



**Sérgio Miguel
Laranjeira Lima**

**Estudo de algoritmos para detectar pólipos em vídeo
de endo-cápsula**



**Sérgio Miguel
Laranjeira Lima**

**Estudo de algoritmos para detectar pólipos em vídeo
de endo-cápsula**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações (M.I.E.E.T.), realizada sob a orientação científica do Doutor João Paulo Trigueiros da Silva Cunha, Professor associado do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

Dedico este trabalho à minha família.

o júri

presidente

Prof. Doutor José Luis Oliveira

professor associado do Departamento de Electrónica e Telecomunicações e Informática da Universidade de Aveiro

vogais

Prof. Doutor Miguel Tavares Coimbra

professor auxiliar convidado do Departamento de Ciência de Computadores da Faculdade de Ciências da Universidade do Porto

Prof. Doutor João Paulo Trigueiros da Silva Cunha

professor associado do Departamento de Electrónica e Telecomunicações e Informática da Universidade de Aveiro

Agradecimentos

O meu principal agradecimento vai para o prof. Doutor João Paulo Cunha, orientador científico deste trabalho, agradeço o empenho colocado, a confiança depositada e a paciência permanente.

À instituição Hospital Geral Santo António e ao Dr. José Soares que permitiram a realização deste trabalho e agradeço a sua colaboração e disponibilidade.

Aos meus pais agradeço tudo aquilo que me têm proporcionado, tão importantes em momentos de maior stress.

Ao meu irmão, avó e tios agradeço a amizade, disponibilidade e paciência sempre presentes.

Aos meus colegas e amigos pelo apoio e incentivo constantes.

Palavras-chave

Cápsula endoscópica, Capview, processamento de imagem, visão por computador, aprendizagem automática, mineração de dados.

Resumo

A cápsula endoscópica está disponível para diagnóstico desde 2001, altura em que recebeu a aprovação da *Food Drug and Administration* (FDA) e a certificação CE. A sua utilização está a crescer criando um novo mercado para ferramentas tecnológicas de detecção automática de eventos, de forma a diminuir a necessidade de visualização do exame completo por parte dos médicos.

O primeiro objectivo a extrair deste trabalho foi o desenvolvimento de uma ferramenta de software de catalogação de imagem de cápsula endoscópica para obter um conjunto de dados bem anotado para ajudar no desenvolvimento de algoritmos computacionais de ajuda ao diagnóstico automático.

Um segundo objectivo deste trabalho foi a validação de duas metodologias de análise de textura aplicadas com o intuito de detectar eventos em vídeo endoscópico.

Os resultados obtidos demonstram a dificuldade de utilização de tais meios para a detecção dos referidos eventos em vídeo de cápsula endoscópica.

Keywords

Endoscopic capsule, Capview, image processing, machine vision, machine learning, data mining .

Abstract

The endoscopic capsule is available for medical use since 2001, when FDA and CE gave their approval as means to diagnose gastrointestinal diseases. Its exponential use is creating a new market for computational tools that can detect events automatically. Therefore aiding doctors on diagnosing exams.

The main goal of this work was the creation of a software tool for catalog images from endoscopic capsule video, obtaining a well annotated dataset for aiding the development of computer aided diagnosis algorithms.

The next objective was the validation of two methodologies of image texture analysis applied and implemented to detect abnormal events on endoscopic video.

The results had shown the difficulty of such means to detect those events using endoscopic capsule video.

Índice

1. Introdução.....	1
1.1. Motivação.....	4
1.2. Objectivos.....	4
1.3. Estrutura do documento.....	5
2. Cápsula endoscópica e software de anotação.....	7
2.1. Características técnicas.....	7
2.2. Características clínicas	9
2.3. Procedimento clínico.....	10
2.4. Software de anotação: CAPVIEW.....	10
3. Espaço de cores e algoritmos de aprendizagem automática.....	13
3.1. Espaço de Cores.....	13
3.1.1. RGB.....	13
3.1.2. Espaço de cor HSV / HSL.....	14
3.1.3. Karhunen-Loève.....	15
3.2. Algoritmos de aprendizagem automática.....	16
3.2.1. Support Vector Machines (SVM).....	16
3.2.2. Redes Neurais Artificiais (Neural Networks).....	17
3.2.3. Bayes Networks.....	19
3.2.4. Métodos de avaliação.....	20
4. Métodos de detecção automática de pólipos.....	23
4.1. Análise de textura.....	24
4.1.1. Matrizes de Co-Ocorrência.....	25
4.1.2. CWC: Color Wavelet Covariance.....	27
4.1.3. LBP: Local Binary Patterns.....	29
5. Ferramenta de anotação de eventos ECCA.....	31
5.1. Objectivo.....	32
5.2. Arquitectura.....	32
5.3. Workflow.....	33
5.4. Aplicações do software.....	34
5.4.1. Gestor conjuntos	35
5.4.2. Ferramenta pedagógica.....	35
5.5. Resultados.....	37
6. Detecção de pólipos em cápsula endoscópica.....	39
6.1. Conjuntos de Dados.....	39
6.2. Metodologia adoptada.....	40
6.2.1. Metodologia 1.....	42
6.2.2. Metodologia 2.....	43
7. Resultados e discussão.....	47
8. Conclusão.....	53
8.1. Trabalho Futuro.....	55
9. Bibliografia.....	57
10. Apêndice I.....	63
11. Apêndice II.....	65

Lista de figuras

Figura 1: Sistema gástrico (digestivo).....	2
Figura 2: Trânsito do intestino delgado, SBFT.....	3
Figura 3: Colonoscopia virtual, com recurso a CT.....	3
Figura 4: Wireless endoscopic capsule.....	8
Figura 6:Localização das antenas e dispositivo de armazenamento.....	8
Figura 7: Capview: janela principal.....	11
Figura 8: RGB colorspace.....	13
Figura 9: Espaço HSV.....	14
Figura 10:Separação máxima do hiperplano.....	16
Figura 11: Linearização das características.....	17
Figura 12: Neurónio projectado por McCulloch.....	18
Figura 13: Rede neuronal com 3 camadas intermédias.....	19
Figura 14: Modelo bayesiano.....	20
Figura 15: Diversos graus de resolução	25
Figura 16: Cálculo das matrizes de co-ocorrência para uma matriz de 4x4.....	26
Figura 17: Representação da decomposição wavelet em 3 níveis.....	28
Figura 18: Cálculo do histograma LBP para uma matriz 3x3.....	29
Figura 19: Arquitectura simplificada do software ECCA.....	32
Figura 20: Janela principal do software.....	33
Figura 21: Janela de catalogação.....	34
Figura 22: Metodologia de avaliação.....	36
Figura 23: Avaliação estratificada.....	37
Figura 24: Arquitectura de componentes do sistema de software utilizado.....	41
Figura 25: Metodologia 1.....	42
Figura 26: Algoritmo de Karkanis et al.....	43
Figura 27: Metodologia 2.....	44
Figura 28: Método usando LBP.....	44
Figura 29: Excerto de código XML - descritor.....	46
Figura 30: Script para seleccionar blocos de eventos.....	46
Figura 31: Comparação visual dos resultados aval-treino.....	49
Figura 32: Comparação visual dos resultados aval-teste.....	49

Lista de Acrónimos

CWC:	Color Wavelet Covariance
LBP:	Local Binary Pattern
RM:	Ressonância Magnética
TC:	Tomografia Computorizada
TID:	Trânsito Intestino Delgado
SBFT:	Small-Bowel Follow Through-6
FDA:	Food Drug and Administration
XML:	eXtended Markup Language
API:	Application Programming Interface
HSV:	Hue Saturation Value
HSL:	Hue Saturation Luminance
RGB:	Red Green Blue
ARFF:	Attribute Relation File Format
K-L:	Karhunen-Loève
CE:	Certificação Europeia

1. Introdução

“Nós somos o que comemos” [Hipócrates].

O sistema digestivo (Figura 1) é de extrema relevância na vida das pessoas pelo facto de ser este que através do processamento dos alimentos fornece os nutrientes e proteínas necessárias para a sobrevivência dos seres vivos. Assim, é de extrema importância conservá-lo da melhor forma minimizando os riscos de contrair algum tipo de doença endogástrica. As doenças mais graves neste sistema são as assintomáticas e prolongadas, sendo o cancro a palavra indesejada por qualquer um. A única maneira de detectar cancros e pólipos em estágios iniciais é através de exames de rotina usando ferramentas de alta fiabilidade, e neste caso essas ferramentas são as endoscopias ou seja a visualização *in loco* do interior do estômago e intestino.

A história da endoscopia teve início há bastante tempo (2500 anos atrás) quando Hipócrates examinou o recto de um paciente. A evolução tecnológica desde esse tempo permite a utilização de ferramentas mais pequenas e precisas, minimizando o desconforto dos pacientes. Em 1805 Bozzini usou uma ferramenta composta por uma vela de cera como fonte de luz e um tubo endoscópico rígido, bastante primitivo, constituído por uma série de espelhos em prata de forma a reflectir a luz e conseguir visualizar o interior do recto do paciente [1]. Desormeux é citado como o pai do termo “endoscopia”, pois foi o primeiro a utilizar o endoscópio com a intenção de diagnosticar e tratar pacientes, este endoscópio era bastante semelhante ao de Bozzini apenas com algumas melhorias a nível da fonte de luz que, neste caso, utilizava uma lâmpada de querosene que queimava álcool e uma espécie de resina (terebentina) [2].

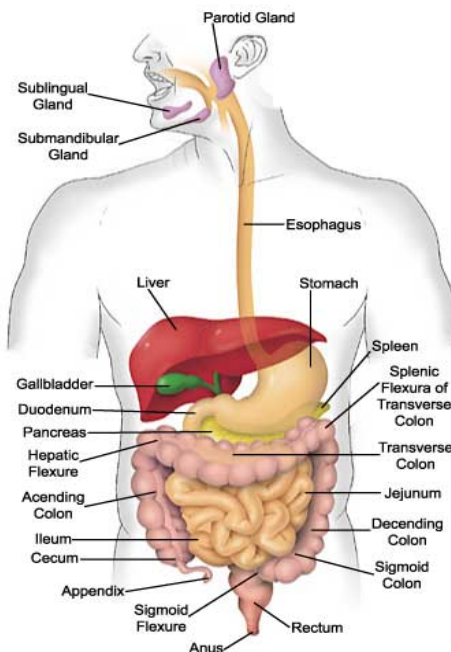


Figura 1: Sistema gástrico (digestivo), <http://www.medicalook.com>

Nas últimas décadas com o desenvolvimento da electrónica e dos sistemas computacionais apareceram muitas ferramentas e técnicas de endoscopia menos dolorosas e mais eficientes. Os tubos endoscópicos são semi-rígidos, mais finos e longos possibilitando a visualização de uma maior área endogástrica, podendo realizar biópsias e micro cirurgias no decorrer do exame. A única desvantagem destas endoscopias é o desconforto promovido ao paciente que em certos casos tem de ser anestesiado enquanto os médicos realizam os exames que podem ter uma duração longa (1 ou 2 horas). O ideal era ter uma ferramenta que conseguisse visualizar o interior dos pacientes sem lhes provocar desconforto mas que tivesse informação médica relevante para efectuar um diagnóstico, questão que levou ao desenvolvimento de métodos de visualização indirectos como sejam endoscopia efectuada com suporte radiológico, endoscopia virtual e ao desenvolvimento da cápsula endoscópica.

Alguns estudos radiológicos como o trânsito do intestino delgado – TID ou SBFT¹ – (Figura 2) e o clister opaco conseguem observar o intestino delgado e grosso mediante a utilização de fluoroscopia e recurso a meios de contraste. No entanto, estes exames envolvem a utilização de radiação ionizante (raios-x) e são em geral exames longos e desconfortáveis para o paciente [3].

1 SBFT: Small Bowel Follow Through



Figura 2: Trânsito do intestino delgado, SBFT, <http://uwmedicine.washington.edu>

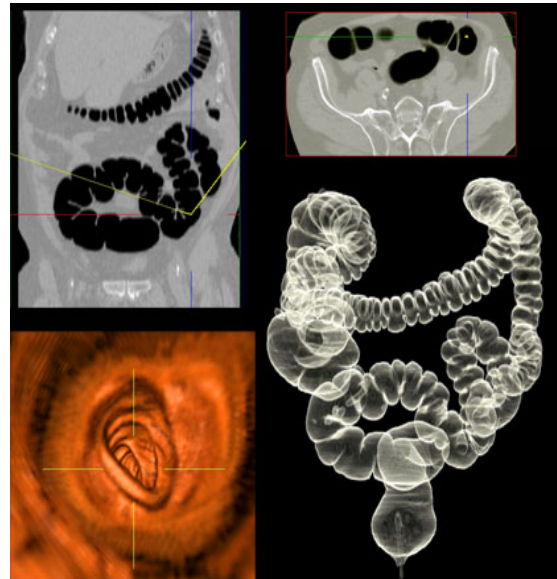


Figura 3: Colonoscopia virtual, com recurso a CT, <http://www.sdirad.com>

A endoscopia virtual (Figura 3) utiliza dados de exames de tomografias computadorizadas (TC) e ressonância magnética (RM) e com recurso a técnicas de processamento de sinal e modelação virtual cria um modelo 3D do interior do intestino e dessa forma o médico consegue ter uma percepção do interior do intestino sem a necessidade de realizar uma endoscopia comum [4,5].

A utilização da cápsula endoscópica, começou com um estudo de 1994 publicado por um grupo de investigação de Londres que comprovou o conceito de uma cápsula endoscópica mostrando nessa altura imagens do interior do estômago de um porco [6]. Este grupo juntou-se a um outro Israelita que foi o primeiro a construir uma cápsula com as dimensões e características necessárias para poder ser usada como ferramenta de diagnóstico em seres humanos. Os primeiros estudos com humanos foram iniciados em 1999 após parecer positivo do comité de ética do *Royal London Hospital*, localizado no Reino Unido [6]. Em Agosto de 2001 recebeu a certificação CE e aprovação da FDA para utilização em pacientes humanos [6]. Desde então a cápsula tem sido utilizada em diversos estudos clínicos e têm sido efectuadas comparações com outras técnicas de endoscopia normais [7-13], demonstrando que este método é o recomendado para a análise do intestino delgado. Esta preferência deve-se ao facto de a endoscopia convencional não conseguir visualizar a totalidade do intestino delgado (média 6.7 metros) ficando-se apenas pelos 37% da totalidade do comprimento [9], enquanto que a cápsula consegue fluir na totalidade do intestino com uma boa qualidade de vídeo, e mais relevante é o facto de ser um exame não invasivo e pouco incomodo para o paciente [14,15].

1.1. Motivação

Este tipo de exames está, hoje em dia, em franca expansão apresentando diversas aplicações práticas descritas nos diversos estudos realizados, sendo a grande desvantagem deste sistema a quantidade de tempo despendido pelos médicos na análise e diagnóstico dos exames/vídeos de cápsula endoscópica.

Assim e anteendo a elevada utilização e massificação deste tipo de exames (entrada no mercado de mais uma companhia japonesa RF-System², para além da GivenImaging e Olympus) é crucial conseguir diminuir o tempo de diagnóstico, campo onde as ferramentas informáticas mais precisamente as técnicas de *computer vision* e *machine learning* têm um grande potencial, tornando possível a abordagem à detecção automática de pólipos.

O facto de os pólipos serem descritos em termos de textura e morfologia torna a análise e computação mais complexa, sendo necessário métodos e algoritmos que consigam simplificar e extrair as características únicas e necessárias de cada imagem para um processamento eficaz. O trabalho descrito por Karkanis et al. é bastante promissor, utilizando o espaço de *wavelets* para extrair informação acerca da textura de imagens a partir de vídeos com origem numa colonoscopia[16].

1.2. Objectivos

O primeiro objectivo desta dissertação foi o desenvolvimento duma ferramenta de catalogação de imagem de cápsula endoscópica, para obter um conjunto de dados bem anotado de modo a ajudar no desenvolvimento de algoritmos computacionais de ajuda ao diagnóstico, facilitando a execução do próximo objectivo deste trabalho.

Um segundo objectivo é estudar, desenvolver e testar metodologias que consigam detectar pólipos e/ou tumores de forma automática, utilizando o vídeo disponibilizado pela cápsula endoscópica. Na literatura ainda não existem trabalhos que façam detecção de modo automático neste género de vídeo, sendo nesse patamar um trabalho inédito. O trabalho efectuado por André Sousa [17] tem como objectivo a ajuda ao diagnóstico automático, mas tem como objecto de trabalho vídeo de endoscopia com magnificação de imagem, com características distintas da cápsula endoscópica. As metodologias existentes e descritas na literatura para essa identificação e detecção cujos resultados foram bastantes promissores estão assentes em vídeo obtido através de colonoscopias, com uma qualidade muito superior à da cápsula. Desta forma, o desafio proposto é o de explorar este tipo de técnicas num tipo de vídeo com pior qualidade.

Os pressupostos na detecção automática de pólipos são os mesmos nas duas situações, o vídeo é filmado do interior do intestino e os pólipos são descritos da mesma forma (cor, morfologia

2 Ver <http://www.rfamerica.com/sayaka/index.html>

e tamanho). Assim vamos verificar a veracidade e aplicabilidade desse método quando temos vídeo extraído a partir da cápsula endoscópica.

No final será necessário avaliar se o método tem fundamentos científicos e de diagnóstico para uma possível integração com o software CAPVIEW.

1.3. Estrutura do documento

A tese inicia com uma breve introdução ao assunto de endoscopia expondo de forma sucinta as diversas técnicas e os problemas a elas associados, bem como uma pequena exposição da história da cápsula endoscópica, núcleo central deste trabalho.

De seguida apresenta-se um capítulo com informação aprofundada relativa ao modo de funcionamento da cápsula, as suas características e o procedimento clínico associado, protocolo este que é necessário cumprir por todos os pacientes. No final apresenta-se o software CAPVIEW de suporte à análise de exames de cápsula endoscópica.

Segue-se uma revisão da literatura dos métodos e teoria necessária para o desenrolar da tese, cobrindo a teoria relacionada com as imagens e respectivo espaço de cores, os métodos de análise e detecção de textura e alguns dos algoritmos de aprendizagem automática (*machine learning*).

O capítulo 5 apresenta o software de catalogação de imagens de cápsula endoscópica desenvolvido, a partir das quais foram avaliadas as metodologias apresentadas neste trabalho. Este software tem uma importância elevada devido à necessidade de obter eventos bem anotados e com elevada confiança científica, por parte de médicos especialistas.

