

M

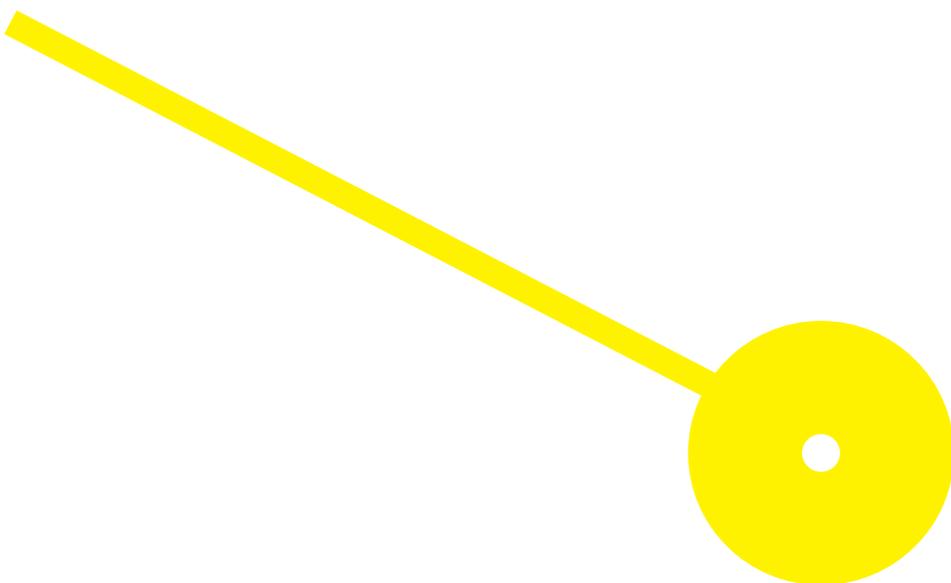
MESTRADO

Técnicas Laboratoriais em Biopatologia – Patologia Molecular

# Diagnóstico assistido por computador no Cancro Colorretal: podemos fazê-lo?

Liliana Filipa Pinho Ribeiro

10/2021



P.PORTO

ESCOLA  
SUPERIOR  
DE SAÚDE



IMP  
DIAGNOSTICS  
MOLECULAR & ANATOMIC PATHOLOGY



CADPath.AI



INESC TEC  
ASSOCIATE LABORATORY  
PORTUGAL

## Diagnóstico assistido por computador no Cancro Colorretal: podemos fazê-lo?

**Autor:**

Liliana Filipa Pinho Ribeiro

**Orientadores:**

Professor Doutor Jaime Cardoso, PhD, do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade De Engenharia da Universidade do Porto

Professora Doutora Regina Augusta Alves Pereira da Silva, PhD, professora coordenadora da Área Científica de Anatomia Patológica, Citológica e Tanatológica da Escola Superior de Saúde do Instituto Politécnico do Porto

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em **Técnicas Laboratoriais em Biopatologia** – Ramo/Área de Especialização em **Patologia Molecular** pela Escola Superior de Saúde do Instituto Politécnico do Porto.

### **Menção de apoio financeiro**

Dissertação inserida no projeto *CADPath.A* do laboratório IMP Diagnostics em parceria com o INESC TEC, realizado com o apoio do programa Compete 2020 do Fundo Europeu de Desenvolvimento Regional.



## Agradecimentos

A realização deste trabalho foi possível apenas devido a uma equipa que me apoiou e ajudou, seja na realização prática ou escrita do mesmo. Assim, os meus agradecimentos pertencem:

- Ao João Monteiro, meu colega e superior no IMP Diagnostics, que confiou em mim o suficiente para me convidar a fazer parte do projeto CADPath.AI, em que se insere o trabalho que aqui apresento;
- À Ana Monteiro, Gestora de Projetos do IMP Diagnostics e responsável pelo projeto CADPath.AI, sem a qual eu não teria um orientador de dissertação e que sempre se disponibilizou para me ajudar em tudo o que precisasse;
- À Doutora Diana Felizardo e ao Doutor João Fraga, patologistas no IMP Diagnostics e parte integrante da equipa do projeto que me forneceram todos os dados adicionais de que fui necessitando e que sempre tiveram disponíveis para responder a todas as dúvidas que fui tendo;
- À Sofia Gonçalves, técnica e colega no IMP Diagnostics e também parte da equipa do projeto, sem a ajuda da qual nada disto teria sido possível! Foi minha companheira de levantamento e digitalização de lâminas, e em conjunto aventuramo-nos em salas e garagens de arquivo e levantamos conjuntos de módulos com quase o nosso peso. No que toca a levantamento de lâminas tenho também de agradecer à Mariana Santos, auxiliar do laboratório e responsável pelo arquivo, que nos guiou no labirinto de ficheiros e módulos, facilitando o nosso trabalho e muitas vezes acompanhando uma de nós nas viagens ao arquivo. Sem ambas este processo teria sido muito mais tedioso e moroso;
- Por último, devo os meus mais sinceros e profundos agradecimentos ao Pedro Carneiro e ao Professor Jaime Cardoso, do INESC TEC, que me ajudaram e guiaram ao longo destes meses e tiveram uma paciência quase divina para as minhas perguntas de leiga informática, respondendo a todas as minhas dúvidas sempre as coloquei. Um especial agradecimento ao Professor Jaime que aceitou o desafio de me orientar neste trabalho e, sem o qual, não teria aqui chegado.

