aplicação de MetaCost nos conjuntos de atributos obtidos no processo de seleção	42
4.14	Resultados obtidos com a aplicação de <i>Stacking</i> no conjunto de dados original . .	43
4.15	Resultados obtidos após seleção sequencial de atributos numa classificação com <i>Stacking</i>	43
4.16	Resultados obtidos com ILP, onde as classes Total e Parcial constituíram o conjunto de exemplos positivos e as restantes classes o conjunto de exemplos negativos	44
4.17	Resultados obtidos com ILP, onde a classe Total constituiu o conjunto de exemplos positivos e as restantes classes o conjunto de exemplos negativos	45
4.18	Resultados obtidos com ILP, onde as classe Total e Parcial constituiu o conjunto de exemplos positivos e as classes Neutra e Negativa o conjunto de exemplos negativos	46

LISTA DE TABELAS

4.19	Resultados obtidos com ILP, onde as classe Total constituiu o conjunto de exemplos positivos e a classe Parcial o conjunto de exemplos negativos	47
A.1	Parâmetros utilizados nos algoritmos de classificação	59

Abreviaturas e Símbolos

PET	Tomografia por emissão de positrões - <i>Positron Emission Tomography</i>
CT	Tomografia Computadorizada - <i>Computed Tomography</i>
PET/CT	Tomografia por Emissão de Positrões com Tomografia Computadorizada
FDG	Fluorodesoxiglicose
GLCM	Matriz de co-ocorrência de níveis de cinzento - <i>Grey-level co-occurrence matrix</i>
GLRLM	Matriz de sequências de níveis de cinzento - <i>Grey-level run-length matrix</i>
GLSZM	Matriz de tamanho das zonas níveis de cinzento - <i>Grey-level size zone matrix</i>
LBP	Padrões Locais Binários - <i>Local Binary Patterns</i>
PCA	Análise da componente principal - <i>Principal component analysis</i>
KDD	Descoberta de conhecimento em bases de dados - <i>Database Knowledge Discovery</i>
ANN	Redes Neurais Artificiais - <i>Artificial Neural Networks</i>
kNN	k Vizinhos Mais Próximos - <i>k Nearest Neighbours</i>
SVM	Máquinas de Vetores de Suporte - <i>Support Vector Machines</i>
ILP	Programação Lógica Indutiva - <i>Inductive Programming Logic</i>
\wedge	Conjunção Lógica (AND)
\models	Implicação Lógica
$\not\models$	Negação da Implicação Lógica
\square	Falsidade
AUC	Área sob a curva - <i>Area under the curve</i>
MRI	Ressonância Magnética - <i>Magnetic Resonance Imaging</i>
NPC	Carcinoma da nasofaringe - <i>Nasopharyngeal carcinoma</i>
RBF	Função de base radial <i>Radial Basis Function</i>
DDSM	Base de Dados Digital de Imagens de Mamografias - <i>Digital Database for Screening Mammography</i>
BI-RADS	<i>Breast Imaging-Reporting and Data System</i>
EkNN	Vizinhos Mais Próximos Evidencial - <i>k Nearest Neighbours Evidential</i>
FCM	<i>Fuzzy C-means</i>
ACM	<i>Active Contour Model</i>

Capítulo 1

Introdução

Em 2012, foram registados 14 milhões de novos casos de cancro e espera-se que esse número aumente em mais de 70% nas próximas duas décadas [SW14]. Não é assim de estranhar que ao longo dos anos, várias abordagens computacionais tenham vindo a ser aplicadas não só a patologias mais frequentes [HAACS⁺14a, HAACS⁺14b] como também a patologias mais raras (com menor incidência na população) [HAGM⁺15, HAAHA⁺16, HAHA⁺16]. Os problemas tratados nestas abordagens vão desde a escolha do melhor método para imputar dados em base de dados incompletas [GLHAHAA15], para balancear os dados [SSHAGL⁺15], para prever a recorrência da doença [HASSHA⁺16], entre outros. Contudo, o avanço tecnológico permitirá no futuro continuar a desenvolver novas abordagens dentro desta área.

1.1 Contexto

Na área oncológica, um dos métodos de diagnóstico utilizado para a deteção de células cancerígenas é a PET/CT - Tomografia por Emissão de Positrões com Tomografia Computadorizada. Esta técnica permite identificar certas doenças oncológicas em diferentes estádios. Uma imagem de PET/CT é composta pela junção de duas técnicas:

PET - Tomografia por emissão de positrões: oferece informação funcional acerca dos tecidos afetados pelas células cancerígenas. Esta imagem é obtida através da injeção de um radiofármaco no sistema do paciente. A imagem dá informação sobre as zonas com maior absorção do radiofármaco, o que pode indicar a existência de uma anomalia nesse tecido. Numa análise de uma PET, é preciso distinguir atividade natural de atividade anómala pois podem existir tecidos em que a elevada absorção de um radiofármaco é natural, isto é, está relacionada com a fisiologia do corpo humano.

CT - Tomografia Computadorizada: permite localizar anatomicamente o tecido afetado. Esta técnica utiliza a combinação de imagens raios X, obtidas de diferentes ângulos, permitindo observar o interior do paciente e verificar a sua anatomia através das diferentes densidades dos tecidos do corpo humano.

Introdução

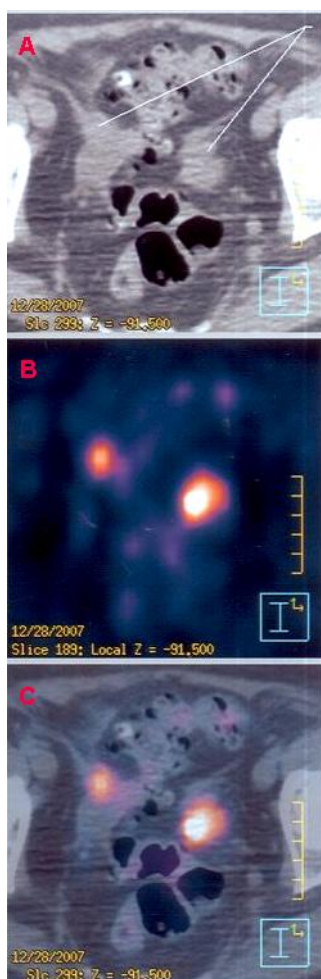


Figura 1.1: Imagens de CT(A), PET(B) e PET/CT(C) de um tumor neuroendócrino, o carcinoma. É possível verificar a combinação das duas técnicas na formação da imagem da PET/CT.

A PET/CT é sobretudo utilizada para diagnóstico e para avaliação da resposta ao tratamento onde tipicamente é realizada a comparação entre duas imagens (pré e pós tratamento).

A avaliação de imagens PET pode ser feita com recurso a medidas quantitativas para avaliar a gravidade da lesão, através da quantidade de radiofármaco absorvido pelo órgão onde a lesão está situada. Esta quantidade pode variar de caso para caso tendo em conta dois fatores: a quantidade de radiofármaco injetado no paciente e do seu peso. Isto é, para pacientes injetados com a mesma quantidade de radiofármaco, a atividade metabólica detetada será maior para o paciente com menor peso; por outro lado, para pacientes com o mesmo peso, o paciente com maior atividade metabólica detetada será aquele que tiver sido injetado com maior quantidade de radiofármaco. De forma a ter em conta este tipo de variações, foi proposta uma norma de quantificação da atividade metabólica que tem em conta estes dois fatores: Valor de Absorção Padrão (SUV - *Standardized Uptake Value*). Esta norma é bastante popular para a avaliação e estudo das imagens obtidas na PET. No entanto, esta métrica está sujeita à introdução de ruído pela ação de vários fatores [KF10]: resolução da imagem obtida, diferentes calibrações do *scanner*, erros no proce-

dimento de realização do exame, entre outros. Desta forma, os analistas acabam por efetuar, na maioria dos casos, uma avaliação manual [FBM⁺14].

As patologias em estudo são os tumores neuroendócrinos e os linfomas de Hodgkin. De forma a perceber a deteção e avaliação destas doenças numa imagem de PET/CT, é efetuada uma breve contextualização.

Tumores Neuroendócrinos: São tumores que têm origem no sistema nervoso e no sistema endócrino. Sendo detetados maioritariamente no pâncreas e no tubo digestivo, estes tumores estão associados a uma absorção acima do normal de substâncias análogas à somatostatina. O radiofármaco mais utilizado é assim o Gálio-68, por ser uma substância análoga à somatostatina. No entanto, o fato de existir uma zona com elevada absorção do radiofármaco, não é suficiente para a existência de uma anomalia. Isto porque o pâncreas é uma glândula onde é natural existir uma elevada absorção de Gálio-68.

Linfomas de Hodgkin: Os linfomas são patologias oncológicas que se desenvolvem no sistema linfático. Havendo órgãos desse sistema dispersos pelo corpo humano, como os nódulos linfáticos, um linfoma pode se desenvolver em qualquer parte do corpo do paciente. A patologia espalha-se depois pelo corpo através dos vasos linfáticos. Esta doença oncológica está associada a uma elevada atividade metabólica, que por sua vez está relacionada com uma elevada produção de glicose. Assim sendo, o radiofármaco utilizado é a Fluorodesoxiglicose (FDG), uma substância análoga da glicose. Existem áreas de maior absorção da FDG, como o coração ou o cérebro, também órgãos de elevada atividade metabólica.

A nível tecnológico, os avanços nos campos de Processamento de Imagem e *Data Mining* podem oferecer uma solução de para este problema.

O Processamento de Imagem consiste num conjunto de técnicas que permitem obter informação de interesse de uma imagem de modo a utilizar essa informação para tarefas de, por exemplo, análise de imagem, extração de características e reconhecimento de padrões, podendo estas tarefas estar, ou não, ligadas entre si.

Neste trabalho, as tarefas de processamento de imagem incidiram sobre segmentação de imagem, que permitiu obter a área da lesão, para de seguida serem extraídas características, recorrendo a alguns dos descritores mais utilizados para retirar informação de imagens médicas [NAM⁺16]. A partir do conjunto de características extraídas, foram aplicados métodos de seleção de atributos, para perceber quais os atributos que otimizam os resultados das tarefas de previsão e classificação.

1.2 Objetivos

O objetivo principal deste projeto é o desenvolvimento de uma abordagem que consiga prever com precisão a resposta de um paciente oncológico ao tratamento, através da análise de imagens de PET/CT pré e pós tratamento, conjuntamente com alguns dados clínicos do paciente. A obtenção destes objetivos compreende várias etapas:

1. Processamento de imagens que permitam extrair as regiões onde as lesões estão situadas, extraíndo características a partir destas regiões;

Introdução

2. Dos atributos extraídos, identificar o conjunto daqueles que são preditores da resposta ao tratamento que posteriormente serão complementados com a informação clínica do paciente;
3. Avaliar quais os algoritmos de *Machine Learning* mais adequados para o problema.

Neste trabalho foram utilizados 63 conjuntos de imagens: 34 de tumores neuroendócrinos e 29 de linfomas de Hodgkin. As imagens e a informação clínica utilizados no trabalho foram fornecidas pelo Instituto Português de Oncologia do Porto (IPO - Porto).

1.3 Estrutura do relatório

O documento tem a seguinte estrutura: no próximo capítulo, apresenta-se um estudo de conceitos necessários à execução do projeto. É feito um estudo da componente tecnológica do trabalho, no que diz respeito a segmentação de imagem, extração e seleção de características e algoritmos de *Machine Learning*.

No Capítulo 3 são apresentados o problema e a metodologia a ser utilizada para obter uma solução. O Capítulo 4 apresenta os resultados obtidos nas tarefas de classificação, apresentando-se no Capítulo 5 as conclusões retiradas, bem como possíveis passos de trabalho no futuro.

Capítulo 2

Revisão Bibliográfica

Neste capítulo apresentam-se os conceitos teóricos relevantes para a metodologia adotada. Para além disso, são também referidas algumas ferramentas que podem ser utilizadas e literatura relevante na área.

2.1 Processamento de imagem

O processamento de imagem pode ser dividido em três tarefas:

1. Pré-processamento - trata de preparar a imagem para o seu processamento em computador.
2. Segmentação de imagem - divisão da imagem em várias regiões através de um certo critério, determinado pelo algoritmo utilizado.
3. Extração de características de interesse - obtenção de atributos através da informação processada diretamente da imagem.

2.1.1 Pré-processamento

O pré-processamento de imagem consiste num conjunto de técnicas que têm por objetivo transformar a imagem de entrada numa outra semelhante, com propriedades mais interessantes para a extração de informação.

Este passo envolve, entre outras, as tarefas de remoção do ruído, escalamento de imagem e aumento de contraste. Neste projeto, uma das motivações é a remoção do ruído a que as imagens de PET/CT estão sujeitas, pelo que foram estudadas algumas alternativas para o fazer.

O ruído existente numa imagem é introduzido durante as etapas de aquisição e transmissão de imagem [GW06]. Exemplos desse ruído são o ruído sal e pimenta, caracterizado pelo aparecimento aleatório de píxeis pretos em áreas claras e de píxeis brancos em áreas escuras da imagem, e o ruído aditivo (gaussiano), em que cada píxel da imagem sofre uma alteração no seu valor de intensidade, para que o histograma de intensidades da imagem com ruído apresente uma distribuição normal.

A remoção de ruído é pode ser feita recorrendo à aplicação de filtros. Existem vários filtros que podem ser utilizados:

- **Passa-alto, passa-baixo e passa-banda** - filtros que passam píxeis de intensidades superiores (passa-alto) ou inferiores (passa-baixo), atenuando os restantes. Um filtro passa-banda consiste no uso simultâneo dos dois filtros anteriores.
- **Média e mediana** - divisão da imagem em regiões mais pequenas onde a intensidade de cada píxel é substituída pela intensidade média ou pela mediana das intensidades dos píxeis dessa região.

2.1.2 Segmentação de imagem

A segmentação de uma imagem é o processo que permite dividir a imagem em regiões de interesse, para que seja possível efetuar a extração das características dessa imagem. Neste caso, o objetivo trata-se de identificar as lesões que as imagens da PET/CT captaram. Entre o conjunto dos algoritmos de segmentação existentes, foram escolhidos dois, por combinarem simplicidade e performance [SS12] na sua utilização: algoritmo de crescimento de regiões [HS85] e algoritmo de limiarização [SSW88].

O algoritmo de crescimento de regiões começa pela definição de pontos semente. Um dos critérios para a escolha destes pontos pode ser a definição de uma escala de intensidades. A partir destes pontos, vão sendo adicionados os pontos vizinhos, caso se encontrem numa escala de intensidades previamente definida, e assim sucessivamente, até que não seja possível adicionar mais pontos, isto é, todos os pontos da vizinhança não cumpram o critério predefinido. Neste tipo de algoritmo é importante saber algum tipo de informação acerca da imagem para que seja possível determinar automaticamente o critério de escolha de pontos semente e de crescimento da região na vizinhança.

Já o algoritmo de limiarização consiste na obtenção de uma imagem binária a partir de uma imagem numa escala de cinzas. A binarização é efetuada pela definição de um limite global utilizado para a transformação. Os pontos abaixo desse limite são substituídos na imagem por um píxel preto, enquanto os pontos acima são substituídos por um ponto branco.

2.1.3 Extração de características

O processamento da imagem permite-nos obter um grande conjunto de dados a serem analisados. Estes dados por si só não contêm informação de interesse, sendo necessário reduzi-los a características que permitam descrevê-los com maior precisão e menor redundância. Podem ser aplicados vários descritores a uma imagem para que lhes sejam extraídas características de interesse. Apresenta-se de seguida uma lista com alguns dos descritores entre os mais utilizados em projetos relacionados com processamento de imagens médicas [NAM⁺16].