Os capítulos seguintes apresentam o trabalho realizado do qual fazem parte, a triagem de dados para avaliação e a descrição das metodologias adoptadas para atingir o objectivo deste trabalho. Segue-se a apresentação e discussão dos resultados obtidos.

O último capítulo está reservado às conclusões retiradas da aplicação destas metodologias para processamento do vídeo de cápsula endoscópica, bem como uma breve descrição do trabalho que se deve realizar nos próximos estudos desta natureza.

2. Cápsula endoscópica e software de anotação

A cápsula endoscópica é uma mini-câmara fotográfica que flui dentro do sistema digestivo [18] do paciente com o objectivo de fotografar todo o caminho possibilitando uma posterior avaliação e diagnóstico por parte do médico/clínico, das doenças e anomalias que afectam o sistema endogástrico do paciente.

2.1. Características técnicas

Este dispositivo é envolvido por um invólucro resistente a ácidos e fluídos gástricos, lacrado e não reutilizável. A cápsula tem as seguintes dimensões 11 x 26 mm e um peso de 3.7 gramas [19].

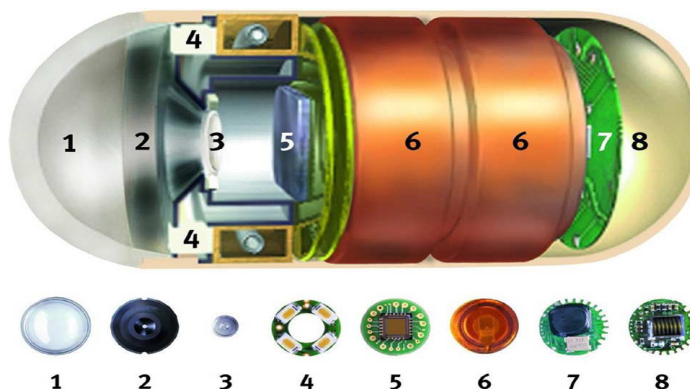


Figura 4: Wireless endoscopic capsule. 1: Optical dome. 2: Lens holder. 3: Lens. 4: Illuminating LEDs. 5: CMOS imager. 6: Battery. 7: ASIC transmitter. 8: Antenna.

Os elementos constituintes da cápsula são visíveis na Figura 4 [20] onde estão presentes as

duas baterias, uma antena UHF, *leds*³ brancos que servem de fonte de iluminação, lentes de focagem curta e um sensor CMOS (câmara). As lentes usadas conseguem ter uma abertura de 140°, ampliação de 1:8 e uma profundidade de 1:30 mm o que lhe permite discriminar objectos com resolução de 0.1 milímetros [19]. A cápsula desloca-se ao longo do intestino devido aos movimentos peristálticos inerentes ao corpo humano. No decorrer do exame capta e transmite para o gravador imagens com uma cadência de duas por segundo com uma resolução de 256x256 *pixels*. A antena UHF utilizada para a transmissão está sintonizada na frequência de 431.1MHz. Toda esta electrónica foi desenhada e dimensionada com o propósito de possibilitar uma autonomia de 7 ± 1 horas [19][15], utilizando apenas as duas baterias existentes.

A Figura 5 [21] apresenta o sistema de aquisição completo. O sistema de cápsula endoscópica é constituído não só pela cápsula mas também pelo sistema de gravação e o software de leitura do vídeo que está armazenado num formato proprietário. O gravador tem que estar fixo ao paciente normalmente na cintura, onde serão conectados os sensores/antenas que recebem os dados emitidos pela cápsula (ver Figura 6), estes sensores possibilitam o cálculo da posição espacial da cápsula no interior do corpo do paciente [6], usando o método de triangulação de sinal recebido entre 3 ou mais antenas.



Figura 5: Sistema de aquisição da Given

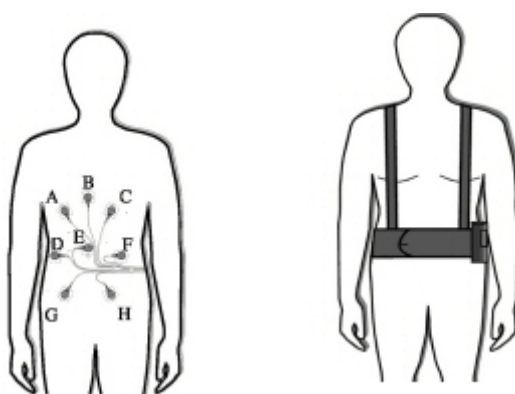


Figura 6: Localização das antenas e dispositivo de armazenamento,
<http://www.mph.ed.ac.uk/endo/patgivilivlabsbce.htm>

O gravador vai recolhendo os dados emitidos pela cápsula e após as 8 horas de exame esses conteúdos são descarregados para o computador onde são posteriormente analisados e

3 LED: light emitter diode – díodo emissor de luz

diagnosticados com ajuda de software de anotação de exames (**Capview** [22], **Rapid** software da Given⁴, **EndoCapsule**⁵ Olympus software). A quantidade de imagens descarregada pela cápsula ao longo das 8 horas é enorme (8h x 2 imagens por segundo = 57600 imagens) logo é necessário software que ajude a manusear, organizar e anotar esses exames. O CAPVIEW é bastante completo, pois possibilita a organização dos exames numa base de dados além de usar uma linguagem normalizada para as anotações dos exames, tornando a revisão por diversos médicos mais simples e compreensível.

Os *softwares* de anotação ajudam na organização dos exames mas ainda existe muito trabalho sistemático e metódico que pode ser feito de modo automático. Essas ferramentas de ajuda ao diagnóstico podem reduzir o tempo que um médico despende na análise de cada exame. Actualmente o médico necessita de despende entre 45 minutos até 2 horas na sua revisão, valores que dependem da experiência e da existência ou não de anomalias clínicas [6], assim sendo todas as ferramentas informáticas que consigam reduzir este tempo são muito apetecidas, abrindo portas para um mercado de negócio especializado.

2.2. Características clínicas

A cápsula fornece informação pertinente sobre o conteúdo do sistema gástrico importante para a detecção de hemorragias internas no intestino delgado [21], entre outras anomalias. Esta técnica fornece conteúdo visual, de vital importância, de zonas do intestino que não são possíveis de alcançar através de técnicas convencionais, técnicas essas que não conseguem percorrer a totalidade da extensão do intestino delgado (média de 6.7 metros) para detectar anomalias, nomeadamente presença de hemorragia, doença de Crohn, detecção de pólipos [21][9] e tumores neuroendócrinos [23] no intestino delgado, ao invés do estômago ou cólon que podem ser analisados utilizando gastro-endoscopia ou colonoscopia, respectivamente.

As desvantagens deste método estão relacionadas com o facto de ser um dispositivo passivo o que impossibilita a realização de biópsias e a impossibilidade de focar a observação numa determinada área de interesse [24]. Uma outra desvantagem que é necessário ter em conta é a possibilidade de retenção o que pode levar à necessidade de intervenção cirúrgica (1.5% a 5% dos casos particularmente na presença de hemorragias e doença de Crohn [19]), assim para evitar cirurgias que não podem ser efectuadas no paciente, devido a restrições médicas ou esse peremptoriamente as não autoriza, convém realizar um exame TID prévio para prever este tipo de situações [24], e prestar os cuidados necessários.

4 Ver em <http://www.givenimaging.com>

5 Ver em <http://www.olympus-europa.com>

2.3. Procedimento clínico

O método de exame através da cápsula endoscópica é não invasivo mas continua a necessitar de alguma preparação específica semelhante à utilizada em outros tipos de endoscopia. É preciso que o paciente esteja em jejum nas 12 horas prévias ao exame e convém fazer uma preparação (limpeza) intestinal no dia anterior ao exame com o intuito de melhorar a qualidade do vídeo e aumentar a probabilidade de a cápsula chegar ao cólon antes da bateria se esgotar, aperfeiçoando desta forma a capacidade de diagnóstico por parte do clínico/médico [19,25].

Após a ingestão da cápsula o paciente pode deixar o hospital e fazer a sua rotina diária normal evitando apenas exercícios físicos ou esforços que possam fazer com que os sensores se descolem do peito do paciente. Duas horas após a ingestão é possível beber alguns líquidos e 4 horas depois é possível fazer uma refeição leve [19].

No final do dia, isto é passado as 8 horas o paciente regressa ao hospital para que o médico retire os sensores e descarregue a informação da cápsula para o computador, informação essa que só é revista, usando software específico, *à posteriori*, devido à sua longa duração facto que impossibilita o médico de efectuar um diagnóstico no imediato.

2.4. Software de anotação: CAPVIEW

O Capview é um software de anotação clínica de exames de cápsula endoscópica que contém a marca do grupo de investigação SIAS (Sistemas de Informação na Área de Saúde) pertencente ao IEETA⁶ (Instituto de Engenharia, Electrónica e Telemática de Aveiro) localizado na Universidade de Aveiro e foi desenvolvido em estreita ligação com especialistas de gastroenterologia do Hospital de S. António no Porto [13,22].

⁶ IEETA: <http://www.ieeta.pt>

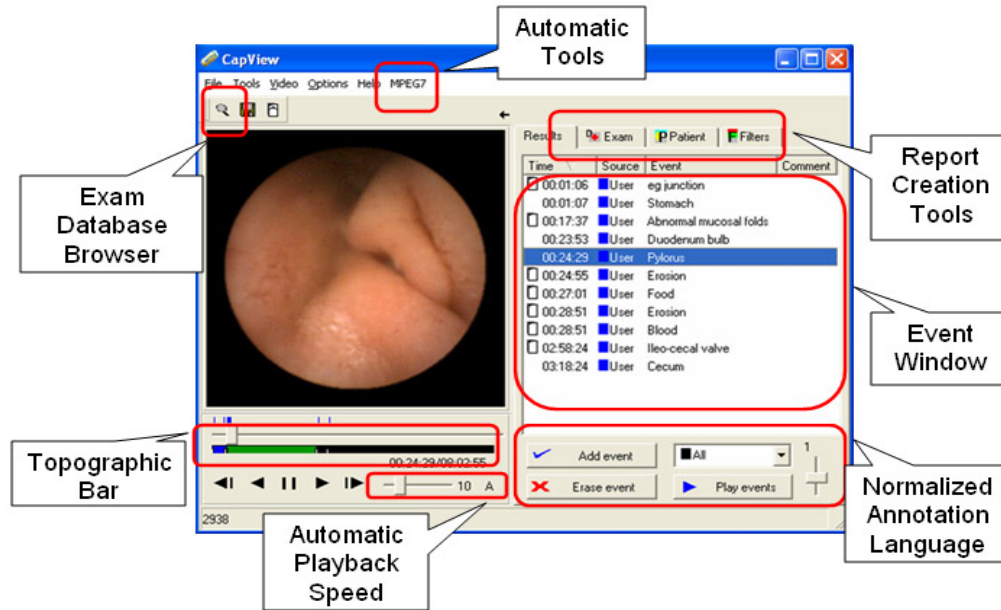


Figura 7: Capview: janela principal, <http://www.capview.org>

O software permite visualizar e anotar os exames de cápsula endoscópica, mas contém outras vantagens como gerir a enorme quantidade de exames efectuados por um departamento de gastroenterologia. Um outro factor de extrema importância que este software tenta promover é a normalização das nomenclaturas usadas na anotação dos exames e respectiva criação dos relatórios [13].

A gestão dos exames, semelhante a uma base de dados, torna possível a realização de operações de pesquisa de exames por paciente, ou por determinado diagnóstico, sendo estes processos realizados de modo expedito, sem confusões ou trocas de processos de pacientes.

A normalização dos termos usados pelos médicos na criação de relatórios tem inúmeras vantagens de que se salienta: 1) a facilidade de revisão, 2) compreensão por outros especialistas 3) em termos informáticos torna a pesquisa e indexação na base de dados do capview mais simples e rápida. Os termos usados na anotação dos exames foram seleccionados de listas da WODE⁷ (*World Organization for Digestive Endoscopy*) [13] e subdivididos em categorias – localização, lúmen, lesão, conteúdo, mucosa, entre outras. A normalização obriga a que o médico indique e caracterize minimamente os eventos (lesões, hemorragias, pólipos) e a sua localização topográfica.

O software incorpora um algoritmo de segmentação topográfica automática que automatiza o processo de detecção de eventos associando cada um à sua localização no sistema gástrico, disponibiliza também informações como o tempo de trânsito que a cápsula demora em cada uma das quatro zonas do sistema gástrico (entrada, estômago, intestino delgado, intestino grosso) das quais é possível retirar informação clínica [13,20,26].

7 Também conhecido como OMED, Organisation Mondiale d'Endoscopie Digestive, www.omed.org

3. Espaço de cores e algoritmos de aprendizagem automática

3.1. Espaço de Cores

O corrente trabalho tem como principal objectivo a análise de vídeo. As respectivas *frames* ou imagens estão codificadas num espaço de cores, sendo o mais comum para armazenamento o RGB (*Red, Green, Blue*) [27]. Este espaço possui algumas desvantagens a nível de processamento de imagem específico e daí a existência e desenvolvimento de outros espaços de cores, como sejam o HSV/L, e o Karhunen-Loeve.

Os sensores digitais usados nas câmaras e outros dispositivos electrónicos são constituídos por CCD's⁸ que captam a quantidade de energia de cada banda de luz que resulta da interacção desses objectos com uma fonte de luz [28], armazenando a imagem na memória do dispositivo no formato definido, normalmente RGB.

3.1.1. RGB

Uma imagem no espaço RGB é composta por uma matriz de tamanho $M \times N$ *pixels* representante dos pontos no espaço, e cada *pixel* é constituído por um conjunto de 3 valores correspondentes às componentes *Red, Green e Blue* (ver Figura 8) [29].

A maneira mais simples de perceber uma

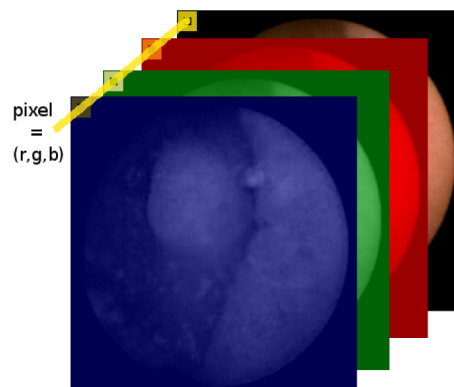


Figura 8: RGB colorspace

⁸ CCD: *charge coupled device*

imagem é imaginar uma sobreposição de 3 imagens em que cada uma delas corresponde à refração da cor Vermelha, Verde ou Azul (RGB respectivamente), e como resultado desta adição temos a noção da cor do objecto e ambiente que o envolve.

Este espaço de cores aditivo é muito utilizado para visualização em dispositivos electrónicos, mas a sua simplicidade [17] é um factor negativo quando utilizado para extracção de características de textura, devido à alta correlação entre as diversas componentes e a sua fraca uniformidade perceptual [30,31].

3.1.2. Espaço de cor HSV / HSL

O espaço de cores *HSV/L* (*Hue, Saturation and Value / Luminance*) é usado num vasto conjunto de análises e processamento de imagem por ser mais intuitivo e perceptual para os humanos. Além disso possibilita isolar o canal de luminância (*value*), quantidade de luz reflectida por um objecto [32], dos canais de saturação que indica a intensidade da cor, e de tom (*hue*) que é descrito pela cor em questão. É esta propriedade que identifica se a cor de um ponto é verde, vermelho ou outra.

Na Figura 9 é possível observar a excursão de cada componente. A componente *hue* descreve a cor dada de forma angular no topo do cone, a *saturation* e *value* são descritas em percentagem, sendo que a *value* descrita pela profundidade ou altura do cone tem como mínimo a cor preta (vértice do cone) e máximo o branco, enquanto que a *saturation* varre a área da secção e que pode ir desde o branco até à cor pura definida pelo *hue*.

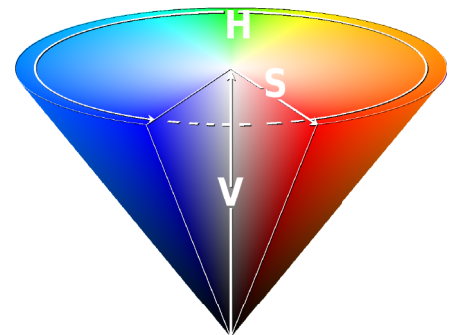


Figura 9: Espaço HSV

A transformação entre o espaço RGB e HSV é dada por:

$$V = \text{Max}(r, g, b)$$
$$S = \begin{cases} \frac{V - \text{Min}(r, g, b)}{V}, & V \neq 0 \\ 0, & V = 0 \end{cases}$$

$$H = \begin{cases} 0, & (S=0) \\ \frac{60(g-b)}{V - \text{Min}(r, g, b)}, & (r=V \wedge g \geq b) \\ 360 + \frac{60(g-b)}{V - \text{Min}(r, g, b)}, & (r=V \wedge g < b) \\ 120 + \frac{60(g-b)}{V - \text{Min}(r, g, b)}, & g=V, \\ 240 + \frac{60(g-b)}{V - \text{Min}(r, g, b)}, & b=V \end{cases}$$

Assumindo que as componentes r,g,b estão normalizadas no intervalo [0,1], as componentes resultantes V e S estão no intervalo [0,1] enquanto que Hue está no intervalo de [0, 360] em graus [33].

3.1.3. Karhunen-Loève

Este espaço de cores foi desenvolvido para eliminar a correlação das componentes existente no espaço RGB.

Assenta numa transformação matricial do espaço RGB, dada por

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 0.333 & 0.333 & 0.333 \\ 0.500 & 0.000 & -0.500 \\ -0.500 & 1.000 & -0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

onde K_1 , K_2 , K_3 são as coordenadas do espaço K-L [16,34].

Na realidade, a transformação Karhunen Loeve é formada pelos vectores próprios da matriz de correlação da imagem, e dessa forma é dependente da imagem, mas Ohta [35] demonstrou que esses vectores são muito semelhantes para imagens da mesma natureza e que essa variação é mínima. Van De Wouwer [34] verificou e mostrou que a maior variância se centra na componente K_1 que está relacionada com a intensidade da imagem. A segunda componente com maior variância é K_2 , restando a componente K_3 com uma quantidade mínima de informação. A transformação de uma imagem para uma referência de eixos ortogonais elimina a correlação existente entre as componentes do espaço RGB.