## Resumo

O número de casos de cancro colorretal tem aumentado exponencialmente, fazendo desta doença oncológica uma das mais diagnosticadas mundialmente. No entanto, o número de patologistas disponíveis para a realização do diagnóstico mantem-se constante, o que leva a um aumento da carga de trabalho destes profissionais, bem como a um aumento no tempo de resposta ao doente. É, desta forma, urgente a criação de ferramentas que ajudem os patologistas a gerir mais eficazmente o seu trabalho. A aplicabilidade crescente da patologia digital e os correntes avanços na tecnologia, fazem do diagnóstico assistido por computador uma ferramenta promissora na patologia.

Este trabalho teve como objetivos a criação de uma base de dados de amostras colorretais a ser usada para treinar um algoritmo de *deep learning* para o diagnóstico de amostras colorretais, e a avaliação da precisão desse algoritmo.

Foi construída uma base de dados com 10% das imagens anotadas. Com esta base foi possível fazer dois eventos de treino e teste do algoritmo, com um número crescente de amostras. Foram utilizadas três categorias de diagnóstico: benigno, lesão de baixo grau e lesão de alto grau.

Para cada evento foi possível atingir uma precisão de 88,42% e 89,89%, respetivamente. Para o diagnóstico diferencial entre benigno e lesão obteve-se uma precisão de 96%.

Apesar de ainda existirem algumas barreiras à implementação desta tecnologia na rotina patológica, o diagnóstico assistido por computador demonstra o potencial de auxiliar os patologistas numa mais eficaz gestão da sua carga laboral.

**Palavras-chave:** cancro colorretal, patologia digital, diagnóstico assistido por computador, *deep learning*, treino semi-supervisionado

## **Abstract**

The number of annual cases of colorectal cancer has increased exponentially, making this one of the most commonly diagnosed cancers in the world. However, the number of pathologists available for diagnosis remains constant, which leads to an increase in the workload of these professionals, as well as an increase in the response time for the patient. It is, therefore, urgent to create tools that help pathologists to manage their work more effectively. The growing applicability of digital pathology and current advances in technology, make computer-aided diagnosis a promising tool in pathology.

This work aims to create a database of colorectal samples, representative of the pathological routine, to be used to train a deep learning algorithm for the diagnosis of colorectal samples, and to evaluate the accuracy of this same algorithm.

A database was created, with 10% of the images annotated. With this base was possible to run two training and test events of the algorithm, with a crescent number of samples. Three diagnostic categories were used: benign, low-grade lesion and high-grade lesion.

For each event it was possible to reach an accuracy of 88.42% and 89.89%, respectively. For the differentiation of benign and lesion a precision of 96% was reached.

Although there are still some obstacles to implement of this technology in the routine pathology, computer-aided diagnosis seems to have the potential to help pathologists in a more effective management of their workload.

**Keywords:** colorectal cancer, digital pathology, computer-aided diagnosis, deep learning, semi-supervised training

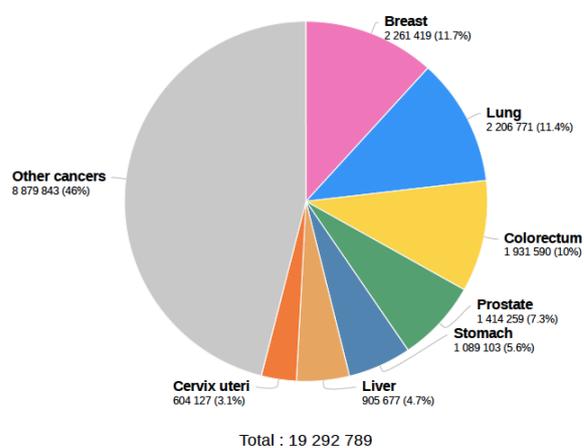
## Índice

1.	Introdução .....	1
1.1.	A Patologia Digital e a análise de imagem.....	2
1.2.	O progresso: o diagnóstico assistido por computador (Computer-aided diagnosis – CAD) e a inteligência artificial .....	4
1.2.1	O CAD no cancro colorretal .....	7
1.3.	Objetivo do estudo .....	9
2.	Metodologia.....	11
2.1.	Amostragem.....	11
2.2.	Processo de digitalização e verificação de imagem .....	11
2.3.	Categorias de diagnóstico.....	11
2.4.	Processo de anotação .....	13
2.5.	Treino do algoritmo.....	13
2.6.	Estatística .....	14
3.	Resultados .....	15
3.1.	Processo de digitalização e verificação de imagem .....	15
3.2.	Base de dados final e anotações .....	15
3.3.	Treino do algoritmo.....	15
3.3.1	Primeiro evento de treino .....	15
3.3.2	Segundo evento de treino .....	19
4	Discussão .....	22
5	Conclusão .....	25
	Referências Bibliográficas .....	27