Histograma de níveis de cinza: consiste num vetor com o número de ocorrências de cada nível de cinza numa região da imagem segmentada. Algumas características extraídas deste descritor são a média, o desvio-padrão e a assimetria. Este descritor oferece uma visão da distribuição dos vários níveis de intensidades presentes na imagem. A partir do histograma, é possível aferir o realce do contraste da imagem, que está relacionado com a percepção da sua qualidade. Imagens que possuem uma baixa variabilidade de níveis intensidade terão um menor contraste, enquanto imagens com uma maior distribuição ao longo dos diferentes níveis de intensidade possuirão um maior contraste. Isto é especialmente importante ao nível de reconhecimento de padrões, pois uma imagem com maior contraste possui uma maior diferença entre os níveis de cinza entre diferentes regiões, o que permite um reconhecimento de padrões mais eficaz. Por isso este histograma é utilizado também ao nível do pré-processamento de imagem, de forma a aumentar a sua qualidade. [AR05]

GLCM: matriz que contém a informação da frequência com que diferentes combinações de níveis de cinza aparecem numa certa região da imagem. Esta matriz é caracterizada por uma direção, definida por um ângulo, e pela distância entre píxeis. Cada elemento m_{ij} da GLCM corresponde ao número de vezes que o nível de cinza j se encontra a uma distância d e na direção de ângulo θ , de um píxel de referência i . A partir das matrizes calculadas são retiradas as características texturais, designadas por características de Haralick [HSD73]. Na figura 2.1 dá-se um exemplo de cálculo da GLCM.

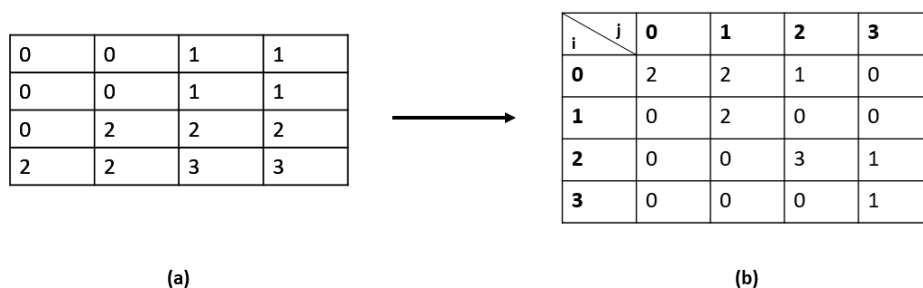


Figura 2.1: Exemplo de cálculo da GLCM (b) para uma matriz de imagem (a), com $\theta = 0^\circ$ e $d = 1$, sendo i a intensidade do píxel de referência e j a intensidade do píxel vizinho, neste caso o píxel à direita do píxel de referência

GLRLM: matriz que contém informação acerca das sequências de níveis de cinza consecutivos numa região da imagem, numa certa direção. Cada elemento m_{ij} da GLRLM corresponde ao número de ocorrências de sequências de comprimento j do nível de cinza i , na direção de ângulo θ . É extraída uma série de características de comprimento das sequências, definidas por Galloway [Gal75]. A figura 2.2 exemplifica o cálculo da GLRLM.

GLSZM [TAM14]: matriz que fornece informação acerca do tamanho das zonas de píxeis da mesma intensidade. Contrariamente ao que acontece com o descritor GLRLM, onde a análise é relativa ao comprimento das sequências de píxeis da mesma intensidade numa determinada direção, o descritor GLSZM computa uma informação acerca da área - número de píxeis consecutivos - das zonas de píxeis adjacentes todos eles com o mesmo valor de intensidade. Assim sendo, é apenas

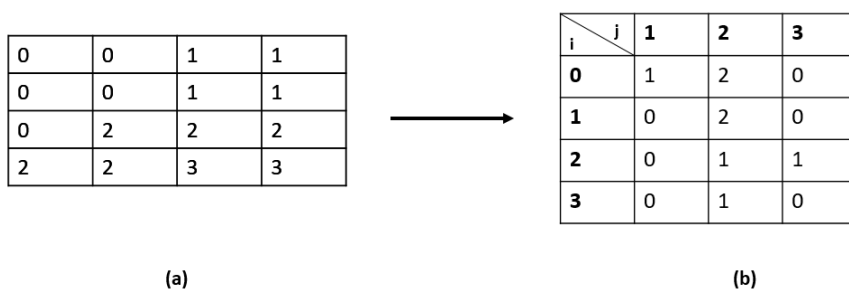


Figura 2.2: Exemplo de cálculo da GLRLM (b) para uma matriz de imagem (a), com $\theta = 0^\circ$ (direção horizontal), sendo i a intensidade do píxel de referência e j o comprimento da sequência do píxel de intensidade i

computada uma matriz, enquanto no descritor GLRLM são computadas matrizes diferentes para diferentes direções.

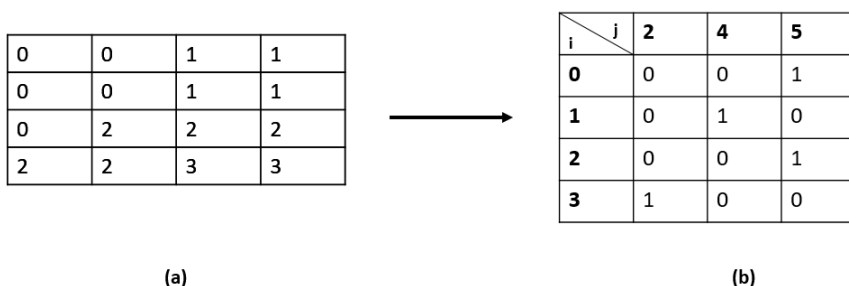


Figura 2.3: Exemplo de cálculo da GLSZM (b) para uma matriz de imagem (a), sendo i a intensidade do píxel de referência e j a área de píxeis adjacentes consecutivos de intensidade i

Wavelets: funções que permitem dividir conjuntos de dados em diferentes domínios de frequência, constituindo ferramentas poderosas de processamento de imagem, no que diz respeito a eliminação de ruído ou detecção de arestas. A decomposição é conseguida através da passagem do sinal da imagem através da aplicação de um conjunto de filtros. Numa primeira fase, as linhas da matriz inicial são passadas por um filtro passa-baixo e por um filtro passa-alto. A informação de baixa frequência contém uma aproximação da representação da imagem inicial enquanto a de altas frequências contém as arestas (são filtros usados também em detecção de arestas, por se partir do princípio que estas têm píxeis de intensidade alta associados). Neste momento, juntando as linhas que sofreram filtragem, temos duas matrizes (uma alta intensidade e outra de baixa intensidade), com o mesmo número de linhas mas metade do número de colunas da matriz inicial. São aplicados novamente os filtros, mas desta vez sobre as colunas destas duas matrizes, obtendo quatro matrizes finais, com metade das linhas e colunas da inicial.

LBP: o descritor Padrões Binários Locais funciona através da divisão da imagem em regiões e é computado para a vizinhança de cada píxel dessa região: aos píxeis vizinhos com uma intensidade inferior ao de referência é atribuído o valor 0, enquanto que aos restantes é atribuído o valor 1. A vizinhança a utilizar para a computação do descritor é definida por um raio r relativo à distância ao píxel de referência. Para $r = 1$, a vizinhança é constituída pelos 8 píxeis em volta.

Para cada píxel dessa região obtém-se uma sequência de caracteres binária que é depois convertida em decimal, criando-se um histograma com a frequência absoluta de cada número na sua forma decimal. Os histogramas obtidos para as várias regiões são concatenados para formar o vetor de características a utilizar em classificação. Existe ainda uma extensão ao algoritmo enunciado que parte do princípio de que a maior parte dos números binários obtidos contém no máximo duas transições de bits (0 para 1 ou 1 para 0), os chamados padrões uniformes. [LC07] Utilizando um histograma apenas com padrões binários uniformes, o comprimento do vetor de características é reduzido de 256 para 58, quando $r = 1$.

Filtros de Gabor: Um filtro de Gabor de duas dimensões é obtido pelo produto de uma onda sinusóide com uma função gaussiana. Para extração de características, são utilizados filtros com diferentes escalas e frequências que são convoluídos com a imagem segmentada. A partir da imagem obtida no final deste processo são obtidas características como a média, desvio-padrão, energia ou entropia.

2.2 Data Mining

Data Mining pode ser definido como o processo de descoberta de conhecimento de interesse em conjuntos de dados gigantescos, guardados em bases ou armazéns de dados. Esta definição pode ser confundida com a de KDD - Descoberta de Conhecimento em Bases de Dados - sendo o *Data Mining* um processo de KDD. [HKP11]

Este processo tem o contributo de várias áreas de conhecimento. Uma dessas áreas é o *Machine Learning*. Dentro das técnicas desta área, temos especial interesse nas de previsão e classificação, que têm por objetivo encontrar modelos que façam a distinção de objetos em valores de classes, permitindo posteriormente classificar objetos que não tenham um valor de classe associado. Um modelo é baseado num conjunto de dados previamente anotados com o valor da classe a que pertencem, denominado de conjunto de treino. Os modelos obtidos podem ser apresentados sob a forma de regras de decisão, árvores de decisão ou redes neuronais. [HKP11]

De seguida, são apresentados não só os métodos de previsão e classificação que podem ser utilizados, mas também técnicas de normalização de dados e de seleção de atributos que podem ser testados, com objetivo de melhorar a eficiência e a performance da classificação.

2.2.1 Normalização de dados

Em conjuntos de dados com atributos de significados e gamas diferentes, torna-se importante recorrer a técnicas de normalização de dados de forma a permitir a comparação mais eficiente entre os valores dos conjuntos de atributos das amostras. Apresentam-se quatro técnicas que podem ser utilizadas para esse fim [HKP11].

Transformação Z: esta técnica de normalização converte os dados fornecidos num conjunto de dados que segue uma distribuição normal de média 0 e variância 1. A transformação de um

valor x é feita segundo a fórmula 2.1, onde μ corresponde à média dos valores do atributo e σ ao seu desvio-padrão.

$$Z = \frac{x - \mu}{\sigma} \quad (2.1)$$

Transformação em gama: esta técnica de normalização tem por objetivo reduzir a gama de valores do atributo ao intervalo $[0,1]$ ou $[-1,1]$ aplicando a fórmula 2.2, onde x corresponde ao valor a ser normalizado:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.2)$$

Transformação em proporção: neste método é calculada a proporção de cada valor do atributo sobre a soma total dos valores do atributo. A soma dos valores do atributo após normalização deve ser 1.

Transformação em gama interquartil: versão semelhante à transformação em gama mas, em vez de se utilizarem os valores máximo e mínimo para a normalização do valor, são utilizados os valores relativos ao primeiro e terceiro quartis dos valores do atributo. A utilização dos valores dos quartis pretende reduzir o efeito da existência de valores discrepantes (*outliers*) nos valores do atributo.

2.2.2 Seleção de atributos

Não serão utilizadas todas as características extraídas das imagens para as descrever. É necessário proceder à seleção das características mais adequadas para o processo de classificação, que são aquelas que possuam uma maior variância, facilitando a separação dos conjuntos de dados em valores de classe. A utilização de características redundantes ou irrelevantes não introduz, portanto, qualquer vantagem para o processo de classificação. Para além disso, a redução do conjunto global de características a um subconjunto mais relevante permite efetuar testes de menor exigência computacional. Existem dois tipos de métodos principais [KJ97] definidos com o objetivo de encontrar os subconjuntos de atributos que otimizam a performance dos modelos de previsão e classificação. Para além disso, apresenta-se um outro método baseado na conversão de atributos em componentes de modo a projetar o conjunto de dados num espaço de menor número de dimensões.

2.2.2.1 Métodos de filtragem

Este tipo de abordagem tem por objetivo a remoção de atributos com base nas suas propriedades intrínsecas, sendo independente do algoritmo de classificação a ser utilizado.

Remoção de atributos correlacionados: Este método trata de eliminar atributos que estejam correlacionados com um outro. Para os pares de atributos são calculados coeficientes de correlação. A remoção de atributos é efetuada caso o coeficiente calculado se situe acima de um determinado limite, sendo eliminado um dos atributos desse par.

ReliefF: ReliefF [KŠRŠ97] é um algoritmo que faz uma estimativa da relevância dos atributos do conjunto de dados, tendo em conta a distinção dos seus valores entre as instâncias, o vizinho mais próximo da mesma classe da instância (*near-hit*) e os vizinhos mais próximos das restantes (*near-misses*). Quanto maior for a capacidade dos valores do atributo de fazerem esta distinção, maior seu o valor de relevância para as tarefas de previsão e classificação do conjunto de dados. Os atributos com um valor de relevância inferior ao limiar escolhido são eliminados do conjunto de atributos.

2.2.2.2 Métodos *wrapper*

Os métodos *wrapper* têm por objetivo encontrar o subconjunto de atributos que apresenta uma melhor performance para um certo método de classificação, reduzindo assim a dimensionalidade do conjunto de dados. Existem vários algoritmos relativamente à obtenção dos subconjuntos de atributos a testar:

Procura exaustiva: consiste na procura do melhor subconjunto de atributos, testando todas as combinações possíveis de atributos. É possível constatar que, para um conjunto de n atributos, existem $2^n + 1$ subconjuntos de atributos, que para conjuntos de dados com um elevado número de atributos, é uma abordagem computacionalmente cara.

Seleção sequencial: ao contrário do que acontece com a procura exaustiva, este algoritmo é ganancioso, isto é, vai localizando a solução ótima localmente na esperança de encontrar a solução ótima global. Mais concretamente, o processo começa por escolher o atributo que apresentou melhor performance no treino do modelo. Continua depois a adicionar atributos ao subconjunto solução, testando em cada ronda a adição de um atributo, selecionando aquele que apresenta uma maior melhoria na performance da classificação. Quando é atingido o máximo de atributos a escolher ou a adição de um qualquer novo atributo não oferece uma melhoria da performance, está obtida a solução.

Eliminação sequencial: este algoritmo é também ganancioso, mas faz o percurso inverso da seleção sequencial. Começa pelo conjunto completo de n atributos e verifica qual o subconjunto de atributos $n - 1$ apresenta melhor performance na classificação após treino. Os atributos vão sendo removidos até que o número máximo de atributos a eliminar for atingido ou se se verificar que a remoção de um qualquer atributo do subconjunto solução leva a um decréscimo na performance do modelo.

2.2.2.3 PCA - Análise da componente principal

Este método trata de encontrar as componentes principais dos dados, isto é, as direções de maior variância dos dados. Com isto, e partindo do princípio que componentes com maior variância de dados permitem facilitar a separação entre valores de classe, os dados são projetados sobre um novo sistema de eixos, correspondentes ao respetivo vetor próprio. As componentes de maior variância são aquelas que tiverem os maiores valores próprios, sendo que os vetores próprios a eles associados têm as direções de maior variância dos dados. As componentes são combinações

lineares dos atributos do conjunto de dados inicial. Assim sendo, podemos concluir que este método trata-se mais exatamente de transformação de atributos, em vez de um método de seleção propriamente dita.

2.2.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (ANN) são inspiradas no sistema nervoso central de um ser vivo animal. Este sistema apresenta um conjunto de neurónios interconectados que recebe, processa e responde a um impulso externo. Assim sendo, podemos definir uma ANN como um conjunto de camadas de neurónios interligadas entre si, que recebem os valores de entrada com pesos atribuídos. Depois disso, uma função de ativação recebe a soma dos pesos da entrada e retorna um valor como resposta a esta soma, que vai corresponder ao valor de saída do neurónio. Este valor vai ser fornecido como entrada aos neurónios da próxima camada ou, caso se trate da última camada, corresponde a um dos valores de saída da rede.

No treino de uma rede neuronal verifica-se se os valores de saída da rede correspondem aos valores da classe da amostra de treino. Em cada amostra do treino, os pesos são modificados de forma a minimizar o erro entre as previsões da rede e o valor esperado para aquela amostra. [HKP11]

2.2.4 k Vizinhos Mais Próximos

A designação deste algoritmo é bastante explícita quanto ao seu comportamento. Utilizando um conjunto de treino como base, são calculadas as distâncias das amostras desconhecidas a cada uma das amostras do conjunto de treino. Depois disso, verificam-se quais são as k amostras com valores mais próximos da amostra a classificar. Neste conjunto, o valor de classe com maior frequência absoluta vai corresponder ao valor de classe atribuído à amostra desconhecida. Pode ser adicionado um peso no processo de votação, para que o voto de instâncias mais próximas seja mais tido em conta que as restantes.