3.2. Algoritmos de aprendizagem automática

3.2.1. Support Vector Machines (SVM)

A técnica de aprendizagem automática SVM foi desenvolvida na década de 70, por Vladimir Vapnik [36] com o objectivo de ser utilizada em problemas de reconhecimento de padrões, e desde então ganhou grande reconhecimento sendo utilizada em diversas aplicações como reconhecimento de escrita, voz, facial [36,37] e em reconhecimento de padrões em vídeo clínico [13,38,39]. Inicialmente desenvolvida como classificador binário, existem técnicas para generalizar a técnica para um classificador de classes múltiplas. Essas generalizações utilizam combinações das N classes, por exemplo n classes contra todas as restantes, sendo que $n = 1...(N-1)$ [36]. Existem extensões ou variantes desta técnica para realizar regressão linear ou não-linear [40].

Esta técnica de aprendizagem supervisionada, está dividida em duas fases. A primeira centra-se no problema de ajuste de curvas ou interpolação, de onde é extraído um modelo matemático dos dados apresentados, modelo este que distingue as duas classes em causa, nesta fase está centrado a maior parte do processamento. Na fase seguinte o modelo criado é utilizado para determinar a classe resultante de um novo conjunto de dados.

O classificador SVM mapeia os vectores num espaço de dimensão N , e cria um hiperplano que separa os dados de forma a manter uma margem de distância das classes o maior possível

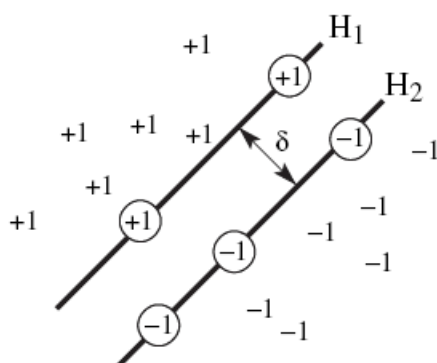


Figura 10: Separação máxima do hiperplano

A imagem Figura 10 (obtida de [37]) demonstra duas classes (+1 e -1) distintas que podem ser facilmente separadas por uma recta (classificador linear). Nos classificadores binários lineares procura-se os hiperplanos que maximizem a separação entre as duas classes (δ) que são únicos e conseguem discriminar ambas as classes, funcionam como margem ou fronteira de cada classe. Cada um desses hiperplanos (representados por H1 e H2 na imagem) são definidos por *support*

vectors representados com um círculo. O modelo gerado pelo algoritmo SVM é representado pelos *support vectors* de cada hiperplano.

No caso de termos um problema que não seja linear (Figura 11), mas no qual queiramos aplicar a técnica de SVM é obrigatório utilizar funções de transformação ou linearização do espaço (Φ) características (*feature space*), podendo de seguida aplicar o algoritmo em estudo.

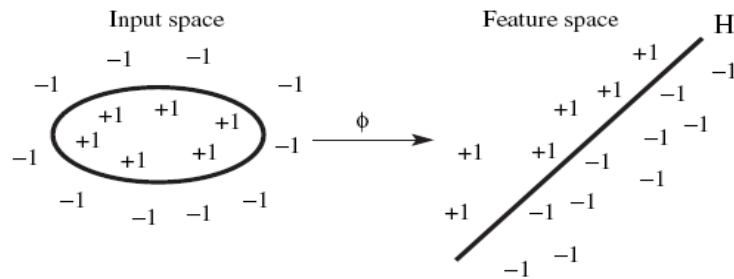


Figura 11: Linearização das características

Estas funções são denominadas por *kernel* e normalmente são funções não lineares sendo as mais comuns [36]:

Linear: $K(x, y) = x \cdot y + 1$

Polinomial: $K(x, y) = (x \cdot y + 1)^p$, p relativo à ordem do polinómio

Rbf: $K(x, y) = e^{-\|x-y\|^2 / 2\sigma^2}$, σ^2 relativo à variância da função gaussiana

Sigmoid: $K(x, y) = \tanh(k \cdot x \cdot y - \delta)$, k e δ relativos ao limite em amplitude e ao deslocamento da função, respectivamente. Sabendo que $p, \sigma, k, \delta \in \mathbb{R} \setminus \{0\}$.

Não existe nenhum método que especifique qual o melhor *kernel* que deve ser usado em cada problema, dessa forma é necessário recorrer à experimentação e ao método empírico para fazer essa escolha [36].

3.2.2. Redes Neurais Artificiais (Neural Networks)

As redes neuronais foram desenvolvidas com o objectivo de imitar os neurónios e sinapses existentes no cérebro humano. Os nós da rede neuronal são comparados aos neurónios e as arestas dessa rede são análogas às sinapses entre neurónios. Este tipo de sistemas não-lineares têm a sua génese no instituto MIT em 1943 por McCulloch e Pitts [41].

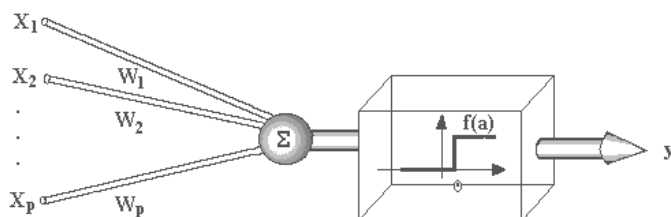


Figura 12: Neurónio projectado por McCulloch

Imagem obtida de <http://www.din.uem.br/ia/neurais/>.

O neurónio da Figura 12 foi projectado de forma a simular o comportamento dos neurónios humanos onde o sinal de saída depende da soma de uma quantidade de sinais de entrada e duma função chamada de “activação”.

A própria Figura 12 mostra que o processo não pode ser efectuado apenas com um neurónio, depende de somatório de diversos sinais de entrada, e se supusermos que cada um desses sinais é a saída de outro neurónio a necessidade de estruturar o sistema como uma árvore, rede ou grafo é evidente.

Uma rede neuronal pode ser disposta em diversas topologias [42], rede em árvores (unidireccionais, *feed-forward*) ou circulares (recorrentes e com realimentação, *back-propagation*). Essas redes são constituídas por diversos elementos de processamento, denominados de funções de activação, que podem ser desde uma simples função degrau, visível na Figura 12, ou podem ser de outro género como *sigmoid*, gaussiana ou linear.

Em geral as redes neuronais estão divididas em 3 tipos de camadas (ver Figura 13). A camada de entrada que pode ser constituída por um ou mais neurónios, uma camada intermédia denominada de *hidden layer* (camada escondida) onde as ligações entre as unidades de processamento (neurónios) podem ser tão complexas quanto necessário, e uma camada de saída com o vector resultado pretendido.

As quatro principais regras para aprendizagem automática são as seguintes: 1) correcção de erro, 2) regra de Boltzmann, 3) regra de Hebbian e 4) aprendizagem competitiva.

O método como estes sistemas conseguem adaptar-se é através da minimização das funções de erro, e para isso é necessário que as unidades computacionais consigam alterar os pesos, ou parâmetros da função de activação. Estes elementos adaptáveis são denominados de perceptrões.

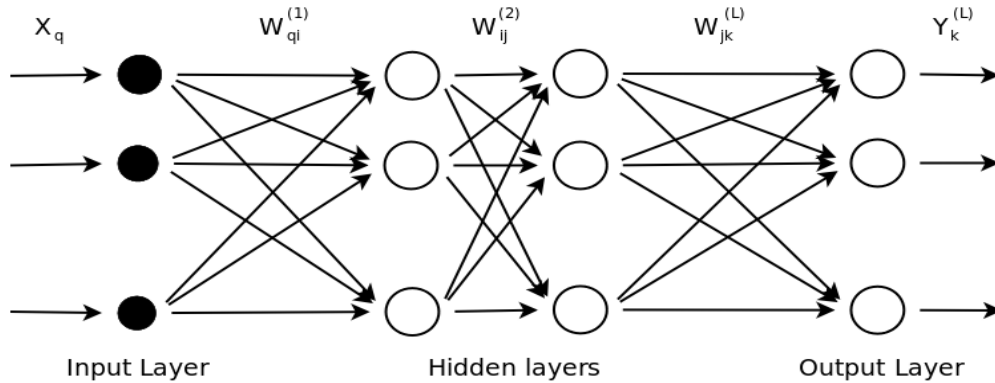


Figura 13: Rede neuronal com 3 camadas intermédias
 Imagem editada de [42].

Os pesos das camadas da rede neuronal são actualizadas seguindo a regra de correcção de erros dada pela expressão, $W_j(t+1) = W_j(t) + \eta(d - y)X_j$ onde d é a saída desejada, t a iteração, e $\eta \in [0,1]$ é o ganho e j o índice da camada da rede neuronal.

A grande vantagem destes métodos é a sua capacidade de aprender regras implícitas do tipo causa-efeito a partir de um largo conjunto de dados de exemplo, e não necessitam de regras explícitas pelos humanos à partida.

As redes neuronais são bastante utilizadas em diversas áreas de aplicação desde bio-medicina, economia e mercado de valores, reconhecimento de padrões de imagem e/ou voz [42].

3.2.3. Bayes Networks

A rede de *Bayes*, usada como classificador estatístico, é representada num grafo acíclico direccionado (DAG *Direct acyclic graph*) onde cada vértice representa uma variável em estudo e as arestas as relações entre elas. Essas arestas representam a probabilidade condicional da variável em estudo conhecendo os eventos passados representados pelos vértices a montante conectados a essa variável [43]. A rede de *Bayes* implementa o teorema de *Bayes* (probabilidade condicional) a sistemas complexos de modo hierárquico.

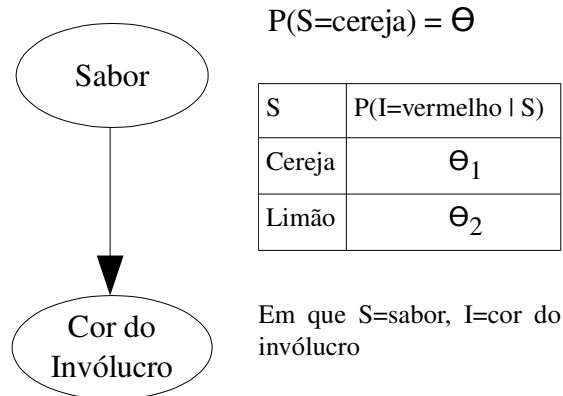


Figura 14: Modelo bayesiano

Figura editada de [44]

Na Figura 14 está presente um exemplo ilustrativo de uma rede de *Bayes* simples, onde a cor do invólucro depende do sabor do rebuçado de forma probabilística. Assim é possível prever qual o sabor de um rebuçado através da cor do seu invólucro.

$$\begin{aligned}
 P(S=cereja, I= verde | \Theta, \Theta_1, \Theta_2) \\
 &= P(S=cereja | \Theta, \Theta_1, \Theta_2) \cdot P(I= verde, S=cereja | \Theta, \Theta_1, \Theta_2) \\
 &= \Theta \cdot (1 - \Theta_1)
 \end{aligned}$$

Equação 1: Probabilidade de um rebuçado num invólucro verde.

Equação obtida de [44]

Este classificador necessita que os modelos, grafos e correspondentes funções distribuição de probabilidades de cada vértice sejam conhecidos *à priori* [44,43], e que sejam de alta confiança pois caso essas probabilidades sejam muito pessimistas ou optimistas podem distorcer a rede e invalidar os respectivos resultados [45].

3.2.4. Métodos de avaliação

A avaliação da qualidade e eficácia destes métodos e algoritmos é essencial para que se consiga discernir se os resultados do modelo criado estão de acordo com o previsto, ou se o melhor é descartar esse modelo.

Aceitar o modelo de forma cega é má ciência, daí a necessidade de um mecanismo de controlo e referência de comparação para com os outros algoritmos e estudos.

A forma ideal de avaliar um algoritmo deste tipo é ter presente dois conjuntos de dados um

para treino e geração do modelo e outro para testar esse modelo. Esses conjuntos de dados têm de ser disjuntos e com um vasto número de instâncias. Cada um desses conjuntos tem de ter uma quantidade de instâncias de diferentes classes mas com uma distribuição equitativa, isto é o número total de instâncias de uma classe deve ser equiparado a qualquer outra classe presente nesse conjunto.

Quando estamos a utilizar dados médicos e clínicos, essa distribuição equitativa das diversas classes é muito difícil de encontrar [46], o que significa que a maior parte dos pacientes que participam nesses estudos são saudáveis. Neste caso a melhor técnica para efectuar a avaliação dos modelos é a utilização de *n-fold cross-validation* que é o método de excelência em casos onde existe escassez de dados ou a relação entre o total de instâncias de cada classe é bastante elevado [47].

A técnica de *cross-validation* assenta na divisão aleatória do conjunto de instâncias em duas partes, sendo que uma delas têm o dobro de instâncias da outra, isto é um conjunto fica com 33% e o outro com 67% das instâncias do conjunto inicial. Dos resultantes, o maior (o de 67 %) é usado para o treino do modelo e o restante é utilizado para testar o modelo. A relação entre os conjuntos pode ser diferente, mas esta é a mais usual.

O teste do modelo usando apenas esta divisão é pouco válida porque sendo a divisão aleatória os conjuntos criados podem não ser representativos, isto é a divisão pode ser efectuada de forma a que no conjunto de treino tenha muito mais instâncias de uma classe do que outra, o que por sua vez influencia o conjunto de teste que vai ter uma relação inversa e consequentemente os resultados não vão ser precisos, devido à sobre-representação de uma determinada classe.

De forma a conseguir minimizar o efeito resultante do processo aleatório utiliza-se o mesmo método mas de forma recursiva ou repetida n vezes, e no final apresenta-se uma média dos resultados de cada *cross-validation* intermédio. Esta técnica é denominada *n-fold cross-validation*.

A técnica de *n-fold cross-validation* mais comum é utilizar $n=10$. Este valor é resultado de diversos estudos empíricos que mostram que esse valor é o que consegue uma melhor taxa de erro, existindo também algum estudo teórico que comprova essa escolha [47].

Os conjuntos criados de forma aleatória, possibilitam a utilização da estatística e probabilidades de forma incondicional, e assim é possível realizar o teste diversas vezes usando o *10-fold cross-validation*, e dessa forma minimizar o erro cometido.

4. Métodos de detecção automática de pólipos

Actualmente as anomalias existentes no sistema digestivo são detectadas pela perícia e visão que os médicos endogastrologistas adquiriram ao longo dos anos, através da experiência na realização de endoscopias em diversos pacientes. Com o evoluir da tecnologia desenvolveu-se técnicas inovadoras para despistar essas anomalias, pólipos e/ou tumores no intestino, com suporte em TC⁹ [48] e RM¹⁰ [4] ambas pertencentes à vertente de endoscopia com suporte radiológico. As técnicas radiológicas geram muita informação imagiológica (visual) o que torna os algoritmos de *computer vision* bastante desejados para realizar este despiste de modo automático e expedito.

Na linha de desenvolvimento de detecção de pólipos utilizando ferramentas automáticas e computadorizadas existem duas abordagens principais.

A primeira é a endoscopia virtual que consegue fornecer imagens de alta resolução em situações onde a endoscopia normal é insuficiente ou inadequada, na presença de obstruções ou quando o paciente não colabora com a execução do exame [49]. Este tipo de exame é efectuado através de estudos de TC, RM, e/ou ultra-som, onde é feito o registro dos diversos dados obtidos por forma a criar um modelo a 3 dimensões (3D) do órgão em causa, neste caso do intestino do paciente. A criação desse modelo é baseada num processo matemático denominado *shaded surface* ou *volume rendering* [49,4]. Este exame é preferencialmente utilizado na zona do cólon (intestino grosso) daí que existam diversos trabalhos acerca de colonoscopia virtual [50,51,3,52] enquanto que os estudos no intestino delgado são bastante mais escassos. Esta escassez deve-se sobretudo à inexistência de ferramentas de processamento capazes de tratar a quantidade de informação geométrica, relativa ao intestino delgado e inerente à estrutura curva e tubular no espaço tridimensional [5].

9 TC: Tomografia computadorizada

10 RM: Imagem de Ressonância Magnética

A outra vertente é usar a vídeo endoscópico e através de técnicas de processamento de vídeo ajudar na detecção de pólipos e tumores no intestino delgado e/ou grosso. No caso da detecção de anomalias a nível do cólon (intestino grosso) a colonoscopia é mais expedita, mas quando se trata de analisar o intestino delgado é bastante complicado, tornando a cápsula endoscópica no exame por excelência.

O maior problema da cápsula é a sua imaturidade a nível de software de ajuda ao diagnóstico automatizado, facto que pode trazer muitas possibilidades de negócio para quem conseguir desenvolver software com eficácia suficiente para ajudar os médicos no diagnóstico clínico. Já existem diversos trabalhos científicos utilizando como objecto de estudo a cápsula endoscópica mas a maioria são estudos médicos, isto é utilização da cápsula em determinadas situações clínicas e comparações com os outros tipos de endoscopia tradicionais [7,8,10,53-55].

Os estudos científicos sobre a cápsula endoscópica, com o intuito de criar ferramentas automáticas de detecção de eventos ou anormalidades, são escassos [13,56]. Existem alguns trabalhos sobre processamento de vídeo e extracção de características usando MPEG-7 [57,58] mas necessitam de maior validação e focalização no plano da utilização de vídeo de cápsula endoscópica, tendo em conta a natureza do sistema gástrico.

No campo de detecção de pólipos e tumores, um dos trabalhos promissores é o da detecção de textura usando como método de co-variância de wavelets [16] , e a escolha da utilização da textura como parâmetro fundamental é explicada pela caracterização das anomalias que têm em conta a cor, morfologia e tamanho.

4.1. Análise de textura

A textura é a característica mais importante nos eventos a detectar pois as suas propriedades definem a superfície ou estrutura de um objecto existente numa imagem, a sua simplicidade de discriminação para os humanos não se reflecte nas ferramentas computacionais, facto que se deve à sua enorme variabilidade e indefinição. Vamos considerar que uma textura é um grupo ou área com um conjunto de *pixeis* contendo duas dimensões, uma descrevendo as primitivas da textura e outra descrevendo a dispersão espacial dessa primitiva ao longo da imagem [28,59].

As primitivas são texturas elementares que compõem uma textura de tamanho superior, constituídas por um conjunto contíguo de *pixeis* com as mesmas propriedades de cor e região espacial, denominadas de *texels* [28]. Pode-se pensar num *texel* como um grão de areia numa fotografia de uma praia.

Assim, uma imagem pode ser interpretada como um conjunto de texturas onde cada uma é composta e descrita por um conjunto de primitivas, pelo seu tipo e sua dispersão espacial [28].

Os processos de análise de textura divergem nos métodos e aplicações de cada um. O

método de descrição de textura mais comum e mais utilizado é o estatístico onde cada textura é descrita usando vectores de característica composto por parâmetros que representam um ponto num espaço multidimensional. Nesta abordagem são utilizados coeficientes de auto-correlação, matrizes de co-ocorrência e detecção de arestas. As matrizes de co-ocorrência utilizam estatísticas de segunda ordem [60] ou seja relações espaciais entre *pixels* (ver abaixo).