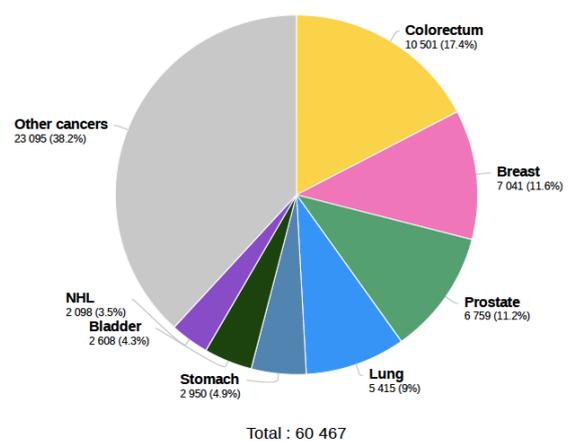
## 1. Introdução

Segundo dados da Globocan, em 2020, foram diagnosticados aproximadamente 1,9 milhões de novos casos de cancro colorretal em todo o mundo, tornando-o a terceira patologia oncológica mais prevalente mundialmente. A mesma entidade estima que em Portugal, no ano de 2020, tenham sido diagnosticados 10501 novos casos, correspondendo a 17,4% de todas as neoplasias diagnosticadas nesse ano (Figura 1) (1). A maior taxa de incidência é observada nos países desenvolvidos e, com a globalização e a adoção do estilo de vida ocidental pela restante população mundial, é expectável que a incidência mundial do Cancro Colorretal continue a aumentar (2,3).

Números estimados de novos casos, em 2020, mundialmente



Números estimados de novos casos, em 2020, em Portugal



**Figura 1.** Números do ano 2020 para os novos diagnósticos de patologias oncológicas, mundialmente e em Portugal. Estes demonstram que o cancro colorretal foi um dos cancros com mais novos casos diagnosticados, correspondendo a 10% de todos os cancros diagnosticados mundialmente. Em Portugal, o cancro colorretal foi a patologia oncológica mais diagnosticada, tendo ultrapassado o número de casos anuais dos cancros da mama e próstata. Fonte: Globocan, 2021

O diagnóstico precoce desta patologia através dos exames de rastreio (aconselhados a indivíduos assintomáticos a partir dos 50 anos) e com recurso a exames endoscópicos é fundamental, seja pela imediata investigação dos sintomas exibidos pelo paciente, seja pela possibilidade da realização de rastreio da população e dos grupos etários de risco, diminuindo a mortalidade em aproximadamente 16% (4,5). Estes exames permitem a visualização e localização de tumores, a identificação de outras neoplasias síncronas e a colheita de amostras para caracterização histológica (6). Desta forma, um grande volume de amostras é enviado, todos os anos, para os laboratórios de Anatomia Patológica. É, ainda, relevante acrescentar que devido à pandemia de Covid-19, os programas de rastreio de cancro colorretal, a nível mundial, sofreram atrasos. Como tal, é crucial que estes programas sejam capazes de compensar

os atrasos sofridos da forma mais rápida e eficaz possível, por forma a diminuir o impacto na progressão da doença nestes doentes (7,8).

Devido à importância do diagnóstico anatomopatológico, este deve estar disponível o mais rapidamente possível, permitindo o rápido início de terapia quando esta se revela necessária. Nos últimos anos, a dupla verificação de lâminas histológicas, especialmente em casos de malignidade, tem sido fortemente recomendada de modo a assegurar um correto diagnóstico. No entanto, enquanto o número de novos casos de Cancro Colorretal demonstra uma tendência crescente, o número de patologistas tem permanecido relativamente constante, aumentando a carga laboral destes profissionais (9). É, assim, da máxima importância a criação de soluções que permitam aos patologistas gerir de forma mais rápida e eficiente o seu trabalho.

### **1.1. A Patologia Digital e a análise de imagem**

O conceito de Patologia Digital, ou Telepatologia, apareceu no final dos anos 60, apesar do termo ter surgido apenas nos anos 80. A Patologia Digital era definida como o diagnóstico anatomopatológico realizado à distância, feito com recurso a imagens digitais de preparações macro e microscópicas, a dados clínicos e a informações relevantes enviados a um patologista que os receberia remotamente. Surgiu, assim, a possibilidade de realizar um diagnóstico à distância e de fazer pedidos de segundas opiniões por parte dos patologistas, eliminando a necessidade do envio físico das amostras (10,11).

Nos dias atuais são muitas as vantagens associadas à Patologia Digital. Num laboratório completamente digital a viagem de uma lâmina termina após a sua digitalização, sendo que, as imagens digitais das lâminas ficam disponíveis de imediato para visualização pelos patologistas no sistema informático. Isto permite reduzir em cerca de 13% o tempo que o patologista dedica a um determinado exame, uma vez que elimina o tempo dedicado à organização, procura e movimentação de lâminas. Existem, também, vantagens para a área técnica da Anatomia Patológica, uma vez que deixa de ser necessária a distribuição de lâminas em tabuleiros e a sua separação para os diferentes patologistas, acelerando a chegada das mesmas à área médica (10).

A Patologia Digital é uma ferramenta essencial para os hospitais mais deslocados dos centros urbanos, tendo o poder de os ajudar a melhorar os seus serviços de prestação de saúde. Ao oferecer supervisão remota à macroscopia, consultas intraoperatórias remotas e acesso à opinião de especialistas, quer no diagnóstico primário quer em pedidos de segunda opinião a Patologia Digital tem a capacidade de reduzir a necessidade de uma segunda cirúrgica, de oferecer suporte a patologistas a trabalhar em hospitais do interior e de facilitar o recrutamento de patologistas e cirurgiões (10). Pode, ainda, ter implicações no ensino, possibilitando a educação virtual (12).