2.2.5 Naive Bayes

Naive Bayes é um método de classificação probabilístico que utiliza o teorema de Bayes (2.1), que trata do cálculo das probabilidades de acontecimentos tendo em conta o conhecimento, *a priori*, dos fatores que podem influenciar a probabilidade da ocorrência desse mesmo acontecimento. Este algoritmo é designado de ingénuo (*naive*) por assumir que os atributos do conjunto de dados são independentes entre si.

$$P(A|B) = \frac{P(B|A) * P(B)}{P(B)} \quad (2.3)$$

Para além das probabilidades calculadas *a priori*, a classificação por intermédio desta técnica faz também uso do conceito de possibilidade (*likelihood*). Este conceito utiliza a noção de proximidade, tal como o algoritmo kNN. Neste caso, é tão mais possível uma instância pertencer a uma classe quanto mais instâncias dessa classe existirem na vizinhança.

A combinação das probabilidades calculadas *a priori*, juntamente com as probabilidades calculadas através do conceito de possibilidade, permitem obter as probabilidades *a posteriori* de uma instância pertencer a cada classe. A classe atribuída corresponderá à maior probabilidade *a posteriori*.

2.2.6 Algoritmos Genéticos

Os algoritmos genéticos são uma classe de algoritmos que tem por base as ideias de evolução e seleção natural. Não são apenas utilizados para classificação, mas também para outros problemas de otimização - como é o caso da seleção de atributos - e para testar a aptidão de outros algoritmos. [HKP11, MMAHA14]

O algoritmo começa pela definição de uma população inicial aleatória, que corresponde a uma variedade de sequências de caracteres binários geradas aleatoriamente. Estas sequências podem representar um grupo de atributos escolhidos, no exemplo da seleção de atributos (onde os 1's representam os atributos selecionados e os 0's aqueles que não foram). Cada uma destas sequências é avaliada por uma função de aptidão -quantifica o quão "boa" é a solução para o problema em questão -, que determina a probabilidade de uma sequência ser escolhida para o processo de reprodução. Uma função de aptidão pode ser, no caso da seleção de atributos, a performance do conjunto de atributos escolhido num certo classificador.

De seguida é simulado o processo de reprodução, que consiste na combinação de pares de sequências binárias para produzir novos descendentes, que vão formar a nova geração. A variabilidade é garantida através do *crossing-over*, onde são efetuadas trocas de subsequências de caracteres binários entre as regras progenitoras, ou por mutação, onde caracteres binários escolhidos aleatoriamente são alterados.

Como condição de paragem da execução do algoritmo, pode ser selecionado um limite máximo para o resultado da função da aptidão, após o qual o algoritmo para, ou um número máximo de gerações a serem produzidas.

2.2.7 Árvore de decisão

Uma árvore de decisão é uma estrutura semelhante a um grafo orientado, onde os nós internos constituem testes a serem feitos aos valores dos atributos, as arestas indicam um dos possíveis resultados do teste e as folhas da árvore são os valores de classe obtidos como resultado final da classificação efetuada pelo modelo. [HKP11]

A construção da árvore de decisão segue os seguintes passos:

1. É criado um nó que representa todas as amostras de treino.

2. Caso as amostras estejam todas anotadas com a mesma classe, esse nó torna-se a única folha da árvore, tendo como resultado o valor de classe das amostras.
3. Havendo amostras pertencentes a mais do que uma classe, é tempo de escolher qual o atributo que melhor vai dividir o conjunto de amostras. Procede-se ao cálculo do ganho de informação baseado na entropia. O atributo com melhor valor de ganho de informação será utilizado para separar as amostras. Este atributo será utilizado para teste no nó criado e, caso seja contínuo, deve ser alvo de discretização.
4. São criadas várias arestas para os diferentes valores conhecidos, ou gamas de valores no caso de se ter procedido a discretização desses valores.
5. O algoritmo executa o processo dos passos anteriores recursivamente, até que seja encontrada uma das condições de terminação da sua execução:
 - As amostras do nó interno a testar pertencem todas à mesma classe
 - Não existe mais nenhum atributo para separar as amostras restantes. Neste caso, é criada uma folha com o resultado correspondente à classe mais representada nesse conjunto de amostras
 - Não existindo amostras correspondentes a uma condição definida numa das arestas, é criada uma folha com o resultado correspondente à classe mais representada no conjunto de amostras do nó interno a que a aresta está ligada

2.2.8 Floresta Aleatória

As árvores de decisão são estruturas que mapeiam os conjuntos de características e que determinam um certo valor de classe, oferecendo uma visualização ampla do processo de classificação das amostras. No entanto, as árvores de decisão tendem a adaptar-se em demasia ao conjunto de amostras de treino. [Ho95]

Para resolver este problema, foi desenvolvido o algoritmo Floresta Aleatória, que baseia-se na construção de várias árvores de decisão. Cada uma delas é formada a partir de subconjuntos do conjunto de treino, formados a partir de substituição com amostras do conjunto original. Esta técnica é conhecida por *bagging* [Bre96]. Para classificar a amostra desconhecida, os vetores de características são fornecidas a cada uma das árvores da floresta. O valor da classe mais escolhido pelo conjunto de árvores corresponde à classificação da amostra.

2.2.9 Máquinas de Vetores de Suporte

Máquinas de vetores de suporte (SVM) são classificadores que constroem um hiperplano entre os conjuntos de dados de diferentes valores de classe, para que a distância entre os conjuntos de amostras de cada uma das classes seja o maior possível. As SVM são classificadores binários, isto é, fazem a divisão das amostras de dados em duas classes. No entanto, o problema em estudo é de múltiplas classes, pois tem como objetivo classificar as amostras em um de quatro resultados de

tratamento possíveis. Existem algumas estratégias para utilizar SVM em classificação de múltiplas classes, das quais destacamos duas baseadas em classificação binária. [HL02]

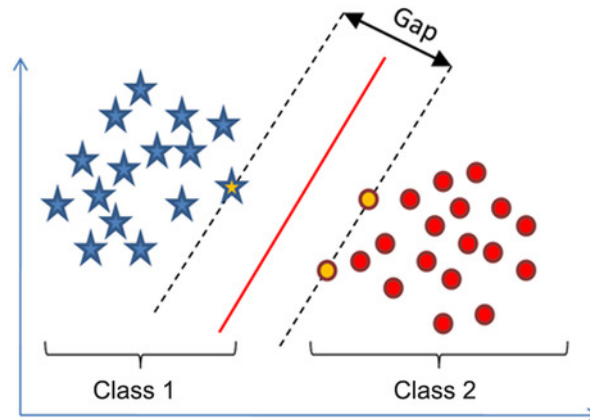


Figura 2.4: Visualização simplificada do modo de operação do SVM. O algoritmo verifica qual a melhor colocação do hiperplano de forma a maximizar a distância entre as amostras mais próximas das duas classes.

- **Decomposição Um-contra-Todos:** esta estratégia trata de criar uma SVM por cada classe, tratando de distinguir as amostras dessa classe das amostras restantes. O valor de classe atribuído à amostra será aquele cujo classificador obteve o valor máximo entre todos.
- **Decomposição Todos-contra-Todos:** esta decomposição trata de criar SVM para cada par de valores de classe. Para obter o valor de classe da amostra desconhecida, procede-se a uma votação com os valores de classe obtidos pela classificação em todas as SVM. Atribui-se à amostra o valor mais votado.

2.2.10 Programação Lógica Indutiva

A programação lógica indutiva é uma área de pesquisa que resulta da intersecção dos campos de Programação Lógica e *Machine Learning*. [Mug91] Da Programação Lógica, a ILP recebe as representações formais de conhecimento, como teorias, hipóteses ou exemplos. O *Machine Learning* contribui com as ferramentas e técnicas para a criação de hipóteses e conhecimento a partir de conjuntos de amostras. Assim, os algoritmos de ILP utilizam um conjunto de exemplos E , com exemplos positivos E^+ e negativos E^- e uma base de conhecimento prévio B para gerar uma hipótese h que explique E segundo B . As condições para a geração correta de uma hipótese h são as seguintes [Mug99]:

- Necessidade: $B \neq E^+$
- Suficiência: $B \wedge h \models E^+$
- Consistência fraca: $B \wedge h \not\models \square$

- Consistência forte: $B \wedge h \wedge E^- \not\equiv \square$

Traduzindo para linguagem natural, a condição de Necessidade impõe que não deve ser gerada uma hipótese caso seja possível explicar os exemplos positivos recorrendo apenas à base de conhecimento prévio. A condição de Suficiência impõe que uma hipótese seja capaz de explicar todos os exemplos positivos. A condição de Consistência Forte impede a geração de hipóteses inconsistentes com o conjunto de exemplos negativos, tendo em conta a base de conhecimento. Por fim, a condição de Consistência Fraca, implicada a partir da condição de Consistência Forte, indica que caso não existam exemplos negativos, a base de conhecimento e a hipótese gerada devem ser consistentes entre si.

2.2.11 Metodologias de avaliação dos modelos e métricas de desempenho

Neste projeto, é de vital importância conseguir avaliar o desempenho dos modelos gerados, de forma a poder comparar e perceber quais os métodos e técnicas mais apropriados para a resolução do problema em mãos. Para isso, são criados subconjuntos de amostras do conjunto original, chamados conjuntos de teste. Neste passo, é preciso ter o cuidado de não utilizar o conjunto de amostras de treino do modelo, para que não sejam obtidos resultados demasiado otimistas [HKP11].

Para evitar este problema, existem várias técnicas de geração de amostras de teste que permitem avaliar o desempenho do modelo com maior precisão:

Holdout: este método consiste na divisão do conjunto de amostras em dois conjuntos disjuntos: um conjunto de treino e o outro é o conjunto de teste. Normalmente, o conjunto de treino corresponde a dois terços do total de amostras. As restantes amostras constituem o conjunto de treino. [HKP11]

Validação cruzada k -fold: esta técnica divide o conjunto de amostras em k conjuntos disjuntos entre si, efetuando o treino e o teste k vezes. Durante cada iteração um dos sub-conjuntos é utilizado para o teste, enquanto os restantes são usados para derivar o modelo. Cada subconjunto é usado para teste do modelo apenas uma vez.

Bootstrapping: esta é uma técnica que consiste em amostragem aleatória com substituição. São criados diferentes conjuntos de amostra, com tamanho igual à original, onde foram introduzidas substituições aleatórias entre amostras. Isto leva ao aparecimento de instâncias repetidas num mesmo conjunto após a aplicação da técnica. Normalmente é utilizado para treino um conjunto com o mesmo número de amostras do conjunto inicial, com as substituições efetuadas. Os valores que foram substituídos e não fazem parte do conjunto de treino são utilizados para o teste do modelo obtido, correspondendo a um terço do número de instâncias do conjunto original. Este processo é repetido várias vezes para uma mesma validação.

No que diz respeito à avaliação numérica do desempenho dos modelos de classificação das amostras de teste, apresentamos uma série de métricas propostas por Sokolova et al. [SL09] para classificação com múltiplas classes. Estas fórmulas baseiam-se na generalização das métricas usadas em classificação binária.

Tabela 2.1: Matriz de confusão para classificação binária. Os valores entre parênteses são utilizados na definição das métricas seguintes.

		Previsto como	
		Positivo	Negativo
Classe Real	Positivo	Real positivo (tp)	Falso Negativo (fn)
	Negativo	Falso Positivo (fp)	Real Negativo (rn)

Nas métricas seguintes, o valor de l corresponde ao número de valores de classes existentes.

Exatidão (*accuracy*) média: taxa média de classificações corretas por cada valor de classe.

$$\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i} \quad (2.4)$$

Error Rate: taxa média de classificações incorretas por cada valor de classe.

$$\sum_{i=1}^l \frac{fp_i + fn_i}{tp_i + fn_i + fp_i + tn_i} \quad (2.5)$$

Precisão (*precision*): média das taxas, para cada valor de classe, de exemplos positivos corretamente classificados sobre os exemplos classificados como positivos pelo modelo

$$\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i} \quad (2.6)$$

Sensibilidade (*recall*): média das taxas, para cada valor de classe, de exemplos positivos corretamente classificados relativamente ao número exemplos positivos do conjunto de amostras

$$\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i} \quad (2.7)$$

F1-score: métrica que combina as métricas precisão e sensibilidade.

$$\frac{2 * \text{Precisão} * \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (2.8)$$

Área sob a curva - AUC: representação gráfica utilizada como medida de desempenho de um classificador binário. Neste caso pode ser feita uma generalização para cada valor de classe, tal como aconteceu nas restantes métricas. Quanto maior for a AUC, melhor terá sido o desempenho do classificador. A figura 2.5 mostra um exemplo de representação gráfica de AUC. [SL09]

2.2.12 O problema dos conjuntos de dados de distribuição assimétrica

Estamos na presença de um conjunto de dados com distribuição assimétrica quando existem diferenças relevantes na frequência relativa de instâncias das diferentes classes. Acontece que se verifica que a performance de certos classificadores é consideravelmente menor quando operam sobre conjuntos de dados desta natureza. [GFB⁺12, LFG⁺13]

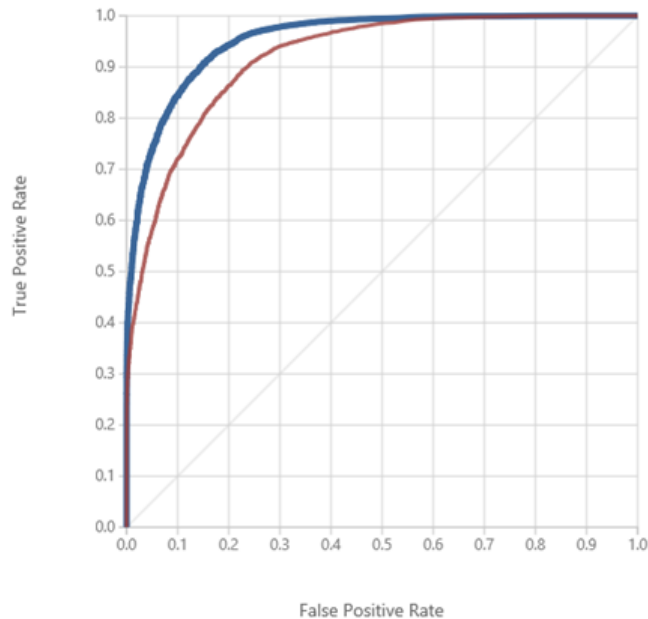


Figura 2.5: Demonstração de uma representação gráfica possível do desempenho de um classificador, medido por AUC. É fácil perceber que quanto mais a curva está à esquerda, maior a AUC e maior é o número de positivos corretamente classificados.

Este problema é bastante discutido no campo de *Machine Learning* neste momento, sendo uma área em claro crescimento a nível de investigação realizada. [HG09] A área de diagnóstico médico conta já com alguma investigação nesse sentido. [FCPB07, MHZ⁺08, LCC⁺08, RD13]

Quanto às causas apontadas para a menor performance de técnicas de previsão e classificação, quando os modelos são derivados a partir de conjuntos de dados assimétricos [LFG⁺13]:

- As medidas globais de performance, como a exatidão (*accuracy*), enviesam a construção e validação de modelos para as classes com maior número de instâncias
- As regras de classificação de instâncias das classes minoritárias são bastante específicas, cobrindo um número muito pequeno de exemplos, pelo que acabam por não fazer parte do modelo final
- Pequenos grupos com atributos semelhantes podem ser considerados como ruído, não sendo tidos em conta pelo classificador, podendo estes pequenos grupos constituir parte das instâncias da classe minoritária

Relativamente às abordagens utilizadas para lidar com o problema em questão, López et al. [LFG⁺13] sugerem três tipos:

- Diferentes amostragens do conjunto de dados
- Aprendizagem sensível ao custo

- Técnicas *ensemble*

No que diz respeito às diferentes amostragens do conjunto de dados, existem abordagens de subamostragem e sobre amostragem. Ao nível da subamostragem, a ideia principal consiste em remover instâncias aleatórias das classes maioritárias de forma a equilibrar a distribuição de dados pelas classes. No entanto, esta abordagem apresenta o problema de poder remover informação potencialmente importante para o processo de aprendizagem, especialmente em conjuntos de dados com poucas amostras. No que diz respeito à sobre amostragem, a ideia é análoga à anterior, com a diferença que o equilíbrio da distribuição dos dados dos diferentes valores da classe é efetuada ao nível da classe minoritária. Uma das abordagens existentes consiste em criar cópias exatas de instâncias aleatórias dessa classe. Outra abordagem, designada SMOTE [CBHK02], consiste na criação de instâncias sintéticas a partir de uma instância real, tendo em conta um número de vizinhos mais próximos a ser escolhido, efetuando interpolações aleatórias entre a instância real e as instâncias vizinhas. A figura 2.6 exemplifica graficamente o processo, onde x_i é a instância real, x_{i1} , x_{i2} , x_{i3} e x_{i4} representam os vizinhos mais próximos e r_1 , r_2 , r_3 e r_4 as instâncias geradas sinteticamente. Um problema existente com as técnicas sobre amostragem é o de poder levar à ocorrência de *overfitting* nos modelos derivados.