4.1.1. Matrizes de Co-Ocorrência

Esta técnica é baseada na frequência de ocorrência do mesmo *pixel* (em níveis de cinza) ao longo da textura. Supondo uma imagem de tamanho MxN, uma matriz de co-ocorrência é uma matriz de frequência relativa $P_{\phi,d}(a,b)$ descrevendo o quão frequente dois *pixels* (a,b) aparecem numa janela de tamanho d e na direcção ϕ . Assim são formalmente definidas quatro matrizes para os ângulos 0°, 45°, 90° e 135°, e são descritas da seguinte forma [60]:

$$P_{0^\circ,d}(a,b) = \left| \left\{ [(k,l), (m,n)] \in D : k-m=0, |l-n|=d, f(k,l)=a, f(m,n)=b \right\} \right|$$

$$P_{45^\circ,d}(a,b) = \left| \left\{ [(k,l), (m,n)] \in D : (k-m=d, l-n=-d) \vee (k-m=-d, l-n=d), f(k,l)=a, f(m,n)=b \right\} \right|$$

$$P_{90^\circ,d}(a,b) = \left| \left\{ [(k,l), (m,n)] \in D : |k-m|=d, l-n=0, f(k,l)=a, f(m,n)=b \right\} \right|$$

$$P_{135^\circ,d}(a,b) = \left| \left\{ [(k,l), (m,n)] \in D : (k-m=d, l-n=d) \vee (k-m=-d, l-n=-d), f(k,l)=a, f(m,n)=b \right\} \right|$$

Equação 2: Matrizes co-ocorrência (4 fórmulas acima)

onde $|\{...\}|$ refere-se à cardinalidade do conjunto e $D = (M \times N) \times (M \times N)$. Da Equação 2 obtém-se que $P_{\phi,d}(a,b) = P_{\phi,d}(b,a)$, levando a que as matrizes sejam simétricas.

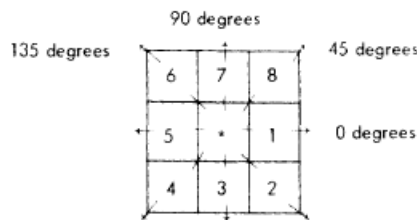


Figura 15: Diversos graus de resolução
Imagem obtida de [60]

O seguinte exemplo (Figura 16) mostra o resultado das matrizes de co-ocorrência de uma matriz 4x4, exemplificativa de uma imagem em nível de cinzentos dada por M.

$$\begin{array}{l}
 \mathbf{M} = \begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix} \\
 \mathbf{P}_{0^\circ,1} = \begin{matrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{matrix} \\
 \mathbf{P}_{135^\circ,1} = \begin{matrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{matrix} \\
 \mathbf{P}_{90^\circ,1} = \begin{matrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{matrix} \\
 \mathbf{P}_{45^\circ,1} = \begin{matrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{matrix}
 \end{array}$$

Figura 16: Cálculo das matrizes de co-ocorrência para uma matriz de 4x4

Haralick descreveu diversas características estatísticas que se podem retirar das matrizes de co-ocorrência, destas as mais relevantes são a energia, entropia, contraste, correlação, *inverse difference moment* (IDM) e dissemelhança [16,59]. Nas equações seguintes (de Equação 3 até Equação 8) assume-se a função $p(i, j) = P(i, j) / R$, em que R é uma constante de normalização, e $P(i, j)$ é uma das matrizes definidas atrás por $P_{\phi, d}(i, j)$, e $Ng = \max(M) + 1$ que corresponde ao número de tons de cinza da imagem.

$$F_1 = \sum_i \sum_j p(i, j)^2, \quad F_1 \in [0, 1]$$

Equação 3: Energia

A energia, também conhecido como *angular second moment*, ou potência da imagem, está relacionado com a homogeneidade da imagem, numa imagem homogenea as transições de nível de cinzento são poucas.

$$F_2 = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}, \quad F_2 \in [-1, 1]$$

Equação 4: Correlação

Onde μ_x, μ_y, σ_x e σ_y correspondem à média e desvio padrão associado às linhas e colunas relativa à iteração ij .

A correlação mede o grau de semelhança entre as linhas e colunas da imagem, o valor é elevado quando os valores estão distribuídos uniformemente na matriz da imagem, e baixo nos outros casos.

$$F_3 = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{p(i, j)}{1 + (i - j)^2}, \quad F_3 \in [0, 1]$$

Equação 5: IDM

O IDM, *inverse difference moment*, está relacionado com a homogeneidade e oferece uma medida global da semelhança em valor entre *pixels* vizinhos. Assim, atinge o máximo (1.0) quando todos os valores são idênticos – imagem sem textura [61].

$$F_4 = - \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i, j) \log(p(i, j)), \quad F_4 \in [0, 2 \log_2 Ng]$$

Equação 6: Entropia

A entropia como medida estatística demonstra a aleatoriedade existente na imagem, quanto maior for a complexidade maior é a entropia da imagem [62], normalmente calculado em imagens de um canal apenas (escala de cinzentos), caso a imagem contenha canais de cor esta medida é realizada em cada um dos canais de forma independente.

$$F_5 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i, j), \quad F_5 \in [0, (N_g - 1)^2]$$

Equação 7: Contraste (difference moment)

Esta medida tal como o nome indica mede o contraste de uma imagem, isto é a quantidade de variações locais, também é denominado de inércia.

$$F_6 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j| p(i, j), \quad F_6 \in [0, 1]$$

Equação 8: Dissemelhança

A dissemelhança mede a falta de semelhança dos *pixels* numa imagem.

Como é possível verificar acima as diversas estatísticas estão relacionadas com mais que um factor ou aspecto de textura. Este fenómeno torna difícil a tarefa de conseguir quantificar a textura de determinada imagem ou objecto.

4.1.2. CWC: Color Wavelet Covariance

A técnica de matrizes de co-variância no espaço wavelet entre cada canal de cor é baseada na co-variância de estatísticas de segunda ordem. Esta informação de segunda ordem foi usada por estar bastante relacionada com a percepção humana e discriminação de textura [16]. As estatísticas utilizadas são as descritas por Haralick (ver Equação 3, 4, 5, 6, 7 e 8) que são calculadas para cada canal da imagem de modo independente seguida do cálculo da co-variância entre essas estatísticas.

A transformação wavelet em duas dimensões é implementada por um banco de filtros, de forma a gerar uma imagem de mais baixa resolução (L_n , em que n é o factor de escala) e um conjuntos de imagens de detalhe ($D_1 \dots D_{3n}$, sendo n o factor de escala). A repetição deste processo leva à decomposição da imagem em diferentes níveis. As equações que descrevem este processo são dadas por [63]:

$$\begin{aligned} L_k(b_i, b_j) &= [H_x * (H_y * L_{k-1})_{\downarrow 2,1}]_{\downarrow 1,2}(b_i, b_j) \\ D_{3k-2}(b_i, b_j) &= [H_x * (G_y * L_{k-1})_{\downarrow 2,1}]_{\downarrow 1,2}(b_i, b_j) \\ D_{3k-1}(b_i, b_j) &= [G_x * (H_y * L_{k-1})_{\downarrow 2,1}]_{\downarrow 1,2}(b_i, b_j) \\ D_{3k}(b_i, b_j) &= [G_x * (G_y * L_{k-1})_{\downarrow 2,1}]_{\downarrow 1,2}(b_i, b_j) \end{aligned}$$

Equação 9: Decomposição wavelet

A imagem original é representada por $L_0 = I(\vec{x})$ (factor de escala ou decomposição igual a 0), os

operadores $\downarrow 2,1$ e $\downarrow 1,2$ representam o processo de redução de taxa de amostragem ao longo das linhas e colunas, respectivamente, o asterisco (*) representa o processo de convolução, H_x e H_y são os filtros passa baixo na direcção das colunas e linhas da imagem, respectivamente responsáveis pelas imagens de baixa resolução, enquanto os G_x e G_y são filtros passa alto na direcção das colunas e linhas respectivamente, sendo estes responsáveis pelas imagens de detalhe [63]. As imagens de detalhe dadas por D_n , $n=1,2,3,\dots,3n$, onde n é o nível de decomposição. A imagem original é representada por conjuntos de imagens de baixa resolução e imagens de detalhe, $I^i=\{L_n^i, D_l^i\}$, sendo $i=1,2,3$, $l=1,2,3,\dots,3n$, sendo n o nível de profundidade, i o índice do canal de cor. A representação visual desta decomposição é visível na Figura 17.

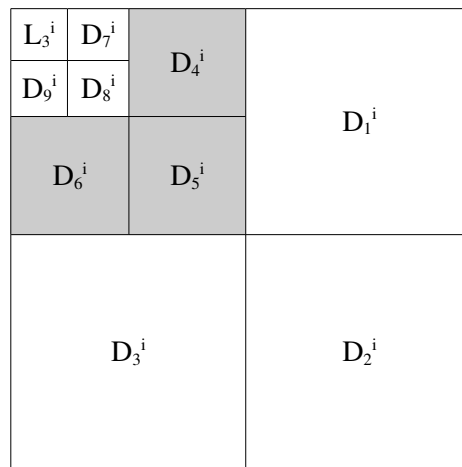


Figura 17: Representação da decomposição wavelet em 3 níveis

A informação de textura está melhor representada nos canais intermédios da decomposição wavelet, assim apenas são considerados os coeficientes de segundo nível, o que resulta num conjunto de nove sub-imagens, dadas por $\{D_l^i\}$, com $i=1,2,3$, $l=4,5,6$ [16]

As matrizes de co-ocorrência codificam as funções de densidade probabilidade $f(i, j, d, \theta)$ calculada, contando a quantidade de pares de *pixels* distantes de d numa direcção Θ e que contenham os níveis de cinza i e j , normalmente os valores para Θ são $\{0, 45^\circ, 90^\circ, 135^\circ\}$. Cada uma destas quatro funções são representadas por matrizes, das quais vão ser calculadas as estatísticas de Haralick (apenas as 4 primeiras F_1 até F_4).

Vamos definir as funções de co-ocorrência por $C_a(D_l^i)$, $i=1,2,3$, $l=4,5,6$, $a=0^\circ, 45^\circ, 90^\circ, 135^\circ$, dos quais o índice i indica os canais de cor, o l indica as imagens de detalhe obtidas do 2º nível de decomposição wavelet, e a variável a indica que as matrizes de co-ocorrência vão ser calculadas nas 4 direcções indicadas anteriormente. Então as estatísticas são definidas por

$$F_m(C_a(D_l^i)), \quad i=1,2,3, \quad l=4,5,6, \quad a=0^\circ, 45^\circ, 90^\circ, 135^\circ, \quad m=1,2,3,4$$

Equação 10: Cálculo das estatísticas das matrizes de co-ocorrência

onde m é a estatística a calcular. Após estas estatísticas é necessário calcular a co-variância das

estatísticas relativas a cada canal de cor, dado pela seguinte fórmula:

$$Cov[F_m(C_a(D_l^i)), F_m(C_a(D_l^j))] = \sum_a [F_m(C_a(D_l^i)) - E(F_m(C_a(D_l^i)))] \times [F_m(C_a(D_l^j)) - E(F_m(C_a(D_l^j)))]$$

Equação 11: Co-variância das estatísticas

$$CWC_m^l(i, j) = \begin{cases} Cov[F_m(C_a(D_l^i)), F_m(C_a(D_l^j))], & i < j \\ Var[F_m(C_a(D_l^i))], & i = j \end{cases}$$

Equação 12: Co-variância wavelet

O processo de extracção do vector CWC (Equação 12) origina, para cada imagem em estudo, um vector de 72 componentes em que 36 correspondem às variâncias por serem calculadas no mesmo canal de cor, e as outras 36 são relacionadas com as co-variâncias de canais de cor distintos. Nos algoritmos de classificação automática são utilizadas estas 72 componentes para construir instâncias (vector de entrada) que identificam a imagem em estudo.

4.1.3. LBP: Local Binary Patterns

O LBP é outro descritor de textura bastante robusto e com diversas aplicações, desde a análise de imagens RM até reconhecimento facial [64,65].

Este método funciona efectuando um limiar à vizinhança de uma matriz usando como referência o valor central e representa o resultado como um valor binário dependente desse limiar.

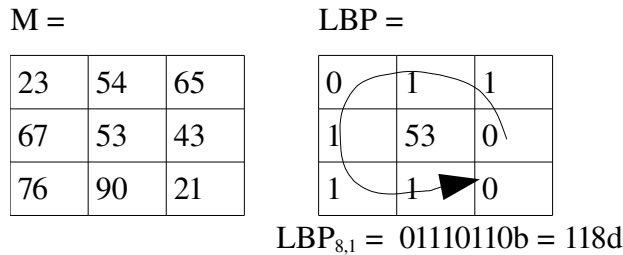


Figura 18: Cálculo do histograma LBP para uma matriz 3x3

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \text{ em que } s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Equação 13: LBP

gc = nível de cinza central
P = total de membros da vizinhança
R = raio

No caso apresentado da Figura 18 está demonstrado o processo para uma região de 3x3 (raio=1), mas este processo é possível de realizar para regiões com áreas superiores. O histograma dos valores resultantes pode ser usado como um descritor de textura (Equação 13). Uma das maiores potencialidades deste método é a sua simplicidade de cálculo, o facto de ser robusta, utilizar imagens de níveis de cinzento e ser invariante a rotações [64-66].

5. Ferramenta de anotação de eventos ECCA¹¹

O conjunto de dados existente na base de dados IEETA-CapDB[22] está organizado por exames, e dentro desses directórios estão presentes os relatórios, vídeo original completo e algumas imagens correspondentes aos eventos detectados e anotados usando o software CAPVIEW. Como este software considera como eventos as fronteiras das zonas gástricas, lesões hemorrágicas e eventos adicionados manualmente, é comum que existam imagens normais ou sem relevância clínica no mesmo exame onde foi diagnosticado existência de algum tipo de anomalia ou lesão. Ou seja, cada directório de um exame pode conter imagens normais ou sem relevância no seio de um exame diagnosticado como contendo algum tipo de lesão.

Sendo o conjunto de dados inicial muito numeroso (1900 imagens) era necessário discriminar quais destas pertenciam efectivamente ao grupo de lesões de pólipos, tumores, normais ou sem relevância clínica. Essa anotação para ter relevância científica teria de ser feita por médicos especialistas em cápsula endoscópica. A anotação poderia ser feita de modo manual onde o médico observava cada imagem e depois conforme o seu parecer, movia essa mesma imagem para uma pasta/directório correspondente a uma das classes acima mencionadas (tumor, pólipo, normal, sem relevância). Este processo seria bastante moroso, motivo por qual foi desenvolvido este software de apoio que mostra todas as imagens existentes no conjunto, de forma aleatória e fornece a possibilidade de o médico escolher a que classe aquela pertence além de recolher informação extra necessária para melhor classificar a imagem.

Cada imagem é classificada em relação à sua classe, e caso seja um tumor ou pólipo é

¹¹ A ferramenta de anotação de eventos ECCA (*Endoscopy Capsule Capview cAtaloguer*) foi descrita no artigo “A Tool for endoscopic capsule dataset preparation for clinical Video event detector algorithms”, aceite para apresentação oral e publicação na conferência BIOSTEC-HEALTHINF, patrocinado pelo IEEE-EMB (Engineering in Medicine and Biology) e por WfMC (Workflow Management Coalition), a realizar no Porto, Portugal em Janeiro de 2009.

acrescida informação acerca do seu tamanho, morfologia e cor, além da qualidade da imagem que indica a existência de restos alimentares ou se a imagem está muito desfocada ou escura. Esta informação é útil para eliminar imagens que não contêm informação relevante ou das quais um computador não consegue extrair informação de forma automática.

5.1. Objectivo

A aplicação foi desenvolvida com o intuito de criar um conjunto de dados de referência para os diversos estudos futuros que irão ser efectuados em torno da cápsula endoscópica, assim e como as imagens são analisadas por diversos médicos especialistas o grau de confiança nessa catalogação é bastante alta e no final obtemos um conjunto *ground truth* referência que deve ser usado para estudos de processamento de imagem e vídeo de cápsula endoscópica, muito útil para comparar resultados de diversos algoritmos de *machine vision* que tentem detectar pólipos e/ou tumores.

5.2. Arquitectura

A ferramenta descrita é baseada numa arquitectura de base dados devido à necessidade de armazenar informação acerca dos diversos eventos e correspondentes imagens objectivando a criação de um catálogo de eventos endogástricos anotados nos pressupostos acima descritos, isto é identificar o tipo e morfologia da lesão.

A arquitectura está descrita na Figura 19:

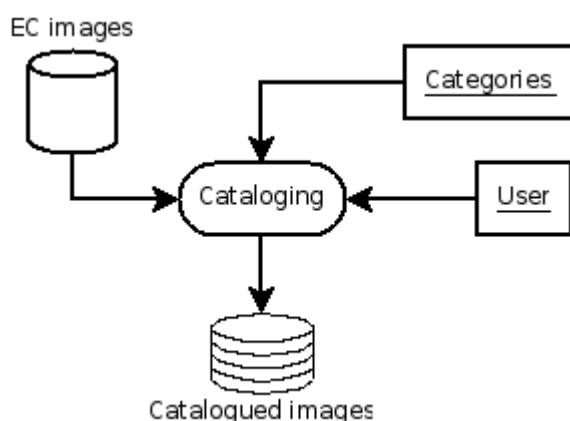


Figura 19: Arquitectura simplificada do software ECCA

Este arquitectura demonstra os eventos extraídos da cápsula endoscópica (IEETA-CapDB), que são o objecto a ser anotado, é necessário um conjunto pré-definido de categorias possíveis das quais o médico necessita de interligar com as imagens ou eventos a serem anotados. O objectivo é

ficar com um conjunto de imagens catalogadas com suporte nas categorias definidas. A aplicação contém uma fina camada de segurança onde os utilizadores (médicos/clínicos) têm um par *username* e *password*, com o objectivo de distinguir o catálogo gerado por cada um dos utilizadores, e para impossibilitar a visualização dos catálogos de outros utilizadores por forma a não obter resultados viciados.

5.3. Workflow

O utilizador acede ao software e após introduzir as suas credenciais é-lhe fornecido o painel principal onde aparecem as imagens.