Em Portugal, a Patologia Digital já é usada para possibilitar a realização de diagnósticos anatomopatológicos de forma mais rápida e eficiente no Centro Hospitalar Universitário da Cova da Beira.

Este projeto, pioneiro no país, intitula-se de “*e-Patologia*” e surge de uma parceria entre este Centro Hospitalar e o Instituto de Patologia e Imunologia Molecular do Porto (IPATIMUP), tendo como objetivo a combinação de meios informáticos de ambas as instituições para a obtenção, gestão, visualização e análise de todas as etapas do exame anatomopatológico, desde a sua entrada no laboratório até ao seu diagnóstico, sendo o diagnóstico realizado totalmente pelo patologista. Este projeto permitiu a realização de diagnósticos rápidos, bem como o aumento do número de diagnósticos realizados num dado período de tempo devido à redução do tempo no transporte das amostras entre as instituições (13).

Nos últimos 30 anos têm sido feitos grandes progressos no campo da ótica, tendo-se observado o surgimento de óticas e de câmaras digitais de grande qualidade. Da mesma forma, desde os anos 90 que têm ocorrido constantes progressos no poder de processamento de computadores, velocidade de transferência de dados e avanços em *software*. Estes progressos têm influenciado positivamente o uso da imagem digital na patologia (14).

Existem, atualmente, três métodos para transformar as preparações histológicas em imagens digitais: o método estático, que permite apenas a transmissão de imagens; o método dinâmico, que utiliza a comunicação em tempo real; e o método digital que permite a digitalização da totalidade da lâmina histológica e a simulação de um microscópio ótico convencional, denominada de *Whole Slide Imaging* (WSI) e que se revela a mais promissora (15).

Digitalizadores automáticos de WSI são, na sua base, microscópicos robóticos capazes de digitalizar uma lâmina histológica na sua totalidade. São realizadas fotografias de todos os campos da lâmina, assegurando que esta é completamente representada. Estas fotografias são depois utilizadas pelo *software* do digitalizador para criar uma imagem possível de se visualizar num computador e que se comporta como uma lâmina num microscópio ótico, podendo a imagem ser vista em diferentes ampliações (14).

A transformação de preparações histológicas em imagens digitais permitiu não só a consulta remota por parte de um patologista com recurso a um ecrã de computador, mas também possibilitou a sua análise digital usando *softwares* de análise de imagem. A introdução do WSI permitiu também avanços nestes programas informáticos, passando a sua aplicação em pequenas e manualmente selecionadas áreas de interesse para a possibilidade de analisar toda a imagem automaticamente (16,17). Estes *softwares* focam-se em 3 grandes categorias de análise: localização, classificação e quantificação de lesões (16). A sua utilização permite a deteção de eventos mitóticos, a segmentação e a classificação de tecidos (em benignos ou malignos) e a quantificação de imunoexpressão membranar e nuclear (14). Desta última são exemplos a possibilidade da quantificação da imunoexpressão do PDL-1 e do HER-2, biomarcadores importantes no diagnóstico dos cancro do pulmão e da mama, respetivamente, bem com da quantificação do sinal obtido na hibridização *in situ* (17). A análise de imagem possibilita, ainda, nas imagens de lâminas coradas pela coloração de Hematoxilina e Eosina (H&E), analisar métricas como a

densidade linfocitária, a composição do estroma tumoral e a morfometria nuclear, bem como a análise de texturas. Esta abordagem tem sido já utilizada em casos de cancro da mama, ovário e colorretal (14,15). Os *softwares* de análise de imagem têm sido apresentados como a melhor forma de eliminar a variabilidade inter e intra-observador no diagnóstico anatomopatológico. Estudos demonstram que a estimativa da percentagem de células tumorais, por exemplo, está sujeita a uma grande variabilidade interobservador (18,19). Em amostras colorretais, vários estudos têm demonstrado uma boa concordância no reconhecimento de características histológicas adenomatosas por diferentes patologistas, existindo, no entanto, uma grande variabilidade interobservador na avaliação do grau de displasia (20–22). Um dos estudos realizados para avaliar a variabilidade interobservador no diagnóstico de displasia de alto grau em pólipos cólicos revelou que, mesmo após a realização de uma conferência de consenso, a concordância no diagnóstico destas amostras foi muito baixa (20). A correta avaliação da arquitetura histológica das amostras colorretais é fundamental para o seguimento clínico do doente e para a decisão terapêutica tendo, assim, um enorme impacto no prognóstico clínico.

O uso de algoritmos demonstra, desta forma, a capacidade de ajudar os patologistas a encontrar nas amostras estruturas morfológicas pré-definidas e a classificá-las de forma mais precisa usando *scores* estabelecidos e, assim, reduzir as diferenças entre diagnósticos de diferentes profissionais. O uso de algoritmos tem a capacidade de tornar a patologia mais comparável mundialmente (11).

A análise de imagem é, no entanto, um processo interativo onde, tipicamente, alguns parâmetros são otimizados manualmente, compilados num algoritmo e testados numa área da imagem da lâmina. Aspectos que falham no controlo de qualidade são melhorados até que o algoritmo atinja os critérios de precisão pré-definidos. Com o surgimento e os avanços da inteligência artificial, também o campo da análise de imagem sofreu alterações uma vez que se tornou possível a criação de algoritmos mais robustos e que necessitam de menos otimizações manuais (16).