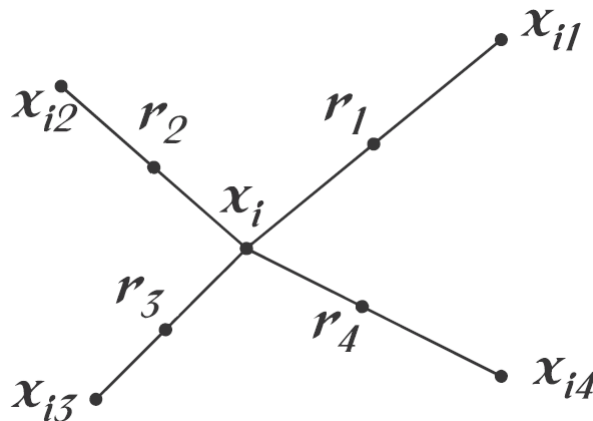


Figura 2.6: Visualização gráfica do processo de geração sintética de amostras com SMOTE [LFG⁺13]

No que diz respeito às técnicas de aprendizagem sensível ao custo, este tipo de métodos consiste na atribuição de custos de penalização a erros de classificação. Com isto pretende-se dar mais importância a certos tipos de erros e evitar que sejam tão frequentes. No problema dos conjuntos de dados de distribuição assimétrica, a ideia consiste em penalizar os erros de classificação das instâncias das classes minoritárias para favorecer a sua correta classificação. Um dos métodos de aprendizagem sensível ao custo existente é o MetaCost. [Dom99] Este método baseia-se em obter a previsão ótima de Bayes para cada exemplo, que consiste na minimização do risco condicional $R(i|x)$, o custo esperado de prever que um exemplo x pertence à classe i , calculado pela equação (2.7). Para este cálculo é necessário saber a probabilidade do exemplo x pertencer à classe j . $C(i, j)$ é o custo de prever que um exemplo pertence à classe j quando na verdade pertence à classe

i. Estes custos são normalmente fornecidos sob a forma de uma matriz de custo. É preciso notar que o cálculo dos custos de erros na classificação nem sempre é direto, nomeadamente nos casos onde os custos não são quantificáveis. Por exemplo, no caso da classificação de dados clínicos, é complicado atribuir custos às avaliações erradas do estado do paciente, contrariamente ao que acontece, por exemplo, com o caso da deteção de fraude de cartão de crédito, onde existem valores monetários associados às falhas da deteção de fraude, que podem ser utilizados para o cálculo da matriz de custos.

$$R(i|x) = \sum_j P(j|x)C(i, j) \quad (2.9)$$

Para estimar as probabilidades de cada classe $P(j|x)$, o algoritmo MetaCost utiliza *bagging*. [Bre96] Com esta técnica, são derivados vários modelos, com conjuntos de treino ligeiramente diferentes entre cada si - obtidos por *bootstrapping* do conjunto de dados original -, fazendo uma previsão através de um processo de votação por parte de cada classificador. No caso do MetaCost, para estimar as probabilidades $P(j|x)$, são calculadas as frações do número de votos de cada classe relativamente ao número total de votos.

No que diz respeito às técnicas *ensemble*, estas utilizam vários classificadores com o objetivo de obter melhores performances relativamente à utilização de um classificador isolado. Dentro do universo deste tipo de técnicas, destacamos *Bagging*, *AdaBoost* e *Stacking*.

Bagging [Bre96]: esta técnica já foi abordada anteriormente, pois é utilizada no algoritmo Floresta Aleatória e no meta-algoritmo MetaCost. A ideia por trás desta técnica é a de gerar vários conjuntos de treino através de *bootstrapping* para derivar vários classificadores do mesmo tipo. Uma instância é depois classificada segundo esses modelos, sendo atribuída ao exemplo a classe mais votada entre todos.

AdaBoost [FS97]: este é um método de *boosting*, tende a beneficiar a correta classificação de exemplos classificados erroneamente em iterações anteriores do algoritmo. No *AdaBoost*, a cada instância do conjunto de dados é atribuído um peso. Este peso é aumentado cada vez que a instância é classificada incorretamente e determina a probabilidade dessa instância fazer parte de um dos conjuntos de treino gerados em cada iteração do algoritmo, para a formação de novos classificadores. Quanto maior for o peso, maior será a probabilidade da instância ser selecionada. Isto significa que as instâncias que tenham sido incorretamente classificadas mais vezes terão uma maior probabilidade de serem escolhidas na derivação dos novos modelos. A execução do algoritmo começa com os pesos iguais para cada instância, o que significa que a escolha de instâncias para os conjuntos de treino da primeira iteração é completamente aleatória. Para além disso, são também atribuídos pesos a cada um dos classificadores conforme a sua performance em rondas anteriores. São atribuídos maiores pesos a classificadores que tenham apresentado melhores performances. Estes pesos influenciam a votação, sendo que o voto de classificadores com um maior peso terá maior influência no resultado final da votação.

Stacking [Wol92]: contrariamente aos dois métodos enunciados anteriormente, *Stacking* permite a combinação de diferentes tipos de modelos de base, cujas previsões são combinadas para

treino de um agregador. Esta técnica opera da seguinte forma:

1. O conjunto de treino é dividido em dois conjuntos disjuntos, isto é, sem elementos em comum
2. Uma das divisões do passo 1. é utilizada para treinar os classificadores de base
3. A outra divisão do passo 1. tem como função testar os modelos de base derivados no passo 2.
4. As previsões efetuadas por cada modelo no passo 3. constituem o conjunto de atributos de treino do modelo de *stacking*

2.3 Literatura Relacionada

Depois de feito um estudo relativo aos conceitos teóricos importantes para o desenvolvimento do projeto, passamos à análise da literatura onde esses mesmos conceitos foram aplicados na prática. Este estudo foi feito ao nível das técnicas utilizadas, dos resultados obtidos e das conclusões retiradas.

No que diz respeito ao processamento e classificação de informação proveniente de imagens obtidas por técnicas de imagiologia médica, Nogueira et al. [NAM⁺16] fazem uma revisão aprofundada do tema, sendo que algumas das conclusões retiradas a partir da literatura revista são as seguintes:

- A combinação dos descritores de imagem juntamente com a informação clínica do paciente pode melhorar a performance da classificação
- Os descritores mais apropriados a serem utilizados dependem da lesão oncológica em estudo
- A técnica de classificação utilizada influencia a performance dos modelos derivados
- Ao nível das imagens da PET, é de notar o problema relacionado com os órgãos mostram maior absorção de radiofármaco naturalmente e não devido à existência de uma lesão oncológica. Assim, combinar valores obtidos nas imagens CT, juntamente com informação relativa à localização anatômica da lesão, pode ajudar a separar as áreas lesionadas das áreas com uma elevada absorção natural de radiofármaco, melhorando a performance da classificação

Relativamente à revisão da literatura, apresentam-se uma série de trabalhos relacionados com classificação de imagens de várias técnicas de imagiologia médica: PET, Ressonância Magnética (MRI) - técnica que utiliza radiação não ionizante com um forte campo magnético para obter imagens detalhadas de órgãos internos - e Mamografia - técnica que utiliza raios x juntamente com compressões da mama, de forma a uniformizar o tecido mamário para facilitar a penetração da radiação.

Revisão Bibliográfica

Farhidzadeh et al. [FKS⁺16] propuseram um método para prever falhas no tratamento de carcinomas da nasofaringe (NPC), sendo o prognóstico dividido em duas classes de pacientes, relativamente ao aparecimento, ou não, de metástases distantes da área da lesão principal. Neste projeto foram utilizadas imagens de 25 MRI de pacientes com NPC. Foram desenhados manualmente os contornos da lesão para cada imagem, sendo depois aplicado um método de segmentação por limiarização de imagens, o método de Otsu [Ots79], dividindo as imagens em regiões de intensidades alta e baixa, tendo sido computados para cada uma das regiões os seguintes descritores: GLCM, GLRLM e LBP. O conjunto de dados final foi formado por 288 características obtidas a partir dos descritores referidos. Relativamente aos métodos de seleção de atributos utilizados, foram utilizados dois métodos de dois tipos diferentes: ReliefF e um método de *wrapping*. No primeiro método foi selecionado o atributo com melhor peso atribuído pelo algoritmo, enquanto a aplicação do método de *wrapping* obteve o par de atributos que otimizava a execução do classificador utilizado, SVM com núcleo de função de base radial (RBF). A técnica de validação do classificador utilizada foi validação cruzada com *leave-one-out* - variante de validação cruzada *k-fold*, onde a iteração, um subconjunto é deixado de fora do processo de validação. LBP foi o descritor que obteve melhores performances para os dois tipos de regiões, com AUC's de 0.6 e 0.76, referentes à seleção com ReliefF.

Kumar et al. [KBV15] utilizaram 480 mamografias de diferentes classes, relativas a 120 pacientes de quatro classificações BI-RADS - padrão de classificação de cancro da mama - diferentes, retiradas da Base de Dados Digital de Imagens Mamográficas (DDSM) [HBK⁺98]. Para cada imagem foi delimitada uma região de 128 por 128 píxeis, centrada no centro da imagem original. A extração de características foi feita por intermédio de *wavelets*, tendo sido utilizados diferentes tipos de filtros. A partir de cada região foram obtidas 16 imagens, correspondentes a dois níveis de decomposição. De cada uma destas foram extraídas 3 características: média, desvio-padrão e entropia. Assim, cada instância do conjunto dados foi composta por 48 atributos. Na fase de experimentação, o classificador utilizado foi SVM, tendo sido utilizados diferentes combinações de atributos para derivar modelos. A abordagem proposta consiste na utilização de filtros de Haar, com um conjunto de 32 atributos, relativos ao desvio-padrão e à entropia calculados, com uma exatidão global de 73.7% e exatidões por classe de 75%, 68.3%, 55% e 96.6% para as classes BI-RADS I, II, III e IV, respetivamente.

Lian et al. [LRD⁺16] propuseram uma versão melhorada do sistema de previsão de tratamento de doenças oncológicas proposto num artigo anterior [LRD15], onde é utilizada a teoria de Dempster-Shafer [Dem67] [Sha76] e os seus conceitos de massa e crença ao nível da seleção de atributos e para introduzir uma formulação alterada no algoritmo kNN, designada de kNN Evidencial (EkNN). Neste trabalho procedeu-se à análise de imagens PET obtidas após injeção de FDG, obtidas antes e durante a aplicação do tratamento de três doenças oncológicas: cancro do pulmão (25 instâncias - 52 atributos), cancro do esófago (36 instâncias - 29 atributos) e linfoma (45 instâncias - 27 atributos). As regiões de interesse foram obtidas ou através de métodos de limiarização ou por delimitação manual por parte dos técnicos analistas. A partir destas imagens foram obtidas características de textura por intermédio da computação do descritor GLSZM, juntamente

com os valores da SUV em cada PET. Além disso, procedeu-se à recolha de informação clínica do paciente e de dados relativos à sua expressão genética. A experimentação consistiu na comparação do método desenvolvido no artigo, comparando com algumas abordagens populares no que diz respeito a seleção de atributos, como ReliefF ou Seleção Sequencial. SVM foi o classificador utilizado para a comparação. Os resultados obtidos, validados através de *bootstrapping* e validação cruzada *leave-one-out* demonstraram que o método proposto superou os restantes, inclusive a abordagem inicialmente proposta em [LRD15], com AUC de 1.00, 0.77 e 0.95, para cada uma das doenças oncológicas referidas.

Natteshan et al. [NAAJ15] desenvolveram um sistema de apoio à decisão médica na deteção de tumores cerebrais. Neste projeto foram utilizadas 92 imagens obtidas por MRI, tendo sido detetado um tumor cerebral em 83 delas. As imagens são inicialmente pré-processadas com recurso a filtros de Wiener [Wie64], uma categoria de filtros que produz uma estimativa da imagem sem ruído através da minimização da estatística de erro ao quadrado, e uma técnica de equalização de histogramas, CLAHE [PAA+87], cujo objetivo é a melhoria do contraste da imagem. Ao nível da segmentação da imagem, foi utilizado *Fuzzy C-means* (FCM) para identificação de 4 *clusters* relativos a diferentes tipos de matéria e fluidos cerebrais. Também foi utilizado *Active Contour Model*(ACM) [KWT88], um algoritmo de desenho de contornos que utiliza um desenho manual da forma do contorno para depois adaptar à lesão, através da minimização de uma função de energia. Foram depois extraídas 35 características, envolvendo descritores de formas, GLCM, GLRLM e estatísticas globais da imagem como, por exemplo, a média de intensidades dos píxeis da região de interesse. No que diz respeito à classificação das amostras, foram utilizadas ANN's e SVM's. A melhor performance, com exatidão de 85.40%, correspondeu à classificação através de SVM com núcleo polinomial.

Wu et al. [WPG+16] desenvolveram um estudo que teve como objetivo apurar quais são os descritores de imagem que melhor classificam cancro pulmonar. Foram utilizadas imagens 3D de CT, obtidas antes do tratamento, relativas a 350 pacientes, onde a lesão foi manualmente segmentada por parte de técnicos analistas. Os descritores computados foram os seguintes: histograma de intensidades, descritores de forma, GLCM, GLRL e *wavelets*. No total, foram extraídas 440 características para cada paciente. Os atributos obtidos foram alvo de um processo de seleção através de vários algoritmos de cálculo de relevância, como ReliefF, índice Gini, ganho de informação, entre outros. Os classificadores experimentados no projeto foram *Naive Bayes*, Floresta Aleatória e kNN. Os resultados finais demonstraram que um conjunto de 5 atributos - todos eles relativos ao uso de *wavelets* - selecionado com ReliefF apresentou a melhor performance, quando o modelo foi derivado com o classificador *Naive Bayes*, com AUC de 0.72.

Wang et al. [WLD+16] fizeram um estudo acerca da capacidade de diagnóstico de certos grupos de descritores em imagens de CT de tumores pulmonares, de forma a distinguir lesões benignas e malignas. O estudo foi feito com recurso a conjuntos de imagens retiradas de uma base de dados de imagens pulmonares, LIDC-IDRI [AMB+11], relativas a 593 pacientes. Ao nível de segmentação de imagem, foi utilizado um método baseado em *tobogganing*, uma técnica que segmenta a imagem tendo por base o seu gradiente. Foram extraídas 193 características, divididas

Revisão Bibliográfica

em 4 grupos: intensidade dos píxeis, textura, forma, tamanho. Para além destas, foram também extraídas características de intensidade e texturas das decomposições da imagem através de *wavelets*. A seleção de características foi realizada com um algoritmo que utiliza floresta aleatória para atribuir pesos relacionados com a importância das características para o processo de classificação. Os modelos foram derivados através de SVM's, após um processo de normalização de dados com Min-Max (Transformação em Gama [0,1]), seguido de otimização do parâmetro C e o núcleo utilizado, recorrendo a algoritmos genéticos, onde a função objetivo definida foi a exatidão do modelo após validação cruzada. Os resultados obtidos indicaram exatidões de diagnóstico 76.1%.

Em jeito de conclusão, apresenta-se de seguida a Tabela 2.2 com uma descrição resumida das abordagens propostas pela literatura revista.