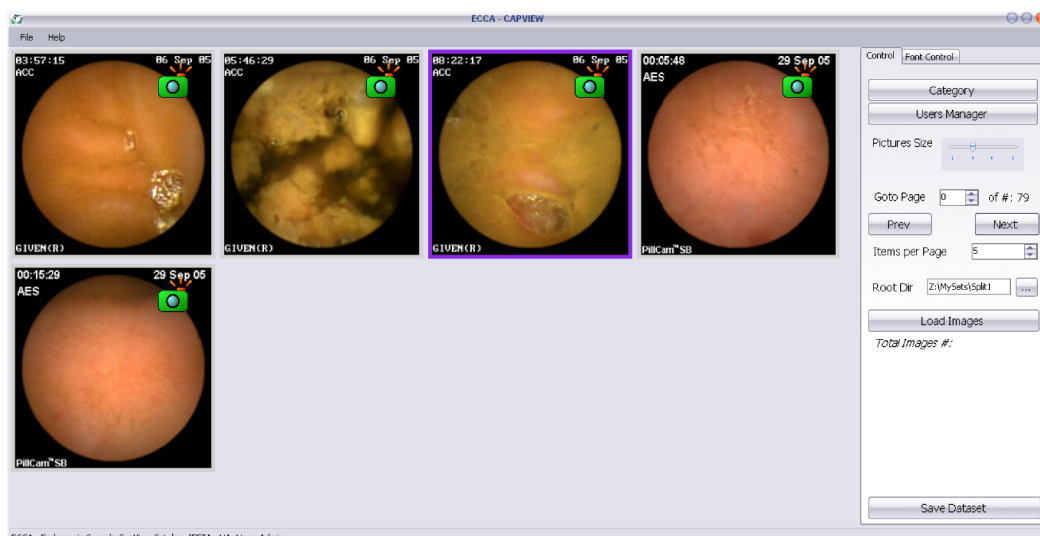


Figura 20: Janela principal do software

A Figura 20 mostra uma janela com 5 imagens, que constitui uma página e um painel lateral que fornece diversas opções de navegação. O utilizador pode escolher o número de imagens contidas na página visível na janela principal, possibilita alterar o tamanho da resolução de cada imagem visível e possibilita navegar para uma página específica. Nesse painel estão também presentes algumas funções de administração, onde é fazer a gestão dos utilizadores e a gestão das categorias que serão utilizadas para a catalogação das diversas imagens e eventos.

Ao utilizador é pedido que seleccione a imagem a catalogar o que leva ao aparecimento da janela de catalogação (Figura 21).

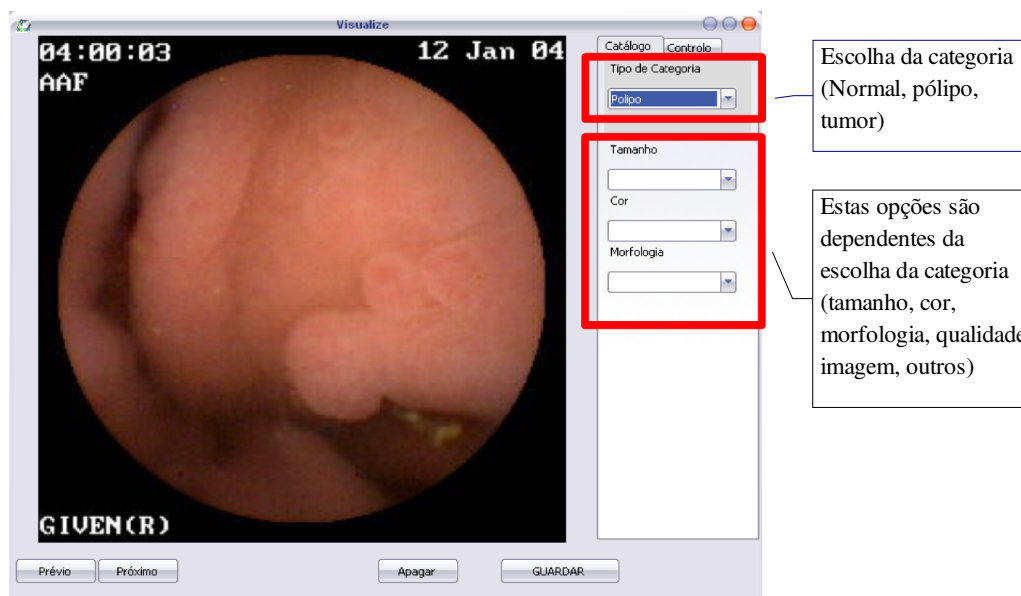


Figura 21: Janela de catalogação

Esta janela possibilita ao utilizador, após observação/análise da imagem, escolher o tipo de categorias a que pertence (normal, pólipo, tumor ou outra), e após essa selecção é-lhe pedido que caracterize melhor o evento. Para caracterizar mais pormenorizadamente o evento tem de escolher as opções que, derivadas do tipo de categoria, aparecem após seleccionar uma destas. No final desta caracterização e após carregar clicar no botão de *guardar*, os dados são guardados numa base de dados (implementada num ficheiro XML pela sua simplicidade). De seguida aparece automaticamente a próxima imagem para ser catalogada, existem dois botões que facilitam a navegação entre as diversas imagens.

Após o utilizador anotar todas as imagens existentes, a base de dados contém um catálogo, associado a esse utilizador, com informação discriminada de cada evento, principalmente com detalhes úteis acerca de tamanho, morfologia e cor nos casos de tumores e pólipos.

Quando diversos médicos especialistas anotarem a totalidade das imagens, e existirem diversos catálogos, utiliza-se a mesma aplicação para realizar uma sobreposição das anotações referentes a estes esses catálogos e extraí-se todos os eventos que sejam anotados de igual modo por todos os especialistas. Esse conjunto de eventos é denominado de *gold standard*, e é escrito num ficheiro de texto (*csv*¹²) para ser passível de ser utilizado noutras aplicações.

5.4. Aplicações do software

Esta ferramenta foi desenvolvida como intuito de ser utilizada como gestor de conjuntos de dados, mas possui também uma vertente pedagógica podendo ser usada como ferramenta de ensino.

¹² CSV – *Comma Separated Value*

5.4.1. Gestor conjuntos

Como foi descrito acima, o software foi desenvolvido para criar um conjunto de dados bem anotados e com alta confiança científica, obtidos a partir da base de dados IEETA-CapDB. As imagens contidas nesta base de dados são maioritariamente de baixa resolução (256x256), existindo algumas de 512x512, devido à utilização das novas cápsulas que contêm uma câmara com maior resolução.

Alguns estudos prévios utilizam imagens de alta resolução para detectar eventos em vídeo de colonoscopia [16,67], por exemplo Kakanis et al.[16], utiliza imagens digitalizadas com resolução de 1K x 1K, que é 16 vezes superior à de 256x256 ou 4 vezes superior à de 512x512, resoluções extraídas da cápsula endoscópica. Esta disparidade leva a que seja necessário compensar com um conjunto elevado de eventos, bem anotados, por forma a que os algoritmos de aprendizagem automática consigam obter resultados validados de forma exaustiva.

O método de criação do conjunto inicial de dados (imagens e vídeos) foi obtido, como já foi dito acima, da base de dados IEETA-CapDB. Os dados existentes nesta base de dados foram submetidos de forma anónima e gratuita, por parte de diversos especialistas nacionais que utilizam o software CAPVIEW como ferramenta de diagnóstico clínico. O CAPVIEW gera pequenos excertos de vídeo a partir dos eventos clínicos especificados pelos especialistas aquando da análise dos exames. Esses pequenos excertos e respectivas imagens são o objecto inicial que vai gerar o conjunto de dados de entrada para a ferramenta ECCA.

O conjunto inicial de 1900 eventos (pólipos, tumores, e normais) obtidos a partir da IEETA-CapDB, foi classificado por dois especialistas residentes no hospital geral de Santo António (HGSA) do Porto, seguindo o *workflow* acima descrito. No final foram filtrados extraindo apenas os resultados validados e designados por *gold standard*. Para um evento ser classificado como *gold standard* tem de ter uma concordância de pelo menos 2 especialistas, e a sua classificação tem de pertencer a uma das seguintes classes: 1) Normal, 2) Pólipo ou 3) Tumor. Esta triagem tem o propósito de eliminar todas as imagens que não tenham conteúdo útil, por exemplo imagens com restos alimentares, ou com muito má qualidade.

5.4.2. Ferramenta pedagógica

A utilização desta ferramenta a nível pedagógico necessita que exista previamente um catálogo *gold standard*, para ser usado como referência. Assim é possível avaliar o desempenho de uma pessoa para avaliar exames de cápsula endoscópica através da metodologia expressa na Figura 22.

A Figura 22 demonstra a necessidade da base de dados *gold standard*, pois a pessoa a ser

avaliada (considerada como o aluno) vai anotar os eventos descritos nessa base de dados. Assim de forma sequencial o aluno vai anotando as imagens, a nível do tipo de lesão e suas sub-categorias (cor, morfologia, tamanho). Após anotar todas as imagens é feita uma avaliação baseada na comparação com as descrições existentes na base de dados *gold standard*.

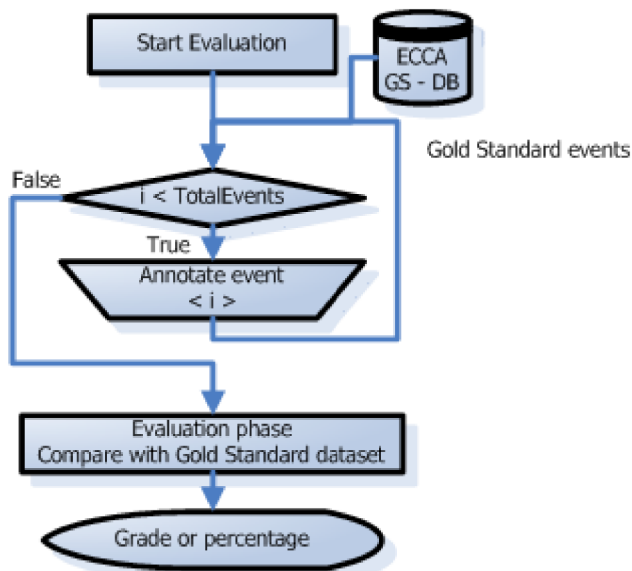


Figura 22: Metodologia de avaliação

A avaliação é realizada com suporte na seguinte equação:

$$\text{AVALIAÇÃO} = 60\% \times \text{Tipo} + 40\% \times \frac{\text{total características correctas}}{\text{total características do gold standard}}$$

Equação 14: Método de avaliação

esta avaliação dá um maior peso (60%) ao tipo de lesão, sendo os restantes 40% distribuídos equitativamente pelas características do respectivo tipo de lesão. Tendo em conta que:

- tipo de lesão pode ser uma de entre normal, tumor ou pólipo
- características podem ser (dependendo do tipo de lesão),
 - tumor: tamanho, cor e morfologia
 - pólipo: tamanho, cor e morfologia
 - normal: não tem características

Existem outras categorias cuja utilidade limita-se a identificar a qualidade das imagens relativa à existência de restos alimentares, a qualidade da imagem e outros. Esta categoria *outros* indica que a imagem não pertence a nenhum dos grupos de pólipos, tumores ou normal. A qualidade de imagem é avaliada usando termos qualitativos como *bom*, *mau* ou *moderado*. No final obtêm-se uma classificação geral resultado da média da avaliação de todos os eventos.

Se fizermos uma avaliação por grupos estratificados em especialistas em EC, especialistas em endoscopia comum, e internos sem experiência em endoscopia, é possível discriminar a capacidade de cada um dos grupos de realizar diagnóstico usando exames de cápsula endoscópica. Para realizar esta avaliação é necessário seguir a metodologia descrita em Figura 23:

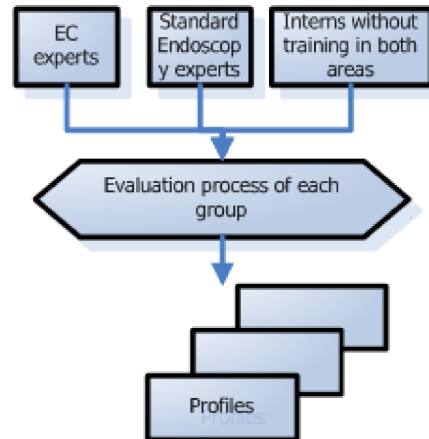


Figura 23: Avaliação estratificada

Para realizar esta avaliação é necessário um conjunto de sujeitos pertencentes a cada um dos grupos para que voluntariamente queiram despende algum tempo a anotar eventos pertencentes à classe *gold standard* para depois obter uma média estratificada, demonstrando a capacidade de cada grupo realizar avaliação usando cápsula endoscópica.

5.5. Resultados

O ficheiro de texto (*csv*) que contém os dados relativos ao conjunto *gold standard*, tem o seguinte formato:

(imagem), (tipo), (GoldStandard), (característica), (valor), {(característica),(valor)}

A primeira coluna contém o nome e caminho da imagem de forma a ser inequívoca, a segunda indica o tipo de evento (tumor, pólipo ou normal), a terceira é um valor binário e indica se a imagem pertence ao grupo de *gold standard*, os campos seguintes são apresentados em pares e cada um destes indica a característica e o valor correspondente associada a cada imagem (por exemplo: {tamanho, 'grande'}). Todas as possibilidades de características e respectivos valores são apresentados no ficheiro *characteristics.xml* (Apêndice II).

Actualmente este software contém um total de 1900 eventos, que estão a ser anotados pelos parceiros do hospital Geral de Santo António. Até a escrita deste documento dois especialistas de cápsula endoscópica e dois médicos internos utilizaram esta ferramenta totalizando um total de 4700 anotações. Destas 900 foram anotadas por 3 utilizadores, enquanto que as restantes 1000 foram anotadas duas vezes. Como todas as imagens foram anotadas pelo menos uma vez, o grau de

confiança nos resultados é bastante bom.

O conjunto de *gold standards* resultante contém 1425 eventos. As restantes 475 imagens foram descartadas por não serem consensuais, por não possuírem boa qualidade de imagem ou por não pertencerem a nenhuma das classes em estudo (normal, pólipos, tumores).

6. Detecção de pólipos em cápsula endoscópica

6.1. Conjuntos de Dados

O conjunto de dados foi construído baseando-se na base de dados *IEETA-CapDB*[22] e a primeira triagem foi realizada a partir dos sumários dos diagnósticos dos diversos exames sendo daí extraído os identificadores para os exames que tinham sido diagnosticados como contendo pólipos, tumor ou normal. Destes foram extraídas as imagens e vídeos correspondentes e colocados em directórios distintos, um para cada classe de diagnóstico e dentro desses directórios foram separados em directórios de imagens e vídeos. As imagens necessitaram de validação clínica/médica e para esse fim foi desenvolvido o software ECCA (descrito anteriormente), cujo objectivo é obter um conjunto bem anotado e com grande confiança científica. O uso do ECCA para este fim revela-se excelente.

Após a anotação por parte dos médicos especialistas e internos, o conjunto inicial continha uma totalidade de 1900 imagens, destas 26% foram descartadas por não pertencerem ao conjunto *gold standard*, resultando um total viável de 1425 imagens, sendo que destas 7% pertenciam à classe de tumores, 25% à classe de pólipos e os restantes 68% pertenciam à classe normal, daqui é fácil de verificar a falta de uniformidade das classes.

	Normais		Pólipos		Tumores	
Total de imagens	967		357		101	
Resolução das imagens	256x256	512x512	256x256	512x512	256x256	512x512
	730	237	270	87	77	24

Tabela 6.1: Conjunto de dados e proporções das classes

A subdivisão dos dados foi distinta nas duas metodologias, como se pode verificar na tabela abaixo.

	Metodologia 1		Metodologia 2	
	Teste	Treino	Teste	Treino
Normais	647	320	647	320
Pólipos	240	117	307	151
Tumores	67	34		
TOTAL	954	471	954	471

Tabela 6.2: Distribuição das imagens nas duas metodologias em estudo

Para a primeira metodologia os dados foram divididos de forma aleatória formando conjuntos de 67% e outro de 33% da totalidade dos dados com o objectivo de treinar e testar o método.

Na segunda onde o objectivo era detectar eventos, sejam eles pólipos ou tumores essas duas classes foram concatenadas e criados conjuntos disjuntos de 67% e 33% para treino e teste respectivamente.

6.2. Metodologia adoptada

No decorrer deste trabalho segui duas abordagens distintas. Na primeira utilizei a informação da imagem completa para extrair as características de textura necessárias para a utilização nos algoritmos de *machine learning*. Na segunda dividi a imagem em sub blocos e tratei cada um deles independentemente, criando o vector de características direccionado ao sub-bloco e não à imagem completa.

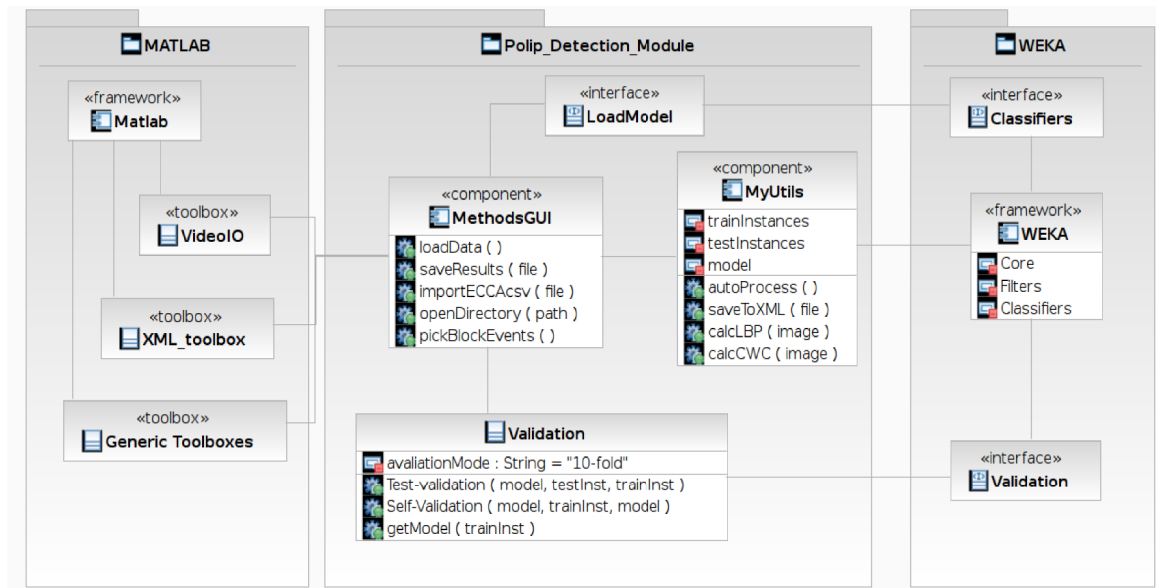


Figura 24: Arquitectura de componentes do sistema de software utilizado.