## **1.2. O progresso: o diagnóstico assistido por computador (*Computer-aided diagnosis* – CAD) e a inteligência artificial**

Nos últimos anos, com a crescente adoção da patologia digital nos laboratórios e, conseqüentemente, com a disponibilização de mais informação, em combinação com a baixa disponibilidade de patologistas para a realização de diagnósticos, foi possível observar uma grande evolução na área da patologia computacional (23–26). O diagnóstico assistido por computador (*computer-aided diagnosis* ou, simplesmente, CAD) pode ser considerado como um dos pioneiros na aplicação de inteligência artificial (AI) no diagnóstico e a sua utilização poderá entrar na rotina como um assistente ao diagnóstico (11).

O CAD refere-se ao uso de modelos de inteligência artificial para apoiar na realização de diagnósticos médicos. No último quarto de século, o CAD tem evoluído de uma ferramenta utilizada na investigação para uma ferramenta usada na prática clínica na área da imagiologia médica. Na patologia, esta tecnologia

traz a promessa da eliminação da necessidade da avaliação por um segundo patologista nos casos de neoplasia (11). No entanto, apesar dos diversos estudos realizados da sua aplicação na Anatomia Patológica, ainda não se verifica a sua utilização na rotina devido à complexidade das imagens e sua interpretação e responsabilidade associadas ao diagnóstico patológico (27).

Na prática, a AI é um termo que compreende o uso de um computador para imitar, e que pode por vezes ultrapassar, a inteligência humana (25). Na Anatomia Patológica, a AI permite a realização de reconhecimento de padrões de forma simples, eficiente e robusta, através do treino de modelos informáticos (11).

Apesar de os termos inteligência artificial e *Machine Learning* (ML) serem muitas das vezes utilizados como sinónimos, a ML é uma subdivisão da AI, que aplica métodos estatísticos para otimizar modelos para uma tarefa específica (25). A ML é, assim, um modelo em que um computador repetidamente aprende com a informação que lhe é dada, tendo depois a capacidade de inferir sobre novos dados apresentados sem a interferência humana. (28)

A *Machine Learning* pode ser dividida em duas categorias: ML supervisionada e ML não supervisionada. Ambas necessitam de extensas bases de dados de treino para a aprendizagem do reconhecimento dos padrões, residindo as suas diferenças na informação dada ao algoritmo durante o seu treino. Nos modelos supervisionados são usados dados emparelhados com anotações, classificações dadas por especialistas, que servem como exemplo de aprendizagem para que o algoritmo possa fazer previsões sobre novos casos inseridos após o treino. Já na ML não supervisionada, esta não necessita de anotações, tendo a capacidade de agrupar casos com características semelhantes. Para que seja possível aplicar estes modelos na clínica, as bases de dados de treino têm de ser suficientemente extensas para contemplar todo o espectro de diagnósticos possíveis de encontrar na rotina patológica, especialmente nos modelos não supervisionados onde as características histológicas relevantes para o diagnóstico são desconhecidas para o algoritmo(29). A *Machine Learning* supervisionada é, neste momento, o modelo dominante nas soluções de inteligência artificial aplicadas ao diagnóstico médico (25).

O *Deep Learning* (DL), surge como uma subdivisão da ML, baseado numa rede neural (*neural network* - NN) com várias camadas, vindo daí o termo *deep* (28). As NNs são modelos inspirados no processamento de informação das redes neuronais biológicas. Os modelos de DL têm um *input* (a base de dados de treino), uma (várias) camada(s) escondida(s) e um *output* (a predição feita pelo modelo). Este tem, assim, a capacidade de aprender com a informação "em bruto" que lhe é fornecida, dando a estes modelos a capacidade de extrair informação de bases de dados extensas e tornando-os mais robustos e mais exatos (19,25).

A *convolutional neural network* (CNN) é uma das NNs com mais sucesso no *Deep Learning*, especialmente no campo da análise de imagem, sendo adequada para o processamento de imagem 2D ou 3D. Um dos componentes das CNNs são camadas de convolução, que permitem a passagem consecutiva de

informação entre camadas do modelo, mimetizando a resposta do córtex visual de um mamífero a um estímulo visual. Esta tecnologia trouxe grandes avanços no processamento de imagem, sendo particularmente útil nas imagens histopatológicas devido à sua complexidade, e é, hoje, aplicada numa grande variedade de modelos de visão computacional (19).

O DL facilita o treino dos modelos uma vez que a extração de informação é uma parte intrínseca da forma como o modelo é otimizado para aprender diretamente dos dados inseridos. Uma vez que o modelo tenha sido otimizado com uma grande quantidade de dados de treino, os padrões aprendidos por estes podem ser aplicados para a realização de predições de categorias em dados previamente desconhecidos. No entanto, dependendo do tipo de modelo desenvolvido, pode ser necessário uma extensa anotação dos dados do *set* de treino o que, quando aplicado a imagens histopatológicas, despende muito do tempo de um patologista especializado (25).