Tabela 2.2: Sumário da literatura revista

Autoria	Objetivo	Técnica	Dados	Segmentação	Descritores	Seleção	Classificadores
Farhizadeh et al. [FKS ⁺ 16]	Previsão de falhas no tratamento de NPC	MRI	25 pacientes NPC	Método de Otsu	LBP	Relieff	SVM
Kumar et al. [KBV15]	Previsão da classe BI-RADS de cancro da mama a partir da análise de mamografias	Mamografia	120 pacientes - 480 imagens	Delimitação de área centrada na lesão	wavelets - Filtrros de Haar	Combinação de diferentes tipos de atributos	SVM
Lian et al. [LRD ⁺ 16]	Sistema de previsão de tratamento a utilizar com a utilização de Evidential kNN	PET-FDG	25 pacientes cancro do pulmão 36 pacientes cancro do esófago 45 pacientes com linfoma	Limiarização ou delimitação manual	GLZSM	Evidential Feature Selection	Evidential kNN
Natteshan et al. [NAAJ15]	Sistema de apoio à decisão médica na deteção de tumores cerebrais	MRI	92 pacientes	FCM ACM	Descritores de forma GLCM GLRLM Estatísticas globais da imagem	-	SVM
Wu et al. [WPG ⁺ 16]	Otimização da escolha de descritores para a classificação de cancro do pulmão	CT	350 pacientes	Delineação manual	wavelets	Relieff	Naive Bayes
Wang et al. [WLD ⁺ 16]	Distinção de tumores pulmonares e benignos e malignos a partir de imagens médicas	CT	593 pacientes	Tobogganing	Intensidade Textura Forma Tamanho wavelets (intensidade e textura)	Atribuição de pesos de relevância de atributos através de Floresta Aleatória	SVM

Revisão Bibliográfica

Capítulo 3

Metodologia

No presente capítulo é efetuada uma descrição pormenorizada do processo experimental utilizado no projeto, desde o conjunto de dados a utilizar até às técnicas de previsão e classificação experimentadas.

3.1 Arquitetura da solução proposta

No que diz respeito à arquitetura da solução proposta, esta pode ser dividida em dois módulos: um primeiro que trata do processamento da imagem, com as fases de segmentação das lesões oncológicas e de extração de características das lesões segmentadas. Este conjunto de características, juntamente com os dados clínicos de cada paciente, constitui o conjunto de dados a ser utilizado para a derivação de modelos de previsão de classificação, com o objetivo de prever a resposta de um paciente com doença oncológica a tratamento.

O segundo módulo está relacionado com o processo de descoberta de padrões nesse conjunto de dados, com as tarefas de normalização dos dados, seleção de atributos e aplicação das técnicas de *Machine Learning*. Também foram experimentadas algumas estratégias para lidar com o caso específico da distribuição desigual de instâncias pelos diferentes valores de classe.

A figura 3.1 oferece uma descrição visual do processo acima descrito.

3.2 Conjunto de dados inicial

Os dados médicos fornecidos pelo IPO Porto contêm séries de imagens PET de 63 pacientes oncológicos, 34 relativos a tumores neuroendócrinos e 29 a linfomas de Hodgkin. Os dados incluem:

- Conjunto de imagens PET/CT de cada paciente, obtidas antes do tratamento
- Valores máximos da SUV no órgão da lesão e num outro órgão de referência

Metodologia

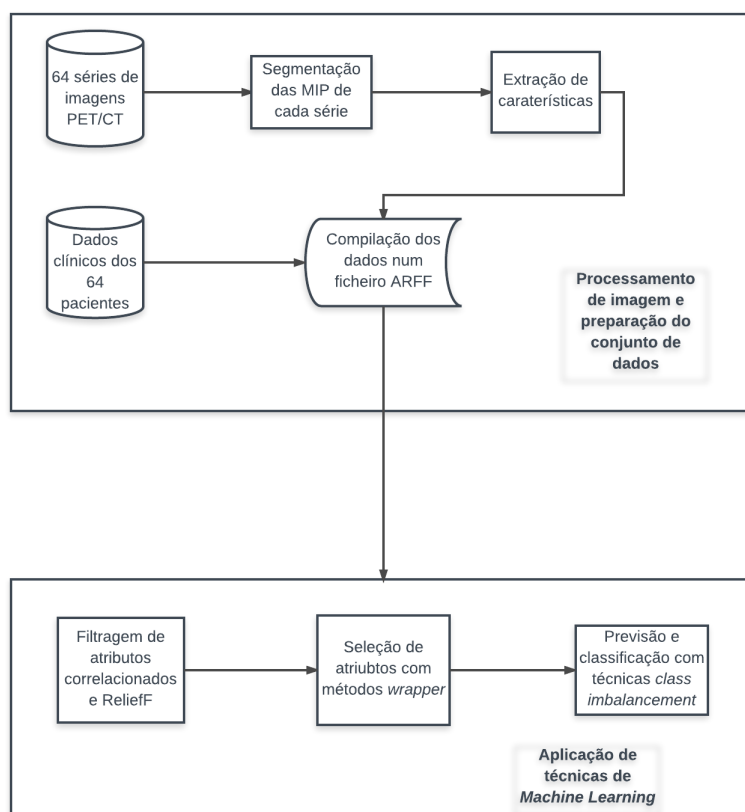


Figura 3.1: Diagrama demonstrativo da arquitetura da solução proposta

- Dados clínicos do paciente, como a idade, estágio do tumor e peso

Embora um paciente possa ter desenvolvido várias lesões relativas às doenças oncológicas em estudo, apenas foi considerada uma lesão por paciente, a principal. Esta é a lesão relacionada com os valores máximos da SUV fornecidos.

Relativamente à evolução da lesão após tratamento, cada amostra do conjunto de dados em estudo pode ser classificada como um dos seguintes tipos:

- Total (28 instâncias) - regressão total da lesão após o tratamento, tendo desaparecido por completo;
- Parcial (27 instâncias) - a gravidade da lesão diminui apesar de não ter sido removida por completo;
- Neutro (6 instâncias) - ausência de resposta a tratamento;
- Negativo (2 instâncias) - a lesão ficou mais maligna após o tratamento;

3.3 Processamento de imagem

As imagens PET fornecidas pela equipa de Medicina Nuclear do IPO-Porto consistem em secções do plano axial - plano horizontal, que divide o corpo em duas partes superior e inferior - pelo que foram utilizados os volumes gerados por essas secções para extrair secções coronais (plano que divide o corpo humano nas partes anterior e posterior).

3.3.1 Segmentação da imagem

O processo de obtenção da lesão segmentada segue três etapas baseado em [Nog15]:

1. Identificação da lesão principal - aquela que foi identificada como mais grave. Para isso, as secções foram agrupadas numa imagem de projeção de intensidades máximas (MIP). Nesta imagem foi identificada a localização anatómica correspondente à maior absorção do radiofármaco, que era aquela onde se verificava o maior nível de intensidade na imagem
2. Depois de identificada a lesão principal na MIP, seguiu-se a deteção da camada onde foi verificado o nível de absorção máximo, descoberto no passo 1
3. Procedeu-se assim ao processo de segmentação propriamente dito, onde foi adotada uma de duas estratégias:
 - Nos casos em que a intensidade máxima da secção se encontrava na região da lesão, foi aplicado um algoritmo de crescimento de regiões na imagem
 - Quando a intensidade máxima da secção se encontrava fora da região da lesão principal, foi manualmente desenhada uma área retangular a delimitar a lesão e aplicado o algoritmo de crescimento de regiões

3.3.2 Extração de características

Finalizado o processo de segmentação, passamos à extração de características da imagem, através dos descritores utilizados para extrair informação de interesse da imagem. Os descritores e respetivas características extraídas estão descritas na Tabela 4.1:

3.4 Conjunto de dados final

O conjunto de dados a ser utilizado na fase de experimentação é constituído pelas características extraídas complementadas pela informação clínica do paciente. Assim, o conjunto de dados final é composto por 63 instâncias, cada uma com 176 atributos, contendo a seguinte informação:

- 171 atributos relativos à informação extraída dos descritores computados das regiões de interesse das imagens obtidas antes do tratamento;

Tabela 3.1: Descritores e características extraídas na caracterização das regiões de interesse

Descritor	Parâmetros	Caraterísticas	Dimensão
Histograma de níveis de cinza		Média Desvio-padrão Assimetria Curtose	4
GLCM	$\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ $d=1$	Caraterísticas de Haralick	22
GLRLM	$\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$	Caraterísticas de Galloway	11
<i>Wavelets</i>	2 níveis de resolução Família Daubechies	Média Desvio-Padrão Energia Entropia	28
Filtros de Gabor	3 frequências 4 orientações	Média Desvio-Padrão Energia Entropia	48
Padrões Locais Binários	raio = 1 (8 pixéis vizinhos)	Histograma de padrões uniformes	58

- 5 atributos relativos à informação clínica do paciente, contendo a idade do paciente, o seu peso e o estágio do tumor na altura da realização da PET/CT. Para além disso, foram recolhidos os valores máximos de SUV do órgão de referência e da lesão do paciente, nas imagens obtidas antes do tratamento.

Uma rápida análise do conjunto de dados final permite antever, logo à partida, 3 potenciais problemas:

1. Elevada dimensionalidade do conjunto de dados: um grande número de atributos aumenta consideravelmente a complexidade do processo de classificação. Foram utilizados vários métodos de seleção de atributos para reduzir o número de atributos de forma a reduzir a complexidade da classificação.
2. Desequilíbrio na distribuição das instâncias pelas classes: no seguimento daquilo que foi referido na secção 2.2.12, torna-se importante ter em atenção as métricas a utilizar para a aferição da qualidade dos modelos. Desta forma, a avaliação foi feita através uma outra métrica, o *f1-score*, que combina as métricas precisão e sensibilidade. Esta estatística foi calculada para cada um dos valores de classe. Para além disso, foram utilizados alguns dos métodos referidos na secção 2.2.12.
3. Número de amostras reduzido: esta situação implica que não estará disponível uma grande quantidade de informação para a construção de modelos, especialmente no que diz respeito às classes minoritárias. Uma das alternativas poderia ser a geração sintética de amostras

utilizando SMOTE. No entanto, testes preliminares levaram à obtenção de resultados demasiado otimistas, pelo que a sua utilização foi descartada.

3.5 Método Experimental

As técnicas experimentadas no decorrer do projeto são descritas de seguida, juntamente com os parâmetros utilizados e os métodos de validação dos modelos.

Foram testados 5 tipos de classificadores distintos: k Vizinhos mais próximos, *Naive Bayes*, Floresta Aleatória, SVM e ANN. Estes algoritmos efetuam classificação de instâncias de dados a partir do treino de modelos com amostras que têm uma classe previamente atribuída para serem capazes de, no futuro, atribuir uma classificação a novas amostras obtidas ainda por classificar. No caso do projeto, a divisão das instâncias em classes é efetuada através da evolução da lesão que se verificou após tratamento. Esta evolução foi dividida em Total, onde houve eliminação da lesão, Parcial, onde a gravidade da lesão diminui parcialmente, Neutra, onde não ocorreu evolução e Negative, onde houve a gravidade da lesão aumentou.

Inicialmente, foi feita uma otimização em grelha dos parâmetros principais dos métodos de classificação e do método de normalização de atributos a utilizar - definidos em 2.2.1. As otimizações de parâmetros efetuadas estão descritas na tabela 3.2. Foi utilizado o método de otimização em grelha, com 10 passos incrementais, excetuando-se o caso do núcleo das SVM, onde foram testados os 4 núcleos referidos na tabela. Estes parâmetros foram utilizados também na etapa de seleção de atributos e na aplicação de métodos sensíveis ao custo e técnicas *ensemble*.

Tabela 3.2: Escalas de valores testados na otimização de parâmetros efetuada

Classificador	Parâmetros	Gama de valores
kNN	k	Valores inteiros entre 1 e 20
Floresta Aleatória	Número de árvores	Inteiros entre 1 e 100
	Tamanho mínimo das folhas	Escala quadrática de inteiros entre 1 e 100
	Profundidade máxima	Inteiros entre 1 e 100
SVM	C	Escala quadrática entre 0 e 100
	Núcleo	Linear, polinomial, rbf ou sigmóide

No que diz respeito aos restantes classificadores, apenas foi verificado qual o método de normalização de dados que apresentava melhor performance. Nas ANN's foi utilizada uma arquitetura com 1 camada de neurónios oculta. Não foi realizada qualquer otimização de parâmetros de ANN's por se ter verificado ser um processo bastante demorado. É de referir também que a utilização de SVM's foi adaptada à classificação de múltiplas classes seguindo uma estratégia um-contratodos. Os restantes parâmetros dos classificadores estão descritos no Anexo A, de forma a possibilitar a replicação do processo de experimentação descrito.

As otimizações e os processos de seleção de atributos empregados terão como objetivo maximizar uma média ponderada do *f1-score* de cada classe (3.1). A utilização desta métrica, em vez de uma qualquer outra global, está relacionada com o desequilíbrio na distribuição de dados

Metodologia

pelas diferentes classes. Os coeficientes da equação correspondem à razão entre o número total de instâncias do conjunto de dados e o número de instâncias de cada classe. O método de validação utilizado para o cálculo das performances foi o *bootstrapping*, com recurso a 100 iterações deste método.

$$x = \frac{2.25f1score_{total} + 2.33f1score_{parcial} + 10.5f1score_{neutro} + 31.5f1score_{negativo}}{46.58} \quad (3.1)$$

Na seleção de atributos foram implementados dois tipos de abordagens:

- PCA: a utilização deste método de análise permite efetuar uma transformação do conjunto de atributos de modo a projetar os dados segundo os eixos (componentes) de maior variância, que são mais relevantes para o processo de aprendizagem. Foi escolhido o menor conjunto de componentes com uma proporção de variância cumulativa de 95%, correspondente às 35 componentes com mais variância. Na figura 3.2 apresenta-se a proporção da variância cumulativa para as 100 componentes com maior variância, em ordem decrescente de variância.

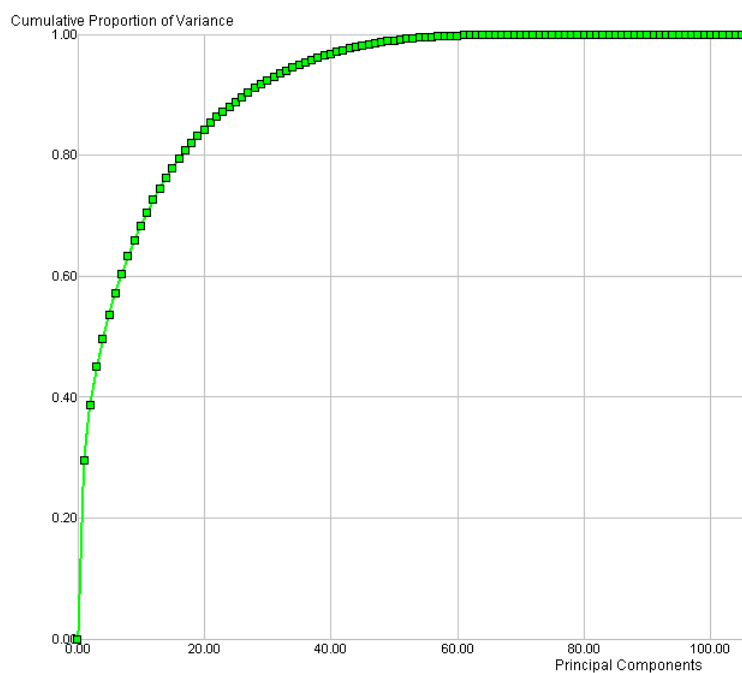


Figura 3.2: Variância cumulativa das componentes calculadas através de PCA

- Aplicação de métodos de filtro seguidos de métodos *wrapping*: foram utilizados dois métodos de filtros para reduzir o conjunto de atributos. Foi feito um cálculo da correlação absoluta entre pares de atributos, tendo sido eliminado um dos atributos dos pares onde a

Metodologia

correlação absoluta foi superior a 0.75. Após a remoção dos atributos correlacionados, foi utilizado ReliefF para atribuir pesos aos atributos de acordo com a sua relevância, tendo sido eliminados os atributos com peso inferior a 0.01. Este processo permitiu reduzir o conjunto de atributos total de 176 para 92. Em seguida foram aplicados os métodos de *wrapping*: Seleção sequencial, eliminação sequencial e algoritmos genéticos cuja função objetivo é a maximização de x na equação 3.1. Os parâmetros utilizados nos primeiros dois métodos são descritos na tabela 3.3, onde as rondas especulativas se referem ao número de vezes que a condição de paragem pode ser ignorada. A sua utilização evita a paragem do algoritmo em certos máximos locais.