A Figura 24 é representativa do conjunto de tecnologias utilizadas e os módulos desenvolvidos para atingir os resultados deste trabalho. Os métodos e atributos presentes nos componentes **MyUtils** e **MethodsGUI** são o cerne do módulo de detecção de pólipos, pois são esses que possibilitam o cálculo dos vectores de características (*calcLBP* - Figura 28 e *calcCWC* - Figura 26), entrada e saída e dados (*loadData*, *saveToXML*), importar dados obtidos do software ECCA (*importECCAcsv*), e interfaces gráficas (*pickBlockEvents* - Figura 30 e *autoProcess*).

Na totalidade do trabalho utilizei o algoritmo SMO-SVM disponível no software WEKA¹³ como classificador. Esta escolha foi ponderada devido à alta escalabilidade deste tipo de algoritmos que consegue tratar vectores com elevado número de características de textura.

O software WEKA foi utilizado por ser uma ferramenta bastante utilizada na comunidade científica quando é necessário recorrer a algoritmos de *data mining*, pois contém implementações de diversos algoritmos, modos de avaliação, é bastante parametrizável e contém uma interface gráfica que facilita a utilização e aprendizagem, fornece também uma API¹⁴ que pode ser usada para desenhar e incorporar módulos em aplicações *ad hoc* específicas de forma simples e transparente.

O MATLAB foi a ferramenta de apoio utilizada devido à simplicidade de desenvolvimento de métodos e rotinas de processamento além de esta ferramenta já contemplar um leque considerável de pacotes (*packages*) para realizar processamento de sinal e imagens, foi tido em conta também a facilidade de integrar código JAVA / WEKA nas rotinas desenhadas ao longo do trabalho.

13 WEKA: <http://www.cs.waikato.ac.nz/ml/weka/>

14 API: Application *Programming Interface*

6.2.1. Metodologia 1

A primeira metodologia adoptada está demonstrada na Figura 25, onde é possível observar que o vector de características é construído baseando-se apenas na região de interesse (ROI) obtida a partir da imagem inicial.

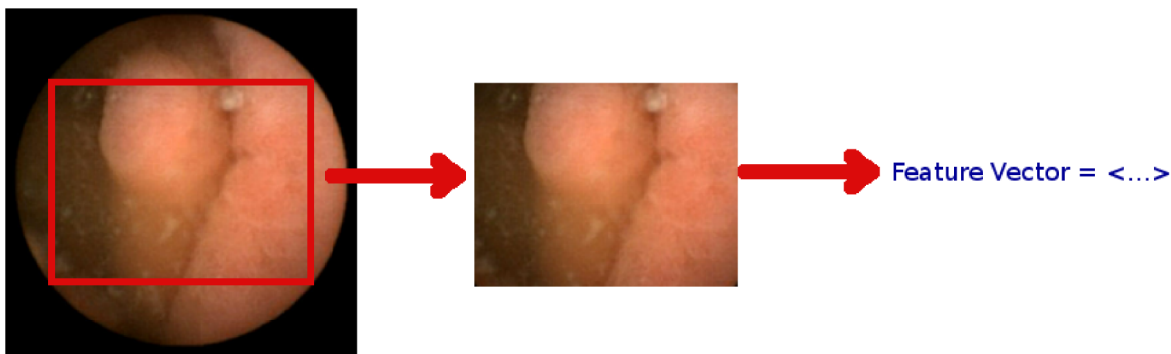


Figura 25: Metodologia 1

A região ROI é utilizada de forma a eliminar a influência que a máscara circular negra tem no processamento. Esta ROI (assumindo um quadrado interior à máscara) tem arestas com 70% do tamanho da imagem inicial o que leva a que a área total da ROI é aproximadamente 49% da imagem inicial, e cerca de 36% de informação visual é desperdiçada.

$$\left\{ \begin{array}{l} \frac{\text{Aresta ROI}}{\text{Aresta Imagem Inicial}} = \frac{1}{\sqrt{2}} = 70\% \\ \frac{\text{Área ROI}}{\text{Área Imagem Inicial}} = 0.7^2 = 49\% \\ 1 - \frac{\text{Área ROI}}{\text{Área Imagem Útil}} = 1 - \frac{2}{\pi} = 36\% \end{array} \right.$$

Equação 15: Área útil de cada imagem do vídeo de endoscopia de cápsula

A garantia que a região ROI contempla o evento de interesse, advém do facto da existência de algumas frames consecutivas do mesmo evento, então devido à dinâmica do video o evento de interesse é contido na ROI.

Após extrair a ROI é necessário aplicar o algoritmo em estudo, neste caso é a metodologia apresentada por Karkanis [16] apresentada atrás como CWC (*Color Wavelet Covariance*) e que pode ser simplificado pelo diagrama da Figura 26.

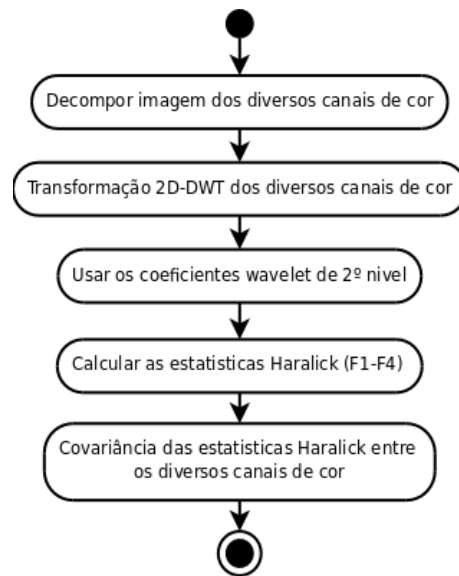


Figura 26: Algoritmo de Karkanis et al.

Após calcular as co-variâncias das características dos diversos canais de cor obtém-se um vector com as 72 componentes e no caso das imagens de teste (supervisionadas) concatena-se o valor da classe correspondente, 0 ou 1 no caso de classes binárias. O modo de criar os vectores através da concatenação da classe é dependente do software WEKA que utiliza o formato de ficheiros ARFF (ver Apêndice I).

O procedimento acima descrito pode ser efectuado de forma sequencial nas directorias que contêm imagens de classes distintas, isto é, existe uma directoria para a classe 1 e outra para a classe 0 e as imagens aí residentes reflectem essa classe. Dessa forma o ficheiro criado utilizando este processo e que contem os vectores de características de cada imagem residente em cada classe, pode ser importado para o software WEKA onde é possível utilizar os algoritmos e classificadores aí existentes para criar um modelo baseado nessas características das diversas imagens. Nesse processo de criação de modelo e avaliação foi utilizado o classificador SMO-SVM e o método de avaliação *10-fold cross validation* [47].

6.2.2. Metodologia 2

A metodologia apresentada na Figura 27, divide a imagem inicial em sub-blocos de tamanho 32 x 32, ou 64 x 64 *pixels*, e extrai o vector de características de cada sub-bloco tendo em atenção se nesse bloco existe algum pólipo ou tumor. Se existir o bloco é catalogado como pertencente à classe 1, e todos os outros pertencem à classe 0.

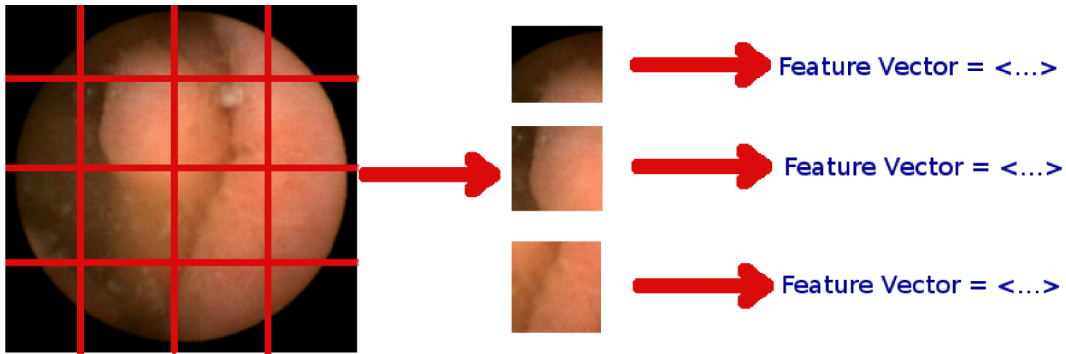


Figura 27: Metodologia 2

Neste método não é aplicado o paradigma da ROI por não ser necessário, pois ao dividir a imagem original em blocos de tamanho mais reduzido e ao realizar o processamento nos blocos individuais consegue-se minimizar o efeito e influência de zonas com informação inútil escassa ou inexistente. Esta minimização é óbvia pois os blocos ao ter um tamanho reduzido a quantidade de informação inútil é menor, e no caso limite de não terem informação útil podem ser facilmente identificados de forma inequívoca, por exemplo os blocos negros. De notar que a imagem completa foi o objecto a ser classificado e na fase de avaliação todos os blocos constituintes da mesma fazem parte apenas de um conjunto, o de treino ou de teste. É garantido que diferentes blocos da mesma imagem não estão presentes nos dois conjuntos na mesma iteração, evitando a possibilidade de *overfitting*.

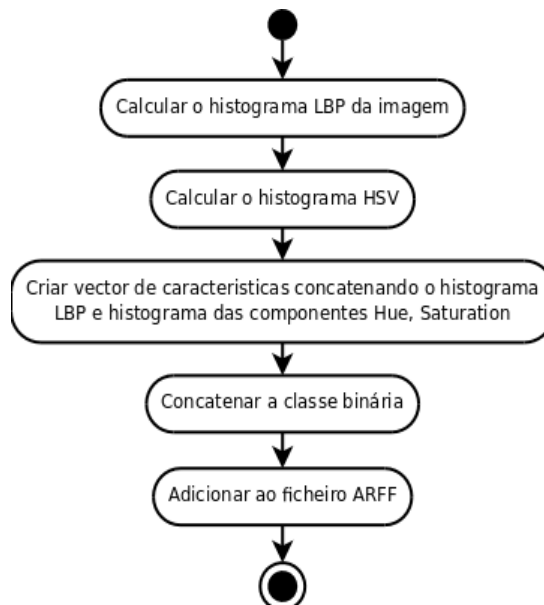


Figura 28: Método usando LBP

A Figura 28 demonstra o algoritmo utilizado neste método. Este algoritmo é aplicado a cada sub-bloco que origina instâncias com informação dos histogramas LBP, HS e a respectiva classe que são adicionados a um ficheiro ARFF, que naturalmente gera mais vectores que o

anterior.

O teste e validação deste método é dividido em duas fases. A primeira é a verificação da eficácia do modelo a qual é realizada aplicando o modelo criado em todos os sub-blocos e verificar a eficácia desse modelo na detecção ou previsão de cada um deles ou seja testar o modelo nos próprios dados, denominado de *self-validation*.

A segunda fase de validação é a aplicação do modelo em dados referentes ao conjunto de teste disjunto do que foi utilizado na criação do modelo. Assim e como as imagens de cada classe estão em directorias diferentes, é simples comparar os resultados do modelo com as classes reais de cada imagem, então de forma sequencial, uma rotina testa cada imagem existente na directoria de uma classe e considera que essa imagem pertence à classe 1 ou 0 da seguinte forma

$$Classe = \begin{cases} 1, & \text{se total blocos de classe 1} \geq 0 \\ 0, & \text{noutros casos} \end{cases}$$

Assim, caso a imagem a ser testada contenha um ou mais blocos que tenham sido classificados como pertencente à classe 1, a classe dessa imagem é também essa. No caso da imagem não conter nenhum bloco identificado como classe 1, então a imagem pertence à classe 0.

Para conseguir comparar e avaliar as diferenças nos resultados dos diversos casos é necessário implementar um método sistemático mantendo a mesma referência base, assim é necessário que os sub-blocos de cada imagem, representativos dos eventos sejam processados pelos diversos modelos de forma semelhante, ou seja garantir que a função de cálculo do vector é o único parâmetro variável, dependendo do caso em estudo.

A solução apresentada assenta na descrição de um ficheiro XML¹⁵, que contém a lista de imagens necessárias para a criação do modelo, e cada item dessa lista contém uma lista dos vértices dos sub-blocos e a que tipo de classe esse sub-bloco pertence.

15 XML: eXtended Markup Language

```
<root>
  <item idx="1">
    <file idx="1"> /path/to/image1</file>
    <event idx="1">
      <point idx="1"> [x,y] </point>
      <type idx="1"> classe </type>
      <point idx="2"> [x,y] </point>
      <type idx="2"> classe </type>
    </event>
  </item>
  <item idx="2">
    <file idx="1"> /path/to/image2 </file>
    <event idx="1">
      <point idx="1"> [x,y] </point>
      <type idx="1"> classe </type>
    </event>
  </item>
</root>
```

Figura 29: Excerto de código XML - descritor

O excerto de código XML na Figura 29 é representativo do descritor usado para os sub-blocos de cada imagem e é construído automaticamente pela biblioteca de MATLAB¹⁶, que é muito mais complexo do que está presente acima, este contém apenas a informação necessária e descritiva para uma percepção simples por parte do leitor.

O ficheiro XML foi criado usando um *script* em MATLAB que apresenta iterativamente as imagens existentes num directório, nas quais o utilizador selecciona os blocos com conteúdo de eventos de pólipos ou tumores para posteriormente ser guardado num ficheiro com a estrutura apresentada acima. Na Figura 30 é possível ver as zonas de eventos (marcadas a vermelho) e as zonas a verde demarcam toda a informação acessória inerente a uma imagem desta natureza.

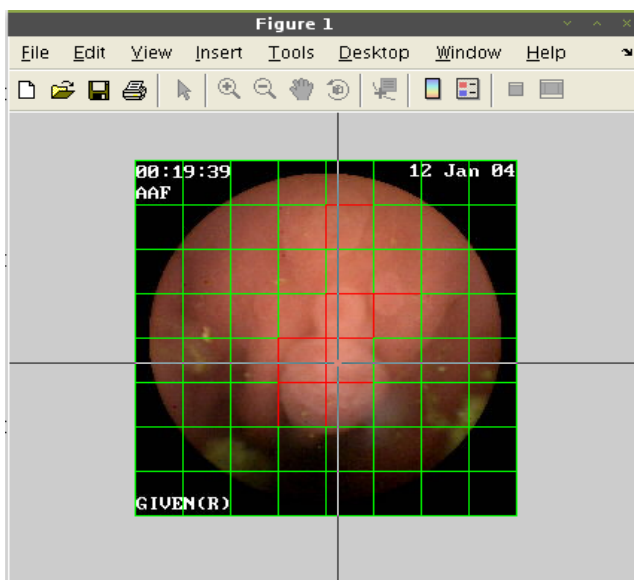


Figura 30: Script para seleccionar blocos de eventos

16 A biblioteca de xml para matlab está disponível em http://www.geodise.org/toolboxes/generic/xml_toolbox.htm

7. Resultados e discussão

Os resultados estão separados nas duas metodologias, e cada uma foi avaliada em duas instâncias: aval-treino e aval-real. A avaliação **aval-treino** que consiste na avaliação do modelo em causa, nos dados responsáveis pela sua criação (self-validation). A avaliação **aval-real** consiste na avaliação do modelo em causa em dados disjuntos dos primeiros, por forma a avaliar o comportamento do modelo em dados desconhecidos. Cada uma das metodologias são constituídas por diversos casos de estudo, que dependem em diversas variáveis como sendo o tipo de imagens utilizadas (maior ou menor resolução), ou a introdução de mais informação nos vectores de características utilizados na criação e validação dos modelos.

Os resultados são apresentados e calculados através das medidas estatísticas seguintes :

$$Eficácia = \frac{TP + TN}{total\ casos}$$

$$Sensibilidade = \frac{TP}{TP + FN} \quad Especificidade = \frac{TN}{TN + FP}$$

$$Observer\ Agreement = \frac{TP + TN}{total\ casos}$$

$$Chance\ Agreement = \frac{(TN + FN) * (TN + FP) + (TP + FP) * (TP + FN)}{total\ casos^2}$$

$$Kappa = \frac{Observer\ Agreement - Chance\ Agreement}{1 - Chance\ Agreement}$$

NOTA: TN = Negativos, TP = Positivos, FN = Falsos Negativos, FP = Falsos Positivos

Equação 16: Estatísticas usadas na apresentação dos resultados.

A sensibilidade mostra a proporção dos casos detectados como positivos no total de casos positivos, este valor convém ser o maior possível indicando que o número total de falsos negativos é baixo [68].

A especificidade mostra a proporção de casos detectados como negativos na totalidade de casos negativos, convém ser o mais alto possível de forma a reflectir a baixa quantidade de falsos positivos [68].

O indicador Kappa quantifica a concordância dos dados, se as predições estão realmente de acordo com a realidade ou se é apenas uma situação do acaso, este indicador é usado em classes nominais, isto é em classes em que os seus valores estão predefinidos e listados à priori. Os valores deste indicador são sempre menores ou igual a 1, sendo que 1 é o máximo de concordância, 0 não existe concordância, e menor que 0 é um caso raro em que a concordância é menor do que a esperada apenas por acaso [69] .

Na Tabela 7.1 são apresentados os resultados obtidos.

Caso		Aval-Treino				Aval-Real			
		Eficácia	Sensib.	Espec.	Kappa	Eficácia	Sensib.	Espec.	Kappa
MÉTODO 1	C1	0.66	0.96	0.11	0.09	0.60	0.49	0.84	0.26
	C1_resampled	0.80	0.52	0.92	0.48	0.73	1.00	0.16	0.21
MÉTODO 2	R1	0.73	0.74	0.73	0.47	0.69	0.27	0.89	0.18
	R2	0.83	0.95	0.38	0.39	0.55	0.48	0.57	0.04
	R3	0.89	0.95	0.63	0.60	0.55	0.48	0.57	0.04
	R4	0.97	1.00	0.13	0.23	0.77	0.23	0.94	0.21
	R10_1	0.75	0.77	0.73	0.50	0.36	0.92	0.19	0.05
	R30_1	0.95	1.00	0.06	0.11	0.71	0.13	0.89	0.03
	R10_2	0.82	0.9	0.50	0.42	0.38	0.88	0.25	0.07
	R30_2	0.79	0.68	0.86	0.56	0.24	1.00	0.00	0.00

Tabela 7.1: Apresentação de resultados.