Vários estudos foram já realizados para demonstrar que a integração de sistemas de DL no diagnóstico anatomopatológico leva a um aumento da precisão no diagnóstico e, conseqüentemente, a uma redução do erro humano (30). Nos últimos anos, foram publicados estudos que demonstram as vantagens da aplicação da inteligência artificial no diagnóstico e prognóstico de diversos tumores, sendo alguns dos exemplos de possíveis aplicações: o diagnóstico do cancro prostático (31,32); a classificação de gliomas (33); a deteção de células pulmonares malignas (34,35); na caracterização histológica e deteção de metástases de cancro da mama (36–39) e no diagnóstico de cancro colorretal (35,40,41). No entanto, ainda não se verifica a aplicação destes modelos informáticos à prática clínica, uma vez que esta tecnologia ainda apresenta algumas limitações na sua concessão e conseqüente utilização.

Um dos desafios é a validação dos modelos de DL. Muitos dos estudos feitos estabelecem os algoritmos de inteligência artificial com bases de dados reduzidas, construídas com dados maioritariamente de uma única fonte. Modelos de aprendizagem supervisionada, como já referido, necessitam de bases de dados extensas e, muitas das vezes, extensamente anotadas. Adicionalmente, entre laboratórios é possível encontrar variações na preparação histológica e nos modelos de digitalizadores, bem como no processo de digitalização. Todas estas variações vão ter impacto na forma como o modelo extrapola a informação e faz as suas predições. Assim, o algoritmo pode beneficiar do treino com amostras de diferentes instituições, ou até mesmo no treino com bases de dados públicas (27).

Outro obstáculo é a infraestrutura necessária para a adoção e uso destes modelos. As imagens histológicas são consideravelmente mais pesadas do que as geradas pelos exames imagiológicos, podendo uma lâmina digitalizada com uma objetiva de 40x gerar uma imagem entre 0,5 Gb a 4 Gb (16). É, assim, necessário um investimento em armazenamento digital bem como um aumento da capacidade de processamento do sistema, o que pode ser um investimento demasiado pesado para alguns laboratórios (27).

Outro desafio é a possível resistência por parte de alguns patologistas na adoção desta tecnologia, uma vez que algumas dúvidas permanecem sobre o funcionamento e utilização destes modelos. Da mesma forma, existem algumas preocupações sobre a forma como um patologista deve reportar e estruturar o relatório patológico aquando da utilização dos modelos CAD como assistentes ao diagnóstico. Também a acreditação destes sistemas pode constituir uma barreira à sua utilização na rotina (27).

Uma possível solução para as dificuldades na construção da base de dados e posterior treino é o uso de modelos de aprendizagem fracamente supervisionada. Nesta, a extensa anotação é substituída por uma “etiqueta” de diagnóstico, algo fácil de fazer uma vez que é uma informação já presente nos sistemas informáticos laboratoriais. A “etiqueta” de diagnóstico dá uma anotação fraca a todos os campos da lâmina digital. Esta anotação funciona sob a premissa de que se o diagnóstico é benigno, então todos os campos da imagem são benignos; se o diagnóstico é de lesão, então pelo menos um campo da imagem possui lesão(42). Este é um exemplo generalista da assunção de múltipla instância ou *Multiple Instance Learning* (MIL). Apesar de este paradigma de treino ser extremamente aplicável no diagnóstico de WSIs, e apesar do seu sucesso em algoritmos clássicos de visão de computador, a MIL é ainda pouco aplicada na patologia computacional devido à escassez de extensas bases de WSIs que possam tirar todo o partido desta tecnologia (43).

### 1.2.1 O CAD no cancro colorretal

Recentemente, as tecnologias do *Deep learning* e do diagnóstico assistido por computador têm sido aplicadas no diagnóstico e classificação do cancro colorretal, embora de forma limitada. Dos estudos realizados na área, muito poucos têm como finalidade o diagnóstico (44).

Em 2012, Kalkan *et al.* (41) propuseram um método automático de deteção do cancro colorretal que pretendia, ao contrário dos métodos de análise de imagem usados até então, a análise da lâmina histológica na sua totalidade para a obtenção do diagnóstico. Para isso, desenvolveram um método que combina a análise das características texturais e estruturais de áreas da imagem, usando depois um esquema de classificação em dois níveis – primeiro as diferentes áreas da imagem são classificadas em quatro possíveis classes (adenoma, inflamação, tumor e normal). Depois, usando a informação dada pela classificação anterior, as lâminas são classificadas em normal ou tumor com recurso a um classificador linear logístico. Foram usadas neste estudo um total de 120 lâminas histológicas coradas por H&E, tendo sido obtida uma precisão média de 87,69%, com precisões de 79,17% e 92,68% para as classes de tumor e normal, respetivamente (41).

De forma semelhante, ainda utilizando técnicas tradicionais de visão de computador, Yoshida *et al.* (9) sugeriram um método de classificação de lâminas histológicas de cancro colorretal, dividindo-as em quatro categorias: benigno, adenoma, carcinoma e não classificável. Neste estudo foram usadas e digitalizadas 1068 lâminas coradas por H&E, perfazendo um total de 1328 secções de tecido analisáveis.

A análise das amostras foi realizada em dois passos: análise da atipia citológica e análise da atipia arquitetural. O primeiro passo recorre a MIL para classificar o grau de atipia citológica, em baixa ou alta, do tecido. A imagem foi então classificada como baixa atipia, atipia intermédia, alta atipia ou inclassificável com base em características nucleares e citoplasmáticas. O resultado final foi obtido através da combinação de ambas as classificações feitas, sendo as amostras classificadas como carcinoma nos casos em que os valores de atipia são mais altos. Este modelo demonstrou uma taxa de 9,3% de carcinomas classificados incorretamente como lesões benignas (9).