Tabela 3.3: Parâmetros dos métodos *wrapper* de seleção de atributos

Método	Nº atributos máximo	Rondas especulativas	Condição de paragem
Seleção Sequencial	20	10	Sem aumento de performance
Eliminação Sequencial	60 eliminações		Decréscimo da performance

Também foi testada a eficácia de algumas das técnicas referidas pela literatura estudada para lidar com a distribuição assimétrica do conjunto de dados. Desta forma, *MetaCost*, *Bagging*, *AdaBoost* e *Stacking* são aplicados em conjunto com os 5 tipos de classificadores em estudo.

Na execução de *MetaCost*, foi testada uma otimização em grelha para os valores da matriz de custo. No entanto, os valores que otimizavam a performance do algoritmo correspondiam ao passo inicial da otimização. Desta forma, foi desenhada uma matriz de custo (tabela 3.4), tendo em conta a razão entre o número de instâncias da classe prevista e da classe real da instância. Nos casos em que esta razão é inferior a 1, foi atribuído um custo de valor 1. Nos casos em que classe prevista coincide com a real (classificações corretas), foi atribuído o custo de valor 0.

Tabela 3.4: Matriz de custo utilizada no algoritmo *MetaCost*

	real Total	real Parcial	real Neutro	real Negativo
previsto Total	0	1	4.7	14
previsto Parcial	1	0	4.5	13.5
previsto Neutro	1	1	0	3
previsto Negativo	1	1	1	0

Foram feitas experimentações com *Stacking*. Este método tem uma arquitetura um pouco diferente dos restantes, utilizando classificadores de base e um classificador agregador das previsões efetuadas pelos classificadores de base. Foram utilizados os 5 classificadores como base, tendo sido cada classificador 1 vez selecionado como agregador. ANN e SVM não foram testados como

Metodologia

classificadores agregadores pelo facto de as previsões fornecidas pelos classificadores de base serem atributos nominais. Foi efetuada novamente seleção de atributos para *Stacking*, pelo facto de ter uma arquitetura diferente dos restantes classificadores.

Finalmente, procedeu-se à utilização de ILP para construção de um conjunto de hipóteses, baseadas em regras. Foram utilizadas duas classes de exemplos, positiva e negativa, tendo sido os exemplos das diferentes classes divididos conforme explicado na tabela 3.5. Os atributos utilizados foram aqueles que apresentaram um peso superior a 0.25, após a remoção de atributos correlacionados e o cálculo da relevância com ReliefF. Este conjunto de atributos é descrito na tabela 3.6. A validação das hipóteses geradas foi feita com recurso a 3 iterações de validação cruzada *holdout*.

Tabela 3.5: Classes dos exemplos utilizados na construção de hipóteses com ILP

Exemplos Positivos	Exemplos Negativos
Total e Parcial	Negativa e Neutra
Total	Parcial, Neutra e Negativa
Total	Negativa e Netura
Total	Parcial

Tabela 3.6: Conjunto de atributos utilizado na construção de hipóteses com ILP

Atributo	Descritor
WV27	<i>Wavelets</i>
suvorgan	SUV do Órgão de Referência
suvlesion	SUV da Lesão
GF2	Filtros de Gabor
GF5	
GF10	
GF24	
GF38	
GF44	
GF46	
GLCM18	<i>Gray level co-occurrence matrix</i>
GLCM19	

Resumidamente, a experimentação efetuada foi a seguinte:

1. Derviação do modelo a partir do conjunto de dados original (com otimização de parâmetros e método de normalização, que são usados em todas as experimentações)
2. Aplicação de PCA no conjunto original de dados
3. Seleção de atributos (filtragem e *wrapping*)
4. Utilização de MetaCost, *Bagging*, AdaBoost e *Stacking* com o conjunto de dados original

Metodologia

5. Utilização de MetaCost, *Bagging* e AdaBoost com os conjuntos de atributos obtidos no processo de seleção de atributos
6. Seleção de atributos (filtragem e *wrapping*) para os modelos de *Stacking*.
7. Utilização de ILP para construção de hipóteses, para as 4 situações descritas na tabela 3.5

Metodologia

Capítulo 4

Resultados e discussão

Neste capítulo são apresentados os resultados obtidos através do processo experimental descrito no capítulo 3, juntamente com a sua análise crítica.

4.1 Otimização de parâmetros e classificação do conjunto de dados original

Como descrito na metodologia, inicialmente procedeu-se à otimização dos parâmetros e do método de normalização de dados dos classificadores. Os parâmetros aqui obtidos foram utilizados durante as experimentações seguintes.

Tabela 4.1: Resultados obtidos no processo de otimização

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{negativo}$	$f1score_{médio}$
kNN	0.5930	0.5451	0.0996	0	0.3094
<i>Naive Bayes</i>	0.5990	0.6740	0.1359	0	0.3522
Floresta Aleatória	0.5316	0.5057	0.1316	0.0158	0.2962
SVM	0.5940	0.5691	0.1870	0	0.3375
ANN	0.6503	0.5670	0.1094	0	0.3317

Tabela 4.2: Parâmetros obtidos no processo de otimização

Classificador	Parâmetros	Normalização
kNN	k = 3	Gama
<i>Naive Bayes</i>	-	Gama
Floresta Aleatória	Nº árvores = 2 Profundidade máxima = 100 Tamanho mínimo da folha = 2	Proporção
SVM	C = 24.0 núcleo = sigmóide	Transformação-Z
ANN	-	Gama Interquartil

Resultados e discussão

A análise dos resultados iniciais permite perceber que não existe um classificador que, nesta fase, se destacasse dos demais a nível de performance. Para além disso, e como já era esperado, existe uma degradação da performance no que diz respeito à classificação de instâncias das classes minoritárias, Neutra e Negativa.

Quanto ao processo de otimização em si, há a destacar o facto de o resultado da otimização do processo de normalização de dados seja diferente entre classificadores.

4.2 Seleção de atributos

Nesta secção são apresentados e discutidos os resultados relativos aos processos de seleção de atributos utilizados: PCA e a combinação entre os métodos de filtros e *wrapping*.

4.2.1 Análise da componente principal

Como já foi discutido no capítulo 3, a utilização de PCA permitiu reduzir o conjunto de atributos a 35 componentes, que acumulam 95% da variância do conjunto de dados original.

Tabela 4.3: Resultados após aplicação de PCA ao conjunto de atributos

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{negativo}$	$f1score$ médio
kNN	0.5885	0.5485	0.1086	0	0.3114
<i>Naive Bayes</i>	0.6360	0.5114	0	0	0.2869
Floresta Aleatória	0.4478	0.4415	0.0749	0.0458	0.2525
SVM	0.6570	0.6166	0.0630	0	0.3342
ANN	0.5763	0.5563	0.1168	0.0243	0.3185

Apesar da redução da dimensionalidade do problema e das vantagens que isso possa trazer a nível computacional, é possível constatar que a utilização de PCA não ofereceu grande vantagem a nível de performance por comparação com os valores obtidos para o conjunto de atributos original.

4.2.2 Filtragem e *wrapping*

Nesta secção apresentam-se os resultados relativos ao processo de seleção combinado, entre técnicas de filtragem de atributos, remoção de atributos correlacionados e ReliefF, e métodos de seleção baseados na performance obtida pelo classificador a que estão associados.

Tabela 4.4: Resultados obtidos após seleção sequencial

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Nº atributos
kNN	0.6872	0.5943	0.5628	0	0.4611	12
<i>Naive Bayes</i>	0.3745	0.7135	0.4143	0.7705	0.5682	11
Floresta Aleatória	0.8088	0.6824	0.0100	0	0.3731	9
SVM	0.7075	0.6644	0.0100	0	0.3432	15
ANN	0.8101	0.6742	0.6221	0	0.5266	4

Resultados e discussão

A visualização dos resultados de seleção sequencial permite-nos concluir que a performance global melhorou visivelmente em 4 classificadores, à exceção de SVM, onde se verificou uma ligeira melhoria. No entanto, de todos estes resultados destaca-se o de *Naive Bayes*, por ser aquele onde as performances melhoraram consideravelmente para as duas classes minoritárias. A melhoria da performance dos classificadores kNN e ANN na classificação da classe Neutra é também de apontar. Relativamente a SVM e Floresta Aleatória, fica a ideia de enviesamento da classificação das duas classes maioritárias.

Tabela 4.5: Resultados obtidos após eliminação sequencial

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Nº atributos
kNN	0.7236	0.6999	0.3230	0	0.4366	41
<i>Naive Bayes</i>	0.6936	0.7323	0.4444	0	0.4676	22
Floresta Aleatória	0.6967	0.6558	0	0	0.3381	68
SVM	0.7227	0.6764	0.1233	0	0.3806	64
ANN	0.6598	0.7201	0.3921	0	0.4430	51

Os resultados de eliminação sequencial não foram tão positivos a nível global como os de seleção sequencial. Houve um decréscimo na performance em quase todos os classificadores, excetuando-se SVM. No entanto, continuou a verificar-se o enviesamento para a classificação das duas classes maioritárias, pelo que a melhoria é sobretudo global e não ao nível da performance de classificação das classes minoritárias, como seria desejável. Não deixa de ser interessante o facto de *Naive Bayes* ter voltado a ser o classificador com melhor performance, tendo em conta não ser tão complexo como outros, como por exemplo ANN. *Naive Bayes* assume independência entre atributos na deriva de modelos [Lew98] e talvez o facto do filtro de remoção de atributos correlacionados ser utilizado tenha alguma influência, ainda que continuem a existir atributos correlacionados entre si. Para além disso, *Naive Bayes* destaca-se de outros atributos na classificação de conjuntos de dados de pequena dimensão [FC04].

Tabela 4.6: Resultados obtidos após aplicação de algoritmo genético

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Nº atributos
kNN	0.7180	0.7488	0.3200	0	0.4467	33
<i>Naive Bayes</i>	0.6203	0.7190	0.4572	0	0.4491	50
Floresta Aleatória	0.5435	0.5253	0.2083	0.3333	0.4026	52
SVM	0.7044	0.6512	0.2241	0	0.3949	43
ANN	0.6470	0.6535	0.2857	0	0.3966	39

Relativamente aos resultados da execução de algoritmos genéticos, há a destacar a melhoria de performance de Floresta Aleatória ao nível da classificação das classes minoritárias. Já ANN e *Naive Bayes* tiveram uma pior performance global em comparação com o que aconteceu nas duas técnicas anteriores.

No que diz respeito aos atributos seleccionados, apresenta-se a tabela 4.7 relativa aos atributos que mais vezes foram seleccionados nas execuções acima descritas.

Tabela 4.7: Lista de descritores mais selecionados nos métodos de seleção de atributos

Atributo	Descritor	Nº vezes
suvorganb	SUV máxima num órgão de referência	11
GF24	Filtro de Gabor	9
GF7	Filtro de Gabor	9
LBP2	Padrão Local Binário	9
suvlesionb	SUV máxima na lesão	9
WVb12	<i>Wavelets</i>	9

Esta tabela realça a conclusão retirada no estado de arte, relativa melhoria na performance com utilização de dados clínicos em conjunto com as características retiradas imagens [NAM⁺16]. As estatísticas da SUV aparecem entre os atributos mais selecionados.

4.3 Métodos *ensemble* e métodos sensíveis ao custo

Na presente secção, discutem-se os resultados relativos às estratégias escolhidas para lidar com o problema da distribuição assimétrica dos dados. Os resultados são comparados com aqueles obtidos com o conjunto de dados original e com os resultados obtidos após seleção de atributos.

As expressões presentes da coluna "Método Seleção" na apresentação dos resultados referem-se ao método de seleção cujo conjunto de atributos teve melhor performance na aplicação dos algoritmos em questão (For - seleção sequencial, Eli - eliminação sequencial e Gen - algoritmos genéticos).

4.3.1 AdaBoost

AdaBoost tem por objetivo favorecer a correta classificação de instâncias incorretamente classificadas em interações anteriores do algoritmo. É com base neste conceito que se espera que este algoritmo possa favorecer a classificação das instâncias de classes minoritárias, tendencialmente mais sujeitas a erros de classificação.

Tabela 4.8: Resultados obtidos com a aplicação do algoritmo AdaBoost sobre o conjunto de dados original

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{negativo}$	$f1score_{médio}$
kNN	0.5930	0.5451	0.0996	0	0.3094
<i>Naive Bayes</i>	0.5919	0.6439	0.1120	0	0.3367
Floresta Aleatória	0.5735	0.5648	0.0636	0	0.3005
SVM	0.6657	0.6262	0.0340	0	0.3315
ANN	0.6162	0.5980	0.1819	0	0.3490

Como se pode verificar, a execução do algoritmo AdaBoost não produziu os efeitos desejados ao nível da classificação das classes minoritárias. Na verdade, os resultados são bastante parecidos com aqueles obtidos para o conjunto de dados original.

Resultados e discussão

Tabela 4.9: Resultados da aplicação de AdaBoost nos conjuntos de atributos obtidos no processo de seleção

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Método Seleção
kNN	0.6872	0.5943	0.5627	0	0.4611	For
<i>Naive Bayes</i>	0.4043	0.6426	0.3463	0.7603	0.5384	For
Floresta Aleatória	0.6853	0.6318	0.2128	0	0.3825	For
SVM	0.7044	0.6512	0.2241	0	0.3949	Gen
ANN	0.7872	0.6796	0.6250	0	0.5230	For

Em linha com o que aconteceu na tabela de resultados 4.8, AdaBoost não produziu qualquer melhoria considerável no processo de classificação, especialmente no que diz respeito às classes minoritárias.

4.3.2 Bagging

A técnica de *Bagging* pretende melhorar a performance de classificação através da deriva de uma série de modelos a partir de dados obtidos por *bootstrapping* do modelo de treino original para aumentar a estabilidade do modelo final.

Tabela 4.10: Resultados obtidos através da utilização de *Bagging* com o conjunto original de dados

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{negativo}$	$f1score$ médio
kNN	0.5865	0.5430	0.0952	0	0.3062
<i>Naive Bayes</i>	0.6000	0.6736	0.1325	0	0.3515
Floresta Aleatória	0.6393	0.6265	0.0769	0	0.3357
SVM	0.6705	0.6056	0	0.1085	0.3462
ANN	0.6263	0.6067	0.1250	0	0.3395

Tal como aconteceu com AdaBoost, a performance de classificação não é alterada significativamente em comparação com o conjunto de dados original. Ainda assim, existiu uma melhoria ligeira da performance de classificação da classe Negativa, por parte do classificador SVM.

Tabela 4.11: Resultados obtidos da aplicação de *bagging* nos conjuntos de atributos obtidos no processo de seleção

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Método Seleção
kNN	0.6849	0.5945	0.5459	0	0.4563	For
<i>Naive Bayes</i>	0.3845	0.6711	0.3341	0.7705	0.5401	For
Floresta Aleatória	0.7690	0.6738	0.1532	0	0.3990	For
SVM	0.7025	0.6352	0.0193	0.2125	0.3924	Gen
ANN	0.7917	0.6863	0.6667	0	0.5362	For

Resultados e discussão

Tal como aconteceu na classificação relativa ao conjunto de dados original, a aplicação de *Bagging* produz um efeito de aumento da performance de classificação da classe Negativa, em comparação com os resultados obtidos no final da seleção de atributos com algoritmos genéticos, para o classificador SVM. Este poderia ser um ponto a explorar na existência de um maior número de amostras.

4.3.3 MetaCost

MetaCost é um algoritmo sensível ao custo, que tem por objetivo penalizar classificações erradas com o custo definido. A atribuição de maiores pesos à classificação errada de instâncias de classes minoritárias pode ser uma maneira de equilibrar o enviesamento existente nos modelos derivados.