NOTA:

R1: modelo criado com suporte a todas as imagens contendo pólipos

R2 e R3: modelo criado usando apenas 15 imagens aleatórias contendo pólipos

R4: aplicação do modelo em R3 no teste de imagens apenas com resolução de 512x512

R10: modelo criado usando as mesmas imagens de R3, mas em que o vector de características contém informação de CWC

R30: modelo criado usando as mesmas imagens de R3, mas em que o vector de características contém histograma no espaço de cores Karhunen-Loeve

R10_1 e R30_1: modelo testado na globalidade do conjunto de teste

R10_2 e R30_2: modelo testado apenas nas imagens de 512x512 do conjunto de teste

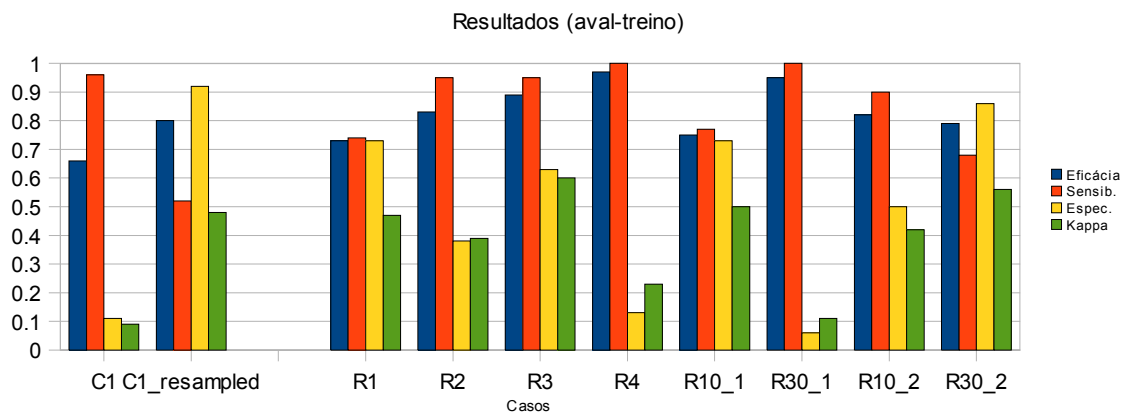


Figura 31: Comparação visual dos resultados aval-treino

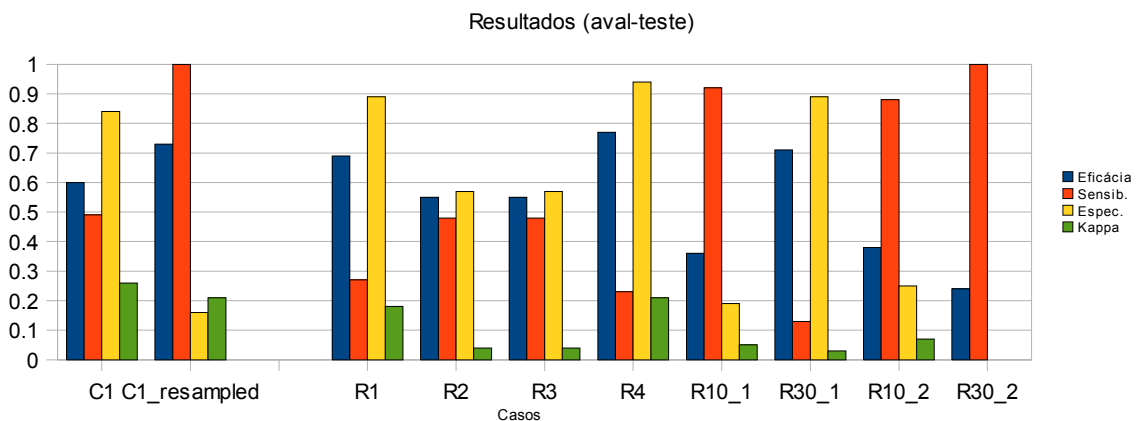


Figura 32: Comparação visual dos resultados aval-teste

As figuras 31 e 32 apresentam os resultados da Tabela 7.1, para ter uma percepção global dos resultados, sendo estes analisados, de seguida, com mais pormenor.

Experiência	Sensibilidade	Especificidade	Positivos	Negativos	Falsos Positivos	Falsos Negativos
Imagem apenas com máscara circular	0.73	0.60	29	24	16	11
Imagem apenas com ROI (50% da área)	0.80	0.60	32	24	16	8

Tabela 7.2: Utilização da ROI versus máscara circular, na metodologia 1.

A Tabela 7.2 demonstra a razão da utilização de uma região de interesse (Figura 25) em detrimento de apenas uma máscara circular no pré-processamento da imagem. A máscara circular consegue utilizar a totalidade da informação de uma imagem, em oposição a região de interesse perde 36% de informação útil, mas consegue obter resultados equiparados na especificidade e ligeiramente superiores na sensibilidade. Nestes rácios só estão contabilizadas 40 imagens aleatórias, mas é indicativo da utilidade da região de interesse. Outra mais valia é a redução no tempo de processamento para a extracção dos vectores de características, pois a área de processamento é 50% inferior à original.

Os resultados obtidos através do método 1 estão aquém do referido na literatura. No caso C1 o melhor resultado da eficácia nas duas avaliações é de 66%, que fica muito distante dos valores de 90-95% obtidos por Karkanis. Observando o índice kappa que está próximo de zero, conclui-se que os resultados de eficácia e precisão são devidos ao acaso, comprovado pelo facto de o modelo classificar as instâncias maioritariamente como positivas, fazendo com que o número de falsos positivos seja elevado e a taxa de especificidade muito reduzida.

Nos mesmos dados de C1 foi aplicado um filtro de pré-processamento existente no software WEKA (de *resampling*), que faz uma amostragem dos dados e remove muitos *outliers*. Esse filtro amostra os dados de forma equiprovável mantendo o número total de instâncias inicial constante. Após esta amostragem a eficácia do modelo melhorou à custa de uma diminuição dos falsos positivos e diminuição da sensibilidade. Estes factores levaram a que o índice kappa tenha aumentado significativamente (de 9% para 48%). Os resultados eram esperados, pois a remoção efectuada tem uma consequente concentração dos dados, fazendo com que o modelo criado se adapte melhor aos dados de treino. Na avaliação em *aval-real*, existiu também uma melhoria de 13%, à custa de um aumento dos falsos positivos que é bastante maior que os positivos detectados. Os resultados desta segunda avaliação demonstram que o modelo está sub-amostrado, ou seja contém pouca informação sobre determinadas instâncias levando a que sejam mal classificadas.

O método 2 foi planeado de forma a verificar a eficácia de diversos modelos variando algumas das variáveis na fase de treino, apenas para analisar a dependência dos dados de entrada na geração de modelos estatísticos para classificação e detecção de eventos em imagens de vídeo de cápsula.

O caso R1 vai servir de referência na comparação com os casos seguintes, para isso o modelo foi gerado a partir de todas as imagens referenciadas como pólipos e tumores, o qual obteve uma eficácia geral de 73%. Os casos R2 e R3 foram criados com suporte a um conjunto de treino com apenas 15 imagens aleatória e testado no conjunto global de teste. No caso R4 foi aplicado o modelo de R3 num conjunto de imagens de teste com resolução de 512x512. Nos testes R10 e R30 foi utilizado o mesmo conjunto de dados de treino que no caso R3 (15 imagens), por apresentar as melhores taxas de eficácia e obter o melhor índice kappa.

Os casos R10_1 e R30_1 utilizaram as mesmas imagens para criar o modelo, apenas diferem no vector de características, que no caso R10_1 é constituído pela concatenação de diversas componentes, sendo estas o histograma obtido do algoritmo LBP, o histograma das componentes *Hue* e *Saturation*, e o vector com as características de CWC.

Nos casos R10_2 e R30_2 foi utilizado o mesmo modelo criado para os casos R10_1 e R30_1, respectivamente mas aplicado e avaliado apenas no conjunto de imagens com tamanho de 512 *pixels*.

A diferença dos casos R2 e R3 foi na escolha das imagens para a criação do modelo, em ambos os casos foram escolhidas 15 imagens de forma aleatória (disjuntas nos 2 casos) do conjunto dos pólipos e tumores. Os resultados obtidos nos dois casos é muito semelhante, com a eficácia a variar entre os 83% e 89%, com sensibilidade de 95%, e com uma diferença mínima na especificidade (ronda os 50%), indicando uma elevada taxa de falsos positivos. Na avaliação *aval-real* obtiveram-se os mesmos resultados, com a sensibilidade e especificidade perto dos 50%, levando o índice kappa para valores mínimos de 4%, indicando a relevância do acaso neste teste.

Os testes acima demonstram a possibilidade de afirmar que as imagens a partir das quais se cria o modelo não é um factor significativo para um melhor desempenho, sendo apenas necessário assinalar correctamente as zonas das imagens que contêm pólipos e/ou outras lesões. A melhoria destes resultados em relação ao caso R1 (em que o modelo foi criado usando todas as imagens de pólipos disponíveis para treino), mostra que um modelo tem melhores taxas de eficácia se não se treinar excessivamente. Facto este que leva a uma agilidade das experiências devido ao menor tempo de treino dos modelos, sendo o único senão a diminuição da taxa de especificidade.

No caso R4 onde foi avaliado o modelo de R3 apenas em imagens com maior resolução (512x512 *pixels*), houve uma detecção correcta de eventos superior à de R3 levando a uma maior eficácia e obteve também uma sensibilidade de 100%. O kappa deste modelo foi de 23% devido à

baixa especificidade. Em *aval-real*, este mesmo modelo obteve a melhor pontuação geral em relação aos outros casos, obtendo uma eficácia de 77%, e uma especificidade de 94%. O valor do índice kappa manteve-se semelhante em *aval-treino*, com 21%, devido à baixa sensibilidade. Este aumento de eficácia em ambos os testes leva considerar que a resolução das imagens é um factor bastante crítico para este tipo de análise. Os resultados dispares entre sensibilidade e especificidade nos dois testes pode ser explicado devido ao reduzido número de imagens de alta resolução, apenas 101 dos conjuntos *pólipos* e *tumores* tornando os rácios e taxas mais instáveis, isto é susceptíveis a alterações bastantes relevantes quando o número de falsos positivos, falsos negativos, positivos e negativos são da mesma ordem de grandeza.

Em relação aos casos R10 e R30, as avaliações *aval-treino* são inconclusivas, pois o intuito destes casos era avaliar a relevância da informação obtida através do histograma no espaço K-L e das características usando o método CWC. Não é possível afirmar que o histograma em K-L acrescenta informação útil, pois apesar de obter uma taxa de eficácia de 95% e não ter encontrado nenhum falso negativo tem uma quantidade de falsos positivos bastante alta (Especificidade ~ 0%), fazendo com que o índice kappa esteja apenas nos 11% com uma eficácia tão elevada.

Os outros casos R10_1, R10_2 e R30_2 obtiveram taxas de eficácia entre 75% e 82%, mas com variações de especificidade e sensibilidade de 36% e 22% respectivamente. Esta variabilidade demonstra uma grande instabilidade nos resultados.

A avaliação nas imagens em *aval-real* reitera a mesma instabilidade existente em *aval-treino*, obtendo índices kappa abaixo do 7%, e taxas de eficácia menores que 38%. A excepção do caso R30_1 em que a eficácia é de 71% mas com uma sensibilidade muito baixa (apenas 13%).

No conjunto de todos os casos testados, os modelos mais promissores e com melhor eficácia em *aval-real*, isto é os modelos que se comportaram melhor na presença de imagens distintas das que tinham sido treinados e apresentados por ordem decrescente de eficácia foram os R4, R30_1 e R1, mas todos eles obtiveram taxas de sensibilidade reduzidas, entre 13% e 27%, impossibilitando qualquer tentativa de incorporar estas metodologias em casos de diagnóstico automático.

O caso C1_resampled obteve uma taxa de 73%, equiparada com as dos casos descritos acima, têm uma taxa de sensibilidade alta (100%), mas com uma especificidade mínima (16%), fazendo com que o índice kappa esteja nos 21%. Este caso obteve esses resultados porque considerou quase todos os eventos como positivos, isto é não obteve nenhum falso negativo (sensibilidade 100%) e a especificidade de 16% indica uma quantidade de falsos positivos 5.25 vezes superior aos negativos detectados.

8. Conclusão

Do software ECCA utilizado para discriminar os dados iniciais posso concluir que a sua implementação foi de extrema importância e utilidade. A necessidade de anotar e catalogar o elevado número de eventos dispostos à partida, e ao mesmo tempo ter um alto grau de confiança nessas anotações obrigava a que um conjunto de especialistas em cápsula endoscópica dispusessem algum do seu tempo nessa anotação, o que seria impossível de conseguir se estes tivessem de realizar esse trabalho de forma manual. De acrescentar que comentários por parte desses médicos/clínicos indicam que esta ferramenta é bastante fácil e intuitiva na sua utilização, tendo como único revés o tempo despendido na anotação (mais de 4 horas, factor proporcional ao número total de instâncias a anotar).

As únicas ferramentas semelhantes ao ECCA são os atlas endogástricos, que contém grandes volumes de informação acerca de doenças e seus tratamentos que estão disponíveis a todos os médicos, enquanto que o ECCA agrega uma vasta quantidade de informação visual (pólipos e tumores) para depois estar disponível para ensinar máquinas e algoritmos de aprendizagem automática.

O conjunto de dados criado por este software irá ser utilizado em futuros estudos, e sendo uma ferramenta complementar ao CAPVIEW está disponível na página de internet deste projecto.

A aplicação e utilidade deste software como ferramenta pedagógica ainda não foi possível de comprovar devido à falta de alunos que queiram dispendir algum do seu tempo a utilizar e experimentar esta aplicação. O mesmo facto se aplica na resposta à questão de quantificar a capacidade de diagnóstico usando a cápsula endoscópica por parte de especialistas em endoscopia normal.

O segundo objectivo deste trabalho era o desenvolvimento de uma ferramenta automática

de ajuda ao diagnóstico, principalmente na detecção de pólipos. Após a revisão da literatura adequada decidiu-se avançar pelas metodologias apresentadas por Karkanis et al.[16], por apresentarem resultados fortes na detecção de pólipos em vídeo endoscópico.

Os resultados apresentados neste estudo demonstram que a implementação dessa técnica não é viável para a avaliação de vídeo extraído da cápsula endoscópica. Na comparação com os dados apresentados por Karkanis et al.[16], existe uma discrepância enorme na eficácia das técnicas, uma vez que na literatura são apresentados resultados com sensibilidades e especificidades de 99% e 93%, respectivamente enquanto que este estudo mostra que os casos com as melhores taxas de sensibilidade e especificidade são o caso C1_resampled que obteve 92% e 52%, e o caso R3_1 com 63% e 95%. Com os resultados actuais ainda não é possível criar um algoritmo com eficácia suficiente para ser utilizado em diagnóstico clínico, mas os resultados indicam que, mesmo aquém do esperado e referido na literatura, existe alguma potencialidade nestas abordagens, necessitando de melhorias nos trabalhos futuros.

Na análise comparativa entre a metodologia 1 e a análise descrita na literatura por Karkanis et al.[16], são visíveis algumas variáveis distintas, nomeadamente o tipo de dados a processar e o modo de aquisição e digitalização dos mesmos. Na literatura utilizou-se vídeo de colonoscopia enquanto que na metodologia 1 é usado vídeo de cápsula endoscópica. O vídeo de cápsula contém uma resolução de 256x256 *pixels* brutos onde apenas 50% dessa área é utilizada, enquanto a digitalização da colonoscopia foi feita com uma resolução de 1K x 1K *pixels*. Na apresentação de resultados já foi descrito que a utilização de um pequeno conjunto de treino tinha no geral melhores resultados do que quando o modelo era treinado com um conjunto com cardinalidade superior, sendo esta ideia reiterada na literatura referenciada, demonstrando que a resolução de aquisição das imagens pode ser um factor muito importante na tarefa de classificação dos eventos.

A utilização do histograma do espaço de cores Kaurhen-Loeve (K-L), casos R30_1 e R30_2 não demonstraram melhorias em relação ao caso R3, e como não foi testado nenhum caso apenas com o histograma em K-L não é possível afirmar que este espaço de cores é melhor que o HSV, mas pode concluir-se que não traz informação adicional relevante para uma melhor classificação dos eventos. Nos casos R10_1 e R10_2 aos quais foi adicionada informação obtida através de CWC aos vectores de características, não é visível nenhuma melhoria em relação à utilização exclusiva da informação obtida através dos algoritmos de LBP. Daqui conclui-se que a utilização apenas de algoritmos de LBP é suficiente e computacionalmente menos intensiva na tarefa de classificar os eventos desejados.

Da análise dos resultados e da verificação do conjunto de imagens é possível afirmar que, um factor de grande influência são as imagens utilizadas e a sua catalogação por parte dos médicos. Devido à perícia e minúcia por parte dos médicos, que catalogaram as imagens usando o software ECCA, é verificável a existência de imagens catalogadas como anormais (com lesão) que devido à

natureza da lesão e à posição da câmara é difícil uma identificação inequívoca por parte do algoritmo de detecção automática.

A incoerência geral dos resultados permite concluir que a análise deste tipo de exames clínicos é bastante mais complexa do que se esperava inicialmente. A natureza dos próprios exames e as variáveis a eles associadas são diversas, nomeadamente a localização topográfica do pólipo, morfologia do evento, posição da cápsula e existência de artefactos no vídeo tornam o processo de uniformização, de análise e de classificação mais difícil.

8.1. Trabalho Futuro

O desenvolvimento deste trabalho demonstra que a utilização de características obtidas a partir de *wavelets* é inadequada para a detecção de eventos em vídeo de cápsula endoscópica. Assim sendo este trabalho suscita uma série de questões que poderão servir como linhas orientadoras de pesquisas futuras, no sentido de aprofundar este problema e apostar nas características de cor, posição topográfica e espacial e outros espaços de cor para conseguir uma descrição mais uniforme destes eventos.

A posição topográfica do evento que indica em que zona do sistema gástrico se encontra a cápsula pode ser informação útil e relevante para adicionar ao vector de características, e dessa forma deve ser estudado em mais pormenor em próximos trabalhos.

Pelo que foi demonstrado acima a resolução das imagens pode ser um factor importante, e dessa forma é necessário adquirir um conjunto mais alargado de imagens com resolução elevada para conseguir discriminar a importância desse factor nos resultados e classificações gerais dos métodos descritos acima ou outros que possam vir a ser desenvolvidos.

No seguimento do estudo da utilidade do software ECCA como ferramenta de ensino é necessário facilitar a utilização desta por parte de um conjunto alargado de estudantes em medicina e avaliar essa potencialidade. Outra questão que necessita e pode ser respondida através da utilização desta ferramenta é avaliar a capacidade de diagnóstico usando vídeo de cápsula endoscópica por parte de especialistas em endoscopia comum.