Em 2017, Korbar *et al* (45) apresentaram o primeiro modelo, com aplicação do *Deep Learning*, para classificar automaticamente pólipos colorretais, em lâminas coradas por H&E, em cinco classes diferentes: normal, hiperplásico, séssil serreado, adenoma serreado tradicional, adenoma tubular e adenoma viloso/tubuloviloso. Cerca de 697 lâminas foram digitalizadas, anotadas por um patologista e divididas em pequenas secções. Cada uma dessas secções foi então classificadas utilizando o *RESNET-152* (uma NN), sendo a predição de cada lâmina obtida como o tipo de pólipo colorretal mais comum entre todas as secções analisadas. Na eventualidade de menos de cinco secções serem identificadas com o tipo de pólipo mais comum e com uma confiança superior a 70%, a lâmina foi classificada como benigna. Este modelo obteve uma precisão de 89,7% (45).

Em 2020 vários autores apresentaram soluções para o diagnóstico do cancro colorretal. Lizuka *et al* (40) apresentaram um modelo para a classificação de amostras colorretais em não neoplásicas, adenoma e adenocarcinoma. Neste estudo foi usada uma base de dados com um total de 4536 WSIs, sendo cada uma destas dividida em secções menores. A cada uma destas secções foi atribuída uma categoria de diagnóstico. O diagnóstico foi conseguido através da agregação das secções e da combinação do diagnóstico efetuados em cada uma delas. Além da base de dados inicial, o modelo foi testado em 547 casos de receções cólicas obtidas no repositório *The Cancer Genome Atlas*, contendo amostras normais e de adenocarcinoma. Para a primeira base de dados foi obtida uma AUC (*area under the curve*) de 0,962 e 0,992 para adenocarcinoma e adenoma, respetivamente. Na base de dados pública, a AUC atingida para adenocarcinoma foi de 0,982 (40). Wei *et al* (46) procuraram criar um modelo que permitisse identificar cinco tipos morfológicos de pólipos – normal, adenoma tubular, adenoma viloso/tubuloviloso, pólipo hiperplásico e adenoma serreado – em lâminas histológicas corada por H&E. Nesse estudo foram usadas 509 lâminas, com anotações em áreas relevantes realizadas por cinco patologistas especializados, para o treino do algoritmo (base de dados interna); e 238 lâminas externas, de diferentes instituições (base de dados externa). Neste modelo, as lâminas foram divididas em secções mais pequenas e foram combinadas com um classificador hierárquico, sendo o diagnóstico da lâmina dado com base na comparação da frequência dos diagnósticos previsto para cada secção. A precisão obtida para a base de dados interna foi de 93,5%, e a obtida para a base de dados externa foi de 87% (46).

Num estudo publicado em 2021, Masud *et al*/(35) propuseram um modelo com o objetivo de diagnosticar não só cancro colorretal, mas também cancro do pulmão. Para tal, utilizaram cinco classes para representar diferentes tipos de tecido histológico: adenocarcinoma cólico, cólon benigno, adenocarcinoma pulmonar, carcinoma de células escamosas do pulmão e pulmão benigno. Para o treino do modelo foram recolhidas 1250 imagens de uma base de dados pública (LC25000), sendo selecionadas 250 imagens de cada um dos tipos histológicos em estudo. Cada uma das imagens foi seccionada em áreas mais pequenas, perfazendo um total de 25000 imagens. Foram usadas duas técnicas de Processamento de Imagem Digital para a extração das características histopatológicas das imagens. A classificação das imagens foi feita com recurso a um grupo de NNs que realizam a extração de características dos dados fornecidos ao modelo. Foi atingida uma precisão de 96,33% no diagnóstico de ambas as patologias (35). Apesar de todos estes estudos mostrarem um grande potencial, algumas limitações ainda podem ser apontadas aos testes realizados com estes modelos. Uma das grandes limitações que se pode apontar é que, na maioria destes estudos, os modelos não foram testados em bases de dados/casos externos. Este é um passo importante pois permite a comparação entre modelos, algo que não é possível quando estes são testados em bases de dados completamente distintas. A adição deste passo pode, também, estar a limitar o estudo da robustez do modelo, uma vez que o teste é feito numa base de dados semelhante à usada para o treino deste. Outro ponto a considerar é o tamanho da base de dados. Uma característica comum a quase todos os estudos aqui abordados são bases de dados compostas por algumas centenas de WSIs. Bases de dados maiores favorecem o treino dos modelos, aumentando a sua precisão e representando de forma mais aproximada as amostras encontradas no dia-a-dia clínico. O tamanho da base de dados pode ser, no entanto, justificado com a necessidade de anotação das lâminas o que, como referido anteriormente, é um processo moroso e minucioso, sendo difícil de realizar em bases de dados com um grande volume de amostras.

É ainda de se mencionar que, em muitos dos estudos, não é indicado a forma como foram recolhidas as amostras ou como foi realizado o processo de anotação. A experiência dos patologistas que realizam as anotações e a forma como são resolvidas possíveis discordâncias nos diagnósticos são, também, raramente descritas.

Assim, apesar de se poder observar que a aplicação do diagnóstico assistido por computador ao Cancro Colorretal em Patologia é possível de ser feito, existe ainda a possibilidade de melhorar a construção e teste destes modelos, tornando-os mais robustos e possibilitando a sua introdução na prática clínica.