Tabela 4.12: Resultados obtidos pela aplicação de MetaCost no conjunto de atributos original

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{negativo}$	$f1score$ médio
kNN	0.5097	0.4756	0.1450	0	0.2826
<i>Naive Bayes</i>	0.5366	0.6483	0.1680	0	0.4510
Floresta Aleatória	0.3925	0.3889	0.1143	0.0133	0.2273
SVM	0.4591	0	0	0.0800	0.1348
ANN	0.6236	0.5740	0.2727	0	0.3676

De uma forma geral, os resultados foram piores do que aqueles verificados para o conjunto de dados original. A degradação estará provavelmente relacionada com a tendência introduzida pelo algoritmo MetaCost para classificar mais dados como pertencentes às classes minoritárias, na maior parte das vezes erradamente. Verificou-se uma melhoria razoável de performance *Naive Bayes*. No entanto, essa melhoria não aconteceu de forma desejada, isto é, através do acerto de classificação de instâncias das classes minoritárias.

Tabela 4.13: Resultados da aplicação de MetaCost nos conjuntos de atributos obtidos no processo de seleção

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Método Seleção
kNN	0.6478	0.4910	0.5121	0	0.4127	For
<i>Naive Bayes</i>	0.3503	0.6120	0.2609	0.7705	0.4984	For
Floresta Aleatória	0.6413	0.5168	0.2600	0.0178	0.3590	For
SVM	0.5085	0	0	0.0861	0.1487	Gen
ANN	0.7273	0.3334	0.3200	0.0571	0.3595	For

Como tinha acontecido no caso do conjunto de dados original, a performance foi pior por comparação com a classificação com a seleção de atributos.

4.3.4 Stacking

Stacking é um método *ensemble* que utiliza classificadores de base cujas previsões efetuadas são fornecidas, como conjunto de dados de treino, a um classificador agregador, sendo esse modelo treinado com essas previsões em vez do conjunto de atributos fornecido aos classificadores de base.

Tabela 4.14: Resultados obtidos com a aplicação de *Stacking* no conjunto de dados original

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio
kNN	0.6774	0.6469	0.1250	0	0.3623
<i>Naive Bayes</i>	0.5893	0.6919	0.4761	0	0.4393
Floresta Aleatória	0.5659	0.5683	0.1111	0	0.3133

Naive Bayes, como classificador agregador, apresenta uma performance melhor do que nas restantes experiências com o conjunto de dados original, à exceção de MetaCost. Para os outros dois classificadores, a sua performance como agregador foi, no geral, similar aquela apresentada nas outras experiências com o conjunto de dados original.

Tabela 4.15: Resultados obtidos após seleção sequencial de atributos numa classificação com *Stacking*

Classificador	$f1score_{total}$	$f1score_{parcial}$	$f1score_{neutro}$	$f1score_{neg}$	$f1score$ médio	Nº atributos
kNN	0.5740	0.5727	0	0.6537	0.4501	14
<i>Naive Bayes</i>	0.3766	0.5625	0.2500	1	0.5473	10
Floresta Aleatória	0.6580	0.6061	0.3333	0.1819	0.4448	14

Relativamente aos resultados obtidos através de seleção de atributos, verificamos que, de uma maneira geral, estão ao nível daqueles obtidos com o processo de seleção de atributos para os outros classificadores. A grande diferença observou-se nos resultados da classe Negativa, onde kNN e *Naive Bayes* obtiveram muito bons resultados, sendo que podemos estar na presença de *overfitting*. A obtenção de mais dados para derivação de modelos pode ajudar a perceber se foi esse o caso. No entanto, as performances de classificação das restantes classes não obtiveram esse nível.

4.4 Programação Lógica Indutiva

A Programação Lógica Indutiva permite a pesquisa e seleção de hipóteses capazes de explicar os exemplos positivos e negativos segundo uma base de conhecimento. Apresentamos as regras obtidas nas experimentações realizadas, bem como os resultados relativos a exatidão, *f1-score* e AUC para cada uma das 3 validações efetuadas.

4.4.1 Total e Parcial vs Neutra e Negativa

As regras obtidas foram as seguintes, para cada um dos conjuntos de validação.

Validação 1

1. SUV órgão de referência ≤ 31.31 , *Gray level co-occurrence matrix* 18 ≤ 1.358
2. *Gray level co-occurrence matrix* 19 ≤ -0.272 , *Gray level co-occurrence matrix* 18 ≤ 1.358 ,
Gray level co-occurrence matrix 18 ≥ 0.835

Validação 2

1. SUV Lesão ≤ 31.41 , *Gray level co-occurrence matrix* 18 ≤ 1.358 ,
2. SUV Lesão ≤ 31.41 , Filtros de Gabor 2 ≤ 0.39 ,
3. Filtros de Gabor 44 ≤ 0.127 , *Gray level co-occurrence matrix* 18 ≤ 1.358 , *Gray level co-occurrence matrix* 18 ≥ 0.867

Validação 3

1. SUV Lesão ≤ 31.41 , *Gray level co-occurrence matrix* 18 ≤ 1.453 ,
2. *Gray level co-occurrence matrix* 19 ≤ -0.401 , Filtros de Gabor 24 ≤ 0.158
3. Filtros de Gabor 46 ≤ 0.144 , *Gray level co-occurrence matrix* 18 ≤ 1.453 , *Gray level co-occurrence matrix* 18 ≥ 0.867

Tabela 4.16: Resultados obtidos com ILP, onde as classes Total e Parcial constituíram o conjunto de exemplos positivos e as restantes classes o conjunto de exemplos negativos

	Exatidão	<i>FI-score</i>	AUC
Validação 1	0.8182	0.9000	0.4737
Validação 2	0.9091	0.95	0.6667
Validação 3	0.8947	0.9444	0.5
Média	0.8740	0.9315	0.5468

Nos resultados obtidos pode-se verificar uma diferença acentuada entre os valores de AUC e exatidão e *fi-score*. Tendo em conta que o cálculo destas últimas métricas não diferencia instâncias positivas de instâncias negativas, - enquanto na AUC isso acontece - os seus valores elevados podem dever-se apenas a um elevado número de previsões corretas de exemplos positivos. Este problema pode estar a ser causado pela diferença no número de exemplos positivos - 55 exemplos - e negativos - 8 exemplos.

4.4.2 Total vs Restantes

As regras obtidas foram as seguintes, para cada um dos conjuntos de validação.

Validação 1

1. SUV órgão de referência ≤ 11.1

Validação 2

1. Filtros de Gabor 44 ≤ 0.101 , Filtros de Gabor 2 ≤ 0.254 , Filtros de Gabor 2 ≥ 0.072 ,
2. *Gray level co-occurrence matrix* 18 ≤ 0.901 , SUV órgão referência ≤ 4.83 , SUV órgão referência ≥ 2.51
3. SUV órgão referência ≤ 4.83 , Filtros de Gabor 44 ≤ 0.121
4. *Gray level co-occurrence matrix* 18 ≤ 0.989 , SUV lesão ≥ 14.95
5. *Gray level co-occurrence matrix* 18 ≥ 0.901 , Filtros de Gabor 38 ≤ 0.261 , GFB38 ≥ 0.103

Validação 3

1. SUV Órgão de Referência ≤ 3.69 , Filtros de Gabor 38 ≤ 0.261 , Filtros de Gabor 38 ≥ 0.064
2. SUV Órgão de Referência ≤ 4.83 , Filtros de Gabor 10 ≥ 0.062
3. SUV Órgão de Referência ≤ 2.72 , Filtros de Gabor 38 ≤ 0.186
4. SUV Órgão de Referência ≤ 23.38 , *Wavelets* 27 ≤ 93.014 , *Wavelets* 27 ≥ 88.547
5. SUV Órgão de Referência ≤ 3.95 , SUV lesão ≥ 11.06

Tabela 4.17: Resultados obtidos com ILP, onde a classe Total constituiu o conjunto de exemplos positivos e as restantes classes o conjunto de exemplos negativos

	Exatidão	<i>F1-score</i>	AUC
Validação 1	0.6818	0.7200	0.7000
Validação 2	0.6364	0.6000	0.6333
Validação 3	0.7368	0.7368	0.7557
Média	0.6850	0.6856	0.6963

No que diz respeito aos resultados obtidos, não se verificam resultados globais tão promissores ao nível da exatidão e *f1-score* como na experiência anterior. No entanto, os resultados da AUC foram melhores, o que pode indicar uma melhoria na previsão de exemplos negativos. Esta melhoria faz sentido, pelo maior equilíbrio entre o número de exemplos positivos (28) e o número de exemplos negativos (35), o que pode ter levado a uma maior capacidade de prever corretamente exemplos negativos.

4.4.3 Total vs Neutra e Negativa

As regras obtidas foram as seguintes, para cada um dos conjuntos de validação.

Validação 1

1. SUV Lesão ≤ 24.6 , SUV Órgão de Referência ≤ 34.95

Validação 2

1. SUV Órgão de Referência ≤ 11.1

Validação 3

1. SUV Órgão de Referência ≤ 23.8

Tabela 4.18: Resultados obtidos com ILP, onde as classe Total e Parcial constitui o conjunto de exemplos positivos e as classes Neutra e Negativa o conjunto de exemplos negativos

	Exatidão	<i>F1-score</i>	AUC
Validação 1	1	1	1
Validação 2	0.9231	0.9474	0.9500
Validação 3	0.8000	0.8889	0.5000
Média	0.9077	0.9454	0.8167

Os resultados obtidos indicam que as regras das hipóteses construídas nas validações 1 e 2 permitem distinguir com muito boa precisão os exemplos de melhoria total do paciente e os exemplos onde não houve evolução ou a evolução foi negativa. Ainda assim, na validação 3 verificou-se que apesar das previsões terem sido, globalmente, corretas, o valor de AUC leva a crer que os resultados não foram tão promissores ao nível dos exemplos negativos.

4.4.4 Total vs Parcial

As regras obtidas foram as seguintes, para cada um dos conjuntos de validação.

Validação 1

1. SUV Órgão de Referência ≤ 4.83 , *Wavelets 27* ≤ 95.148
2. Filtros de Gabor $5 < = 0.087$, *Wavelets 27* ≤ 95.029 , *Wavelets 27* ≥ 92.454
3. SUV Órgão de Referência ≤ 4.83 , *Gray level co-occurrence matrix 18* ≤ 0.901

Validação 2

1. SUV Órgão de Referência $\leq 3,96$, Filtros de Gabor $38 \leq 0.186$, Filtros de Gabor $38 \geq 0.064$
2. Filtros de Gabor $38 \leq 0.186$, SUV Órgão de Referência ≤ 3.1 , SUV Órgão de Referência ≥ 1.63

Resultados e discussão

3. SUV Órgão de Referência ≤ 4.83 , SUV Órgão de Referência ≥ 1.63

Validação 3

1. *Gray level co-occurrence matrix* 18 ≤ 0.886 , SUV Órgão de Referência ≤ 11.1
2. *Gray level co-occurrence matrix* 18 ≤ 0.901 , SUV Órgão de Referência ≤ 11.1 , SUV Órgão de Referência ≥ 2.51
3. SUV Órgão de Referência ≤ 2.82 , Filtros de Gabor 38 ≤ 0.186 , Filtros de Gabor 38 ≥ 0.064
4. SUV Órgão de Referência ≤ 2.72 , Filtros de Gabor 44 ≤ 0.121
5. Filtros de Gabor 2 ≥ 0.157 , SUV Lesão ≤ 22.91 , SUV Lesão ≥ 10.17
6. *Gray level co-occurrence matrix* 18 ≤ 0.989 , *Gray level co-occurrence matrix* 19 ≤ -0.447

Tabela 4.19: Resultados obtidos com ILP, onde as classe Total constituiu o conjunto de exemplos positivos e a classe Parcial o conjunto de exemplos negativos

	Exatidão	<i>F1-score</i>	AUC
Validação 1	0.7368	0.7826	0.7278
Validação 2	0.7895	0.8182	0.7833
Validação 3	0.8235	0.8421	0.8333
Média	0.7833	0.8140	0.7815

Talvez devido ao equilíbrio existente entre o número de exemplos das duas classes, os resultados acabaram por ser mais consistentes entre as 3 validações. Assim sendo, podemos estar perante uma hipótese capaz de distinguir, com boa precisão, o tipo de evolução positiva que ocorreu em pacientes que apresentaram melhorias.

Resultados e discussão

Capítulo 5

Conclusões e Trabalho Futuro

Neste projeto foram exploradas abordagens para previsão eficaz da evolução da doença oncológica após a administração de um determinado tratamento, tendo como base imagens PET e informação clínica dos pacientes. A motivação para o desenvolvimento deste tipo de abordagens prende-se com a necessidade de lidar com o ruído inerente às imagens médicas e com a capacidade de poder auxiliar a tomada de decisão médica, nomeadamente no tratamento a administrar aos pacientes oncológicos.

A análise da imagem foi feita através da extração de características de descritores de imagem, computados sobre as lesões já devidamente segmentadas.

A fase de experimentação envolveu a utilização de diferentes tipos de técnicas de *Machine Learning*, destinadas às tarefas de previsão e classificação. Estas técnicas foram alvo de processos de otimização de parâmetros e foram também selecionados os atributos que apresentavam melhor performance para cada classificador. Alternativamente, devido ao problema existente com a distribuição irregular das instâncias pelas diferentes classes, foram utilizados métodos *ensemble* e um algoritmo sensível ao custo de forma a combater esse problema.

Os melhores resultados apontam para *f1-score* de 0.3745, 0.7135, 0.4143 e 0.7705 para cada um dos resultados previstos de tratamento. No que diz respeito à utilização de ILP para a obtenção de regras de associação, os resultados foram bastantes promissores, particularmente no que diz respeito à distinção dos exemplos de melhoria Total e Parcial, onde foi obtida uma AUC de 0.7815.

5.1 Satisfação dos Objetivos

O trabalho desenvolvido não permitiu a obtenção de resultados positivos na generalidade das experimentações efetuadas. O desequilíbrio entre o número de instâncias das classes, combinado com o número reduzido de instâncias do conjunto de dados, podem ter impedido a obtenção de bons resultados, especialmente no que concerne às classes minoritárias. Ainda assim, foram obtidos resultados bastante promissores no que diz respeito à utilização de um sistema ILP para a obtenção de regras de associação para classificação dos exemplos.

Conclusões e Trabalho Futuro

Destaca-se também o facto de se ter desenvolvido uma abordagem que pode constituir um ponto de partida para a análise e classificação de imagens PET e dados clínicos reais.

5.2 Trabalho Futuro

O seguimento da linha de desenvolvimento deste projeto teria sempre de passar pela obtenção de informação relativa a novos pacientes, para melhorar os resultados obtidos e aferir os efeitos dos problemas relacionados com o conjunto de dados na performance das tarefas de previsão e classificação. A obtenção de mais amostras pode também permitir validar os resultados positivos obtidos com ILP e testar a construção de hipóteses para todas as restantes classes, numa estratégia de uma classe contra as restantes.

Referências

- [AMB⁺11] Samuel G. Armato, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. van Beek, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P.-Y. Qing, Rachael Y. Roberts, Amanda R. Smith, Adam Starkey, Poonam Batra, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Castele, Sangeeta Gupte, Maha Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, Barbara Y. Croft e Laurence P. Clarke. The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38(2):915–931, 2011. URL: <http://dx.doi.org/10.1118/1.3528204>, doi:10.1118/1.3528204.
- [AR05] Tinku Acharya e Ajoy K. Ray. *Image Processing - Principles and Applications*. Wiley-Interscience, 2005.
- [Bre96] Leo Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, August 1996. URL: <http://dx.doi.org/10.1023/A:1018054314350>, doi:10.1023/A:1018054314350.
- [CBHK02] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall e W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, June 2002. URL: <http://dl.acm.org/citation.cfm?id=1622407.1622416>.
- [Dem67] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38(2):325–339, 04 1967. URL: <http://dx.doi.org/10.1214/aoms/1177698950>, doi:10.1214/aoms/1177698950.
- [Dom99] Pedro Domingos. Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '99, pages 155–164, New York, NY, USA, 1999. ACM. URL: <http://doi.acm.org/10.1145/312129.312220>, doi:10.1145/312129.312220.