9. Bibliografia

- [1] B. Philip, "Lichtleiter, eine Erfindung zur Anschauung innerer Teiler and Krankheiten [Abstract]," *Journal der Practischen Arzneykunde and Wundarzneykunst*, vol. 24, 1806, pp. 107-124.
- [2] M. Classen, A. Axon, J. Baillie, H. Barr, K. Binmoeller, A. Bitoun, Y. Bouhnik, L. Brandt, A. Burroughs, L. Buscail, O. Cass, S.C.S. Chung, J. Colombel, G. Costamagna, and P. Cotton, *Gastroenterological Endoscopy*, Georg Thieme Verlag, 2002.
- [3] Xing Sun, Tianbai Qian, Minglu Li, and Lixu Gu, "Interactive virtual colonoscopy system based on CT volume images," *Communications, Circuits and Systems, 2005. Proceedings. 2005 International Conference on*, 2005, p. 915.
- [4] A.G. Schreyer and S. Feuerbach, "Virtual endoscopy using MRI," *Minimally invasive therapy & allied technologies : MITAT : official journal of the Society for Minimally Invasive Therapy*, vol. 12, Sep. 2003, pp. 217-21.
- [5] M. Harders, S. Wildermuth, D. Weishaupt, and G. Székely, "Improving Virtual Endoscopy for the Intestinal Tract," *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*, 2002, pp. 20-27.
- [6] A. Fritscher-Ravens and C. Swain, "The Wireless Capsule: New Light in the Darkness," *Digestive Diseases*, vol. 20, 2002, pp. 127-133.
- [7] L.B. Gerson and J. Van Dam, "Wireless Capsule Endoscopy and Double-Balloon Enteroscopy for the Diagnosis of Obscure Gastrointestinal Bleeding," *Techniques in Vascular and Interventional Radiology*, vol. 7, Sep. 2004, pp. 130-135.
- [8] S. Johanssen, M. Boivin, H. Lochs, and W. Voderholzer, "The yield of wireless capsule endoscopy in the detection of neuroendocrine tumors in comparison with CT enteroclysis," *Gastrointestinal Endoscopy*, vol. 63, Apr. 2006, pp. 660-665.
- [9] O. Haluszka, "Small-bowel endoscopy," *Clinical Update*, vol. 15, Apr. 2008, pp. 1-4.
- [10] G.M. Cobrin, R.H. Pittman, and B.S. Lewis, "Increased diagnostic yield of small bowel tumors with capsule endoscopy," *Cancer*, vol. 107, 2006, pp. 22-27.
- [11] B.S. Lewis, G.M. Eisen, and S. Friedman, "A pooled analysis to evaluate results of capsule endoscopy trials," *Endoscopy*, vol. 37, Oct. 2005, pp. 960-5.

- [12] M. Mylonaki, A. Fritscher-Ravens, and P. Swain, “Wireless capsule endoscopy: a comparison with push enteroscopy in patients with gastroscopy and colonoscopy negative gastrointestinal bleeding,” *Gut*, vol. 52, Aug. 2003, pp. 1122–1126.
- [13] J. Cunha, M. Coimbra, P. Campos, and J. Soares, “Automated Topographic Segmentation and Transit Time Estimation in Endoscopic Capsule Exams,” *Medical Imaging, IEEE Transactions on*, vol. 27, 2008, pp. 19-27.
- [14] G.S. Raju, L. Gerson, A. Das, and B. Lewis, “American Gastroenterological Association (AGA) Institute Technical Review on Obscure Gastrointestinal Bleeding,” *Gastroenterology*, vol. 133, Nov. 2007, pp. 1697-1717.
- [15] G. Iddan, G. Meron, A. Glukhovskiy, and P. Swain, “Wireless capsule endoscopy,” *Nature*, vol. 405, May. 2000, p. 417.
- [16] S. Karkanis, D. Iakovidis, D. Maroulis, D. Karras, and M. Tzivras, “Computer-aided tumor detection in endoscopic video using color wavelet features,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, 2003, pp. 141-152.
- [17] André Sousa, “Analysis of colour and texture features of vital-stained magnification-endoscopy images for computer endoscopy computer-assisted diagnosis of precancerous and cancer lesions,” Master Thesis, Faculdade de Engenharia da Universidade do Porto, 2008.
- [18] G.D. Meron, “The development of the swallowable video capsule (M2A),” *Gastrointestinal Endoscopy*, vol. 52, Dec. 2000, pp. 817-9.
- [19] D. Mishkin, R. Chuttani, J. Croffie, J. DiSario, J. Liu, R. Shah, L. Somogyi, W. Tierney, L. Kee Song, and B. Petersen, “ASGE Technology Status Evaluation Report: wireless capsule endoscopy,” *Gastrointestinal Endoscopy*, vol. 64, 2006, pp. 536-545.
- [20] M. Coimbra, P. Campos, and J. Cunha, “Topographic Segmentation and Transit Time Estimation for Endoscopic Capsule Exams,” *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 2006, p. II.
- [21] J.A. Erber, “Wireless capsule endoscopy: where and how to learn?,” *Gastrointestinal Endoscopy*, vol. 68, Jul. 2008, pp. 115-117.
- [22] J.P. Cunha, M. Coimbra, and J. Soares, “Capview: computer-aided diagnosis for endoscopic capsule,” *Revista do DETUA*, vol. 4, Apr. 2007, pp. 809-812.
- [23] S.A. van Tuyl, J.T. van Noorden, R. Timmer, M.F. Stolk, E.J. Kuipers, and B.G. Taal, “Detection of small-bowel neuroendocrine tumors by video capsule endoscopy,” *Gastrointestinal Endoscopy*, vol. 64, Jul. 2006, pp. 66-72.
- [24] A. Caunedo, M. Rodríguez-Téllez, J.M. García-Montes, B.J. Gómez-Rodríguez, J. Guerrero, J.M. Herrerías, F. Pellicer, and J.M. Herrerías, “Usefulness of capsule endoscopy in patients with suspected small bowel disease,” *Revista Española De Enfermedades Digestivas: Organó Oficial De La Sociedad Española De Patología Digestiva*, vol. 96, Jan. 2004, pp. 10-21.
- [25] Ning Dai, Christoph Gubler, Peter Hengstler, Christa Meyenberger, and Peter Bauerfeind, “Improved capsule endoscopy after bowel preparation,” *Gastrointestinal endoscopy*, vol. 61, Jan. 2005, pp. 28-31.
- [26] M. Coimbra, P. Campos, and J. Cunha, “Extracting clinical information from endoscopic capsule exams using MPEG-7 visual descriptors,” *Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. The 2nd European Workshop on the (Ref. No.*

2005/11099), 2005, pp. 105-110.

- [27] C. Vertan and N. Boujemaa, "Color texture classification by normalized color space representation," *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, pp. 580-583 vol.3.
- [28] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing: Analysis and Machine Vision*, Cengage-Engineering, 1998.
- [29] R.C. Gonzalez, R.E. Woods, and S.L. Eddins, *Digital Image Processing Using MATLAB*, Prentice Hall, 2003.
- [30] C. Li and P. Yuen, "Regularized color clustering in medical image database," *Medical Imaging, IEEE Transactions on*, vol. 19, 2000, pp. 1150-1155.
- [31] M. Tkalcic and J. Tasic, "Colour spaces: perceptual, historical and applicational background," *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, 2003, pp. 304-308 vol.1.
- [32] G. Paschos, "Perceptually uniform color spaces for color texture analysis: an empirical evaluation," *Image Processing, IEEE Transactions on*, vol. 10, 2001, pp. 932-937.
- [33] Hua-Wen Chang and Shu-Duo Lan, "Image fusion based on addition of wavelet coefficients," *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. International Conference on*, 2007, pp. 1585-1588.
- [34] G. Van De Wouwer, P. Scheunders, S. Livens, and D. Van Dyck, "Wavelet correlation signatures for color texture characterization," *Pattern Recognition*, vol. 32, 1999, pp. 443--451.
- [35] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation," *Computer Graphics and Image Processing*, vol. 13, Jul. 1980, pp. 222 - 241.
- [36] C.J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, Jun. 1998, pp. 121-167.
- [37] O. Ivanciuc, "Applications of Support Vector Machines in Chemistry," *Reviews in Computational Chemistry*, Lipkowitz: Wiley-VCH, Weinheim, 2007, pp. 291-400.
- [38] M. Hafner, A. Gangl, F. Wrba, K. Thonhauser, H. Schmidt, C. Kastinger, A. Uhl, and A. Vecsei, "Comparison of k-NN, SVM, and NN in Pit Pattern Classification of Zoom-Endoscopic Colon Images using Co-Occurrence Histograms," *Image and Signal Processing and Analysis, 2007. ISPA 2007. 5th International Symposium on*, 2007, pp. 516-521.
- [39] S. Sidhu and K. Raahemifar, "Texture classification using wavelet transform and support vector machines," *Electrical and Computer Engineering, 2005. Canadian Conference on*, 2005, pp. 941-944.
- [40] A. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, 2003, p. 199.
- [41] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology*, vol. 52, Jan. 1990, pp. 99-115.
- [42] A. Jain, Jianchang Mao, and K. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, 1996, pp. 31-44.
- [43] D. Heckerman, "A tutorial on learning with Bayesian networks," *Learning in Graphical Models*, 1995.

- [44] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2002.
- [45] D. Niedermayer, “An Introduction to Bayesian Networks and their Contemporary Applications,” 1998.
- [46] B. Namee, P. Cunningham, S. Byrne, and O. Corrigan, *The problem of bias in training data in regression problems in medical decision support*, 2002.
- [47] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*, Morgan Kaufmann, 2005.
- [48] J.G. Fletcher, F. Booya, R.M. Summers, D. Roy, L. Guendel, B. Schmidt, C.H. McCollough, and J.L. Fidler, “Comparative Performance of Two Polyp Detection Systems on CT Colonography,” *Am. J. Roentgenol.*, vol. 189, Aug. 2007, pp. 277-282.
- [49] J. Moore and G. Zouridakis, *Biomedical Technology and Devices Handbook*, CRC, 2003.
- [50] Z. Wang, Z. Liang, X. Li, D. Eremina, and H. Lu, “An Improved Electronic Colon Cleansing Method for Detection of Colonic Polyps by Virtual Colonoscopy,” *Biomedical Engineering, IEEE Transactions on*, vol. 53, 2006, pp. 1635-1646.
- [51] R. Van Uitert and R. Summers, “Automatic Correction of Level Set Based Subvoxel Precise Centerlines for Virtual Colonoscopy Using the Colon Outer Wall,” *Medical Imaging, IEEE Transactions on*, vol. 26, 2007, pp. 1069-1078.
- [52] L. Zhao, C. Botha, J. Bescos, R. Truyen, F. Vos, and F. Post, “Lines of Curvature for Polyp Detection in Virtual Colonoscopy,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, 2006, pp. 885-892.
- [53] B.S. Lewis and P. Swain, “Capsule endoscopy in the evaluation of patients with suspected small intestinal bleeding: Results of a pilot study,” *Gastrointestinal Endoscopy*, vol. 56, Sep. 2002, pp. 349-53.
- [54] F. Graepler, M. Wolter, R. Vonthein, and M. Gregor, “Accuracy of the size estimation in wireless capsule endoscopy: calibrating the M2A PillCam (with video),” *Gastrointestinal Endoscopy*, vol. 67, May. 2008, pp. 924-931.
- [55] M. Mylonaki, A. Fritscher-Ravens, and P. Swain, “Wireless capsule endoscopy: a comparison with push enteroscopy in patients with gastroscopy and colonoscopy negative gastrointestinal bleeding,” *Gut*, vol. 52, Aug. 2003, pp. 1122-6.
- [56] M. Coimbra, M. Mackiewicz, M. Fisher, C. Jamieson, J. Soares, and J.P.S. Cunha, “Computer Vision Tools for Capsule Endoscopy Exam Analysis,” *EURASIP Newsletter*, 2007, pp. p. 1 - 19.
- [57] K. Duda, T. Zielinski, R. Fraczek, J. Bulat, and M. Duplaga, “Localization of Endoscopic Capsule in the GI Tract Based on MPEG-7 Visual Descriptors,” *Imaging Systems and Techniques, 2007. IST '07. IEEE International Workshop on*, 2007, pp. 1-4.
- [58] M. Coimbra and J. Cunha, “MPEG-7 Visual Descriptors—Contributions for Automated Feature Extraction in Capsule Endoscopy,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, 2006, pp. 628-637.
- [59] R. Haralick, “Statistical and structural approaches to texture,” *Proceedings of the IEEE*, vol. 67, 1979, pp. 786-804.
- [60] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,”

Systems, Man and Cybernetics, IEEE Transactions on, vol. 3, 1973, pp. 610-621.

- [61] M.G. Turner and R.H. Gardner, *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity*, Springer, 1994.
- [62] K. Mohamed, B. Abdou, E. Mohamed, and H. Hassan, "Automated Image Analysis System for Homogeneity Evaluation of Nuclear Fuel Plates," *Neural Networks and Brain, 2005. ICNN&B '05. International Conference on*, 2005, pp. 1178-1183.
- [63] G. Van de Wouwer, P. Scheunders, and D. Van Dyck, "Statistical texture characterization from discrete wavelet representations," *Image Processing, IEEE Transactions on*, vol. 8, 1999, pp. 592-598.
- [64] D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil, "Robustness of Local Binary Patterns in Brain MR Image Analysis," *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 2098-2101.
- [65] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, 2006, pp. 2037-2041.
- [66] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, 2002, pp. 971-987.
- [67] Sae Hwang, JungHwan Oh, W. Tavanapong, J. Wong, and P. de Groen, "Polyp Detection in Colonoscopy Video using Elliptical Shape Feature," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2007, pp. II - 465-II - 468.
- [68] A.R. Feinstein, *Principles of Medical Statistics*, Chapman & Hall/CRC, 2001.
- [69] A.J. Viera and J.M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Family Medicine*, vol. 37, May. 2005, pp. 360-3.

10. Apêndice I

Ficheiro ARFF

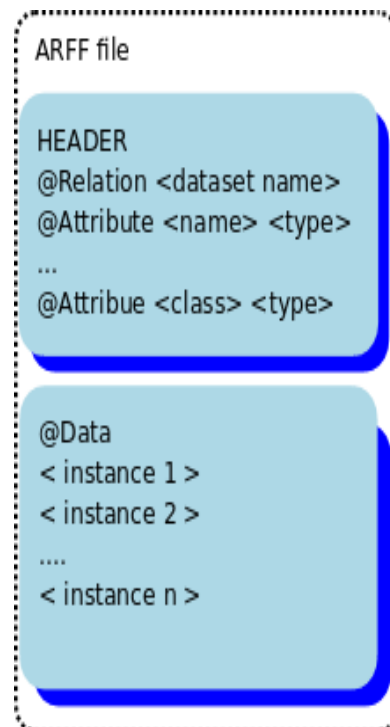
Formato desenvolvido pelos criadores do software WEKA para utilização com o mesmo, o ficheiro é preenchido usando texto ASCII onde estão descritos os vectores com características a serem processadas pelo software, este formato contém duas secções constituintes: o cabeçalho e o corpo onde estão presentes os dados.

O cabeçalho descreve o conteúdo o nome do conjunto de dados a que se refere, e os atributos de cada vector (instância) presente na secção de dados. Os atributos são descritos por um nome (por ex. Entropia, média, cor,...) e o tipo indica o tipo de dados desse atributo, que pode ser um número, *string*, uma data, ou um conjunto de dados nominais. Normalmente o último atributo corresponde à classe, e de forma a distinguir dos outros o nome do atributo é 'class' e o tipo da classe pode ser um dos descritos acima.

A secção de dados contém as instâncias propriamente ditas, e seus respectivos valores. Esta secção começa pela palavra reservada "@data" e todas as linhas seguintes correspondem a uma instância ou vector que contém todos os atributos discriminados no cabeçalho e estes atributos são separados por uma vírgula ','.

Exemplo, ficheiro *demo.arff*:

```
@Relation "medias anuais"  
@Attribute "nome" string  
@Attribute "média" numeric
```



```
@Attribute "class" {1,0}
@data
aluno1, 13, 1
aluno2, 9, 0
aluno3, 10, 1
...
```

O exemplo acima mostra a utilização das diversas palavras reservadas para delimitar as secções de cabeçalho e de dados, bem como a utilização de tipos distintos nos atributos (string e numéricos) e o atributo da classe nominal que indica os dois tipos possíveis, 1 ou 0. As instâncias seguem a estrutura descrita no cabeçalho.

O ficheiro *demo.arff* está de acordo com os requisitos mínimos descritos, pelo que está preparado para ser lido e processado no software WEKA.-6

11. Apêndice II

Ficheiro *characteristics.xml*, com todas as características e respectivos valores possíveis utilizados no software ECCA. Este ficheiro contém todas as possibilidades de escolha aquando da anotação das imagens nesse software.

```
<?xml version="1.0" standalone="yes"?>
<characteristics>
  <characteristic class="Pólipo" id="Tamanho">
    <value id="Gigante ( maior que 3cm )"/>
    <value id="Grande (1,5cm - 3cm)"/>
    <value id="Médio (5mm - 1,5cm)"/>
    <value id="Diminuto ( 0 - 5mm )"/>
  </characteristic>
  <characteristic class="Pólipo" id="Morfologia">
    <value id="Séssil"/>
    <value id="Semi-pediculado"/>
    <value id="Pediculado"/>
  </characteristic>
  <characteristic class="Pólipo" id="Cor">
    <value id="Vermelha"/>
    <value id="Pálida"/>
    <value id="Avermelhada"/>
  </characteristic>
  <characteristic class="Tumor" id="Tamanho">
    <value id="Gigante ( maior que 3cm )"/>
    <value id="Grande (1,5cm - 3cm)"/>
    <value id="Médio (5mm - 1,5cm)"/>
    <value id="Diminuto ( 0 - 5mm )"/>
  </characteristic>
  <characteristic class="Tumor" id="Morfologia">
    <value id="Polipoide"/>
    <value id="Ulcerado"/>
    <value id="Estonosante"/>
  </characteristic>
</characteristics>
```

```
</characteristic>
<characteristic class="Tumor" id="Cor">
  <value id="Congestivo"/>
  <value id="Pálida"/>
  <value id="Sangrante"/>
</characteristic>
<characteristic class="Normal"/>
<characteristic class="Outras" id="Restos alimentares">
  <value id="Sim"/>
  <value id="Não"/>
</characteristic>
<!-- quando não tem atributo 'class' a característica é global-->
<characteristic id="Qualidade da imagem">
  <value id="Excelente"/>
  <value id="Boa"/>
  <value id="Razoável"/>
  <value id="Má"/>
</characteristic>
</characteristics>
```