REFERÊNCIAS

- [FBM⁺14] Brent Foster, Ulas Bagci, Awais Mansoor, Ziyue Xu e Daniel J. Mollura. A review on segmentation of positron emission tomography images. *Computers in Biology and Medicine*, 50:76 – 96, 2014. URL: [//www.sciencedirect.com/science/article/pii/S0010482514001000](http://www.sciencedirect.com/science/article/pii/S0010482514001000), doi:<http://dx.doi.org/10.1016/j.compbiomed.2014.04.014>.
- [FC04] George Forman e Ira Cohen. *Learning from Little: Comparison of Classifiers Given Little Training*, pages 161–172. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. URL: http://dx.doi.org/10.1007/978-3-540-30116-5_17, doi:10.1007/978-3-540-30116-5_17.
- [FCPB07] Alberto Freitas, Altamiro Costa-Pereira e Pavel Brazdil. *Cost-Sensitive Decision Trees Applied to Medical Data*, pages 303–312. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. URL: http://dx.doi.org/10.1007/978-3-540-74553-2_28.
- [FKS⁺16] Hamidreza Farhidzadeh, Joo Y. Kim, Jacob G. Scott, Dmitry B. Goldgof, Lawrence O. Hall e Louis B. Harrison. Classification of progression free survival with nasopharyngeal carcinoma tumors, 2016. URL: <http://dx.doi.org/10.1117/12.2216976>, doi:10.1117/12.2216976.
- [FS97] Yoav Freund e Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139, 1997. URL: <http://www.sciencedirect.com/science/article/pii/S002200009791504X>, doi:<http://dx.doi.org/10.1006/jcss.1997.1504>.
- [Gal75] Mary M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4(2):172 – 179, 1975. URL: <http://www.sciencedirect.com/science/article/pii/S0146664X75800086>, doi:[http://dx.doi.org/10.1016/S0146-664X\(75\)80008-6](http://dx.doi.org/10.1016/S0146-664X(75)80008-6).
- [GFB⁺12] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince e F. Herrera. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4):463–484, July 2012. doi:10.1109/TSMCC.2011.2161285.
- [GLHAHAA15] P. J. García-Laencina, P. Henriques Abreu, M. Henriques Abreu e N. Afonso. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Computers in Biology and Medicine*, 59:125–133, 2015.
- [GW06] Rafael C. Gonzalez e Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [HAACS⁺14a] P. Henriques Abreu, H. Amaro, D. Castro Silva, P. Machado e M. Henriques Abreu. Personalizing breast cancer patients with heterogeneous data. In *Proceedings of the International Conference on Health Informatics*, pages 39–42. Springer International Publishing, 2014.

REFERÊNCIAS

- [HAACS⁺14b] P. Henriques Abreu, H. Amaro, D. Castro Silva, P. Machado, M. Henriques Abreu, N. Afonso e A. Dourado. Overall survival prediction for women breast cancer using ensemble methods and incomplete clinical data. In *Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pages 1366–1369. Springer International Publishing, 2014.
- [HAAHA⁺16] M. Henriques Abreu, N. Afonso, P. Henriques Abreu, F. Menezes, P. Lopes, R. Henrique, D. Pereira e C. Lopes. Male breast cancer: Looking for better prognostic subgroups. *The Breast*, 26:18–24, 2016.
- [HAGM⁺15] M. Henriques Abreu, M. Gomes, F. Menezes, N. Afonso, P. Henriques Abreu, R. Medeiros, D. Pereira e C. Lopes. Cyp2d6* 4 polymorphism: A new marker of response to hormonotherapy in male breast cancer? *The Breast*, 24(4):481–486, 2015.
- [HAHAA⁺16] M. Henriques Abreu, P. Henriques Abreu, N. Afonso, D. Pereira, R. Henrique e C. Lopes. Patterns of recurrence and treatment in male breast cancer: A clue to prognosis? *International Journal of Cancer*, 139(8):1715–1720, 2016.
- [HASSHA⁺16] P. Henriques Abreu, M. Seoane Santos, M. Henriques Abreu, B. Andrade e D. Castro Silva. Predicting breast cancer recurrence using machine learning techniques: A systematic review. *ACM Computing Surveys*, 49(3):1–40, 2016.
- [HBK⁺98] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang e S. Munishkumaran. *Current Status of the Digital Database for Screening Mammography*, pages 457–460. Springer Netherlands, Dordrecht, 1998. URL: http://dx.doi.org/10.1007/978-94-011-5318-8_75, doi:10.1007/978-94-011-5318-8_75.
- [HG09] Haibo He e Edwardo A. Garcia. Learning from imbalanced data. *IEEE Trans. on Knowl. and Data Eng.*, 21(9):1263–1284, September 2009. URL: <http://dx.doi.org/10.1109/TKDE.2008.239>, doi:10.1109/TKDE.2008.239.
- [HKP11] Jiawei Han, Micheline Kamber e Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition, 2011.
- [HL02] Chih-Wei Hsu e Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Trans. Neur. Netw.*, 13(2):415–425, March 2002. URL: <http://dx.doi.org/10.1109/72.991427>, doi:10.1109/72.991427.
- [Ho95] Tin Kam Ho. Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*, ICDAR '95, pages 278–, Washington, DC, USA, 1995. IEEE Computer Society. URL: <http://dl.acm.org/citation.cfm?id=844379.844681>.
- [HS85] Robert M. Haralick e Linda G. Shapiro. Image segmentation techniques, 1985. URL: <http://dx.doi.org/F10.1117/12.948400>, doi:10.1117/12.948400.
- [HSD73] R. Haralick, K. Shanmugam e I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 1973.

REFERÊNCIAS

- [KBV15] Indrajeet Kumar, H.S. Bhadauria e Jitendra Virmani. Wavelet packet texture descriptors based four-class birads breast tissue density classification. *Procedia Computer Science*, 70(Proceedings of the 4th International Conference on Eco-friendly Computing and Communication Systems):76 – 84, 2015. URL: <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso,url,uid&db=edselp&AN=S1877050915032068&lang=pt-br&site=eds-live&scope=site&authtype=sso>.
- [KF10] Paul E. Kinahan e James W. Fletcher. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Seminars in Ultrasound, {CT} and {MRI}*, 31(6):496 – 505, 2010. PET-CT in {RT} Planning and Assessment. URL: <http://www.sciencedirect.com/science/article/pii/S0887217110000880>, doi:<http://dx.doi.org/10.1053/j.sult.2010.10.001>.
- [KJ97] Ron Kohavi e George H. John. Wrappers for feature subset selection. *Artif. Intell.*, 97(1-2):273–324, December 1997. URL: [http://dx.doi.org/10.1016/S0004-3702\(97\)00043-X](http://dx.doi.org/10.1016/S0004-3702(97)00043-X), doi:10.1016/S0004-3702(97)00043-X.
- [KŠRŠ97] Igor Kononenko, Edvard Šimec e Marko Robnik-Šikonja. Overcoming the myopia of inductive learning algorithms with relief. *Applied Intelligence*, 7(1):39–55, 1997. URL: <http://dx.doi.org/10.1023/A:1008280620621>, doi:10.1023/A:1008280620621.
- [KWT88] Michael Kass, Andrew Witkin e Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988. URL: <http://dx.doi.org/10.1007/BF00133570>, doi:10.1007/BF00133570.
- [LC07] S. Liao e A. C. S. Chung. Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 1, pages I–1221–I–1224, April 2007. doi:10.1109/ICASSP.2007.366134.
- [LCC⁺08] Hung-Yi Lo, Chun-Min Chang, Tsung-Hsien Chiang, Cho-Yi Hsiao, Anta Huang, Tsung-Ting Kuo, Wei-Chi Lai, Ming-Han Yang, Jung-Jung Yeh, Chun-Chao Yen e Shou-De Lin. Learning to improve area-under-froc for imbalanced medical data classification using an ensemble method. *SIGKDD Explor. Newsl.*, 10(2):43–46, December 2008. URL: <http://doi.acm.org/10.1145/1540276.1540290>, doi:10.1145/1540276.1540290.
- [Lew98] David D. Lewis. *Naive (Bayes) at forty: The independence assumption in information retrieval*, pages 4–15. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. URL: <http://dx.doi.org/10.1007/BFb0026666>, doi:10.1007/BFb0026666.
- [LFG⁺13] Victoria López, Alberto Fernández, Salvador García, Vasile Palade e Francisco Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113 – 141, 2013. URL: <http://www.sciencedirect.com/science/article/pii/S0020025513005124>, doi:<http://dx.doi.org/10.1016/j.ins.2013.07.007>.

REFERÊNCIAS

- [LRD15] Chunfeng Lian, Su Ruan e Thierry Dencœur. An evidential classifier based on feature selection and two-step classification strategy. *Pattern Recognition*, 48(7):2318 – 2327, 2015. URL: [//www.sciencedirect.com/science/article/pii/S0031320315000382](http://www.sciencedirect.com/science/article/pii/S0031320315000382), doi:<http://dx.doi.org/10.1016/j.patcog.2015.01.019>.
- [LRD⁺16] Chunfeng Lian, Su Ruan, Thierry Dencœur, Fabrice Jardin e Pierre Vera. Selecting radiomic features from fdg-pet images for cancer treatment outcome prediction. *Medical Image Analysis*, 32:257 – 268, 2016. URL: [//www.sciencedirect.com/science/article/pii/S1361841516300366](http://www.sciencedirect.com/science/article/pii/S1361841516300366), doi:<http://dx.doi.org/10.1016/j.media.2016.05.007>.
- [MHZ⁺08] Maciej A. Mazurowski, Piotr A. Habas, Jacek M. Zurada, Joseph Y. Lo, Jay A. Baker e Georgia D. Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2–3):427 – 436, 2008. Advances in Neural Networks Research: {IJCNN} '072007 International Joint Conference on Neural Networks {IJCNN} '07. URL: <http://www.sciencedirect.com/science/article/pii/S0893608007002407>, doi:<http://dx.doi.org/10.1016/j.neunet.2007.12.031>.
- [MMAHA14] P. Machado, T. Martins, H. Amaro e P. Henriques Abreu. An interface for fitness function design. In *Proceedings of the International Conference on Evolutionary and Biologically Inspired Music and Art*, pages 13–25. Springer Berlin Heidelberg, 2014.
- [Mug91] Stephen Muggleton. Inductive logic programming. *New Generation Computing*, 8(4):295–318, 1991. URL: <http://dx.doi.org/10.1007/BF03037089>, doi:10.1007/BF03037089.
- [Mug99] Stephen Muggleton. Inductive logic programming: Issues, results and the challenge of learning language in logic. *Artificial Intelligence*, 114(1):283 – 296, 1999. URL: <http://www.sciencedirect.com/science/article/pii/S0004370299000673>, doi:[http://dx.doi.org/10.1016/S0004-3702\(99\)00067-3](http://dx.doi.org/10.1016/S0004-3702(99)00067-3).
- [NAAJ15] N. V. S. Natteshan e J. Angel Arul Jothi. *Automatic Classification of Brain MRI Images Using SVM and Neural Network Classifiers*, pages 19–30. Springer International Publishing, Cham, 2015. URL: http://dx.doi.org/10.1007/978-3-319-11218-3_3, doi:10.1007/978-3-319-11218-3_3.
- [NAM⁺16] Mariana A. Nogueira, Pedro Henriques Abreu, Pedro Martins, Penousal Machado, Hugo Duarte e João Santos. Image descriptors in radiology images: a systematic review. *Artificial Intelligence Review*, pages 1–29, 2016. URL: <http://dx.doi.org/10.1007/s10462-016-9492-8>, doi:10.1007/s10462-016-9492-8.
- [Nog15] Mariana Antunes Nogueira. Creating evaluation functions for oncological diseases based on pet/ct. Master's thesis, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 2015.

REFERÊNCIAS

- [Ots79] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan 1979. doi:10.1109/TSMC.1979.4310076.
- [PAA⁺87] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart Ter Haar Romeny e John B. Zimmerman. Adaptive histogram equalization and its variations. *Comput. Vision Graph. Image Process.*, 39(3):355–368, September 1987. URL: [http://dx.doi.org/10.1016/S0734-189X\(87\)80186-X](http://dx.doi.org/10.1016/S0734-189X(87)80186-X), doi:10.1016/S0734-189X(87)80186-X.
- [RD13] M. Mostafizur Rahman e D. N. Davis. Addressing the class imbalance problem in medical datasets. *International Journal of Machine Learning and Computing*, 3(2):224–228, 2013.
- [Sha76] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
- [SL09] Marina Sokolova e Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.*, 45(4):427–437, July 2009. URL: <http://dx.doi.org/10.1016/j.ipm.2009.03.002>, doi:10.1016/j.ipm.2009.03.002.
- [SS12] Amandeep Singh e Jaspinder Sidhu. Article: Performance analysis of segmentation techniques. *International Journal of Computer Applications*, 45(23):18–23, May 2012.
- [SSHAGL⁺15] M. Seoane Santos, P. Henriques Abreu, P. J. García-Laencina, A. Simão e A. Carvalho. A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. *Journal of biomedical informatics*, 58:49–59, 2015.
- [SSW88] P.K Sahoo, S Soltani e A.K.C Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233 – 260, 1988. URL: <http://www.sciencedirect.com/science/article/pii/S0734189X88900229>, doi:http://dx.doi.org/10.1016/0734-189X(88)90022-9.
- [SW14] Bernard Stewart e Christopher P. Wild, editors. *World Cancer Report 2014*. International Agency for Research on Cancer, Lyon, France, 2014.
- [TAM14] G. Thibault, J. Angulo e F. Meyer. Advanced statistical matrices for texture characterization: Application to cell classification. *IEEE Transactions on Biomedical Engineering*, 61(3):630–637, March 2014. doi:10.1109/TBME.2013.2284600.
- [Wie64] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
- [WLD⁺16] J. Wang, X. Liu, D. Dong, J. Song, M. Xu, Y. Zang e J. Tian. Prediction of malignant and benign of lung tumor using a quantitative radiomic method. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1272–1275, Aug 2016. doi:10.1109/EMBC.2016.7590938.

REFERÊNCIAS

- [Wol92] David H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241 – 259, 1992. URL: <http://www.sciencedirect.com/science/article/pii/S0893608005800231>, doi:[http://dx.doi.org/10.1016/S0893-6080\(05\)80023-1](http://dx.doi.org/10.1016/S0893-6080(05)80023-1).
- [WPG⁺16] Weimiao Wu, Chintan Parmar, Patrick Grossmann, John Quackenbush, Philippe Lambin, Johan Bussink, Raymond Mak e Hugo J. W. L. Aerts. Exploratory study to identify radiomics classifiers for lung cancer histology. *Frontiers in Oncology*, 6:71, 2016. URL: <http://journal.frontiersin.org/article/10.3389/fonc.2016.00071>, doi:10.3389/fonc.2016.00071.

REFERÊNCIAS

Anexo A

Parametrização dos classificadores

Neste anexo é disponibilizada a parametrização dos algoritmos de classificação utilizados com maior detalhe, de forma a permitir a replicação da experimentação efetuada no projeto. Para além dos parâmetros utilizados, os parâmetros restantes têm os valores por defeito do *software* Rapidminer. A tabela A.1 contém essa informação detalhada.

Tabela A.1: Parâmetros utilizados nos algoritmos de classificação

Classificador	Parâmetro	Valor
kNN	k	3
	weighted vote	false
	mesaure types	Mixed Measure
	mixed mesaure	Mixed Euclidean Distance
<i>Naive Bayes</i>	laplace correction	true
Floresta Aleatória	n ^o of trees	2
	criterion	information_gain
	maximal depth	100
	pruning	true
	confidence	0.25
	prepruning	true
	minimal gain	0.1
	min leaf size	2
voting strategy	confidence vote	
SVM(LibSVM)	type	C-SVC
	kernel	sigmoid
	gamma	0.0
	C	24.0
	epsilon	0.001
ANN	hidden layers	1
	training cycles	500
	momentum	0.3
	error epsilon	0.00001