### **1.3. Objetivo do estudo**

Com este estudo pretende-se testar uma ferramenta informática, um algoritmo de *deep learning*, com potencial para realizar a classificação de amostras colorretais com um desempenho comparável ao de um

patologista permitindo a sua utilização uma melhor gestão e uma redução da carga laboral atual destes profissionais de saúde.

Para tal será criada uma base de dados de imagens digitalizadas de preparações colorretais para o treino de um algoritmo de *Deep Learning* capaz de classificar os casos em benigno ou com lesão. O método de treino a utilizar será uma combinação dos métodos supervisionado e fracamente supervisionado. Neste estudo serão, ainda, realizados dois treinos para avaliar a precisão do algoritmo e inferir sobre a sua utilização na prática clínica.

## 2. Metodologia

### 2.1. Amostragem

Para o presente estudo, foram utilizadas um total de 15147 lâminas histológicas, coradas por H&E, relativas a biópsias e polipectomia colorretais, que se encontravam arquivadas no Laboratório IMP *Diagnostics*, referentes a casos do período temporal de 2018 a 2021.

### 2.2. Processo de digitalização e verificação de imagem

As lâminas foram digitalizadas com recurso a dois digitalizadores Aperio GT450DX, da Leica Biosystems, a uma ampliação de 40x. Antes da digitalização, todas as lâminas foram limpas com recurso a uma solução alcoólica. Por forma a garantir a qualidade da digitalização, as lâminas foram também inspeccionadas para a eliminação de fatores que pudessem levar a erros durante o processo de digitalização, nomeadamente:

- As lâminas em que o código se encontrava riscado ou com baixa qualidade de impressão foram impressas novas etiquetas;
- Nos casos em que se observou retração de meio sintético de montagem, duplas lamelas, bolhas de ar e lamelas desalinhas procedeu-se à remontagem das lâminas.

Todos os erros detetados pelos digitalizadores foram registados e corrigidos, tendo-se procedido a novas digitalizações e remontagens de lâminas quando necessário.

A qualidade das imagens resultantes foi avaliada por dois patologistas num ecrã de uso médio, LG 27HJ712C-W, recorrendo ao software Aperio e Slide Manager para visualização da imagem. Foram eliminadas do estudo imagens com baixa qualidade de digitalização e/ou com demasiados artefactos técnicos, bem como imagens repetidas (correspondentes a cortes mais fundos).

### 2.3. Categorias de diagnóstico

Três categorias de diagnóstico foram definidas para o algoritmo: benigno, lesões de baixo grau (correspondendo a adenomas convencionais com displasia de baixo grau) e lesões de alto grau (incluindo carcinomas) (**Tabela 1**). Na categoria “benigno” não foram incluídos casos com suspeita/história de doenças inflamatórias. Foram selecionados os casos de adenomas convencionais, uma vez que estes correspondem a grande parte da rotina diária de um laboratório de Anatomia Patológica, tendo sido excluídas do estudo as amostras com lesões serradas e outros tipos de pólipos.

Aquando da revisão dos casos estes foram, na sua maioria, coincidentes com o diagnóstico dado inicialmente. Os casos em que ocorreu discordância com o diagnóstico inicial foram revistos por ambos os patologistas tendo sido o diagnóstico decidido por ambos. Os critérios utilizados para a revisão dos diagnósticos, mais especificamente para a classificação das amostras em displasia de baixo ou alto grau,

encontram-se sumariadas na **Tabela 2**. Em caso de discordância entre os patologistas e em casos mais complexos ocorreu a intervenção de um terceiro patologista com mais anos de experiência.

**Tabela 1.** Definição das categorias de diagnóstico para o algoritmo.

<b>Categorias para o algoritmo</b>	<b>Diagnóstico patológico</b>
<b>Benigno</b>	Mucosa colorretal normal, inflamação, hiperplasia
<b>Lesão de baixo grau</b>	Adenoma com displasia de baixo grau
<b>Lesão de alto grau</b>	Adenoma com displasia de alto grau e adenocarcinoma invasivo.

**Tabela 2.** Caracterização das lesões de baixo e alto grau para amostras colorretais.

<b>Parâmetros</b>	<b>Lesão de baixo grau</b>	<b>Lesão de alto grau</b>
<b>Extensão</b>	--	Alterações devem envolver mais do que duas glândulas (exceto em pequenas biópsias de pólipos).
<b>Caraterísticas observadas a baixa ampliação</b>	Ausência de complexidade estrutural.	Anomalias estruturais complexas, epitélio com aspeto espessado, coloração azul, desorganização arquitetural.
<b>Citologia/arquitetura</b>	Não combina displasia citológica de alto grau com caraterísticas estruturais de alto grau.	Combina displasia citológica de alto grau com caraterísticas estruturais de alto grau.
<b>Caraterísticas arquiteturais</b>	Aglomerção de glândulas, em disposição paralela, não complexas (não se observa <i>back-to-back</i> nem forma de crivo). Arquitetura global pode variar entre tipos tubular e viloso.	Aglomerção glandular complexa e irregular, com aparência cribiforme e <i>back-to-back</i> .
<b>Caraterísticas citológicas</b>	Núcleos aumentados e hipercromáticos, alguns com forma alongada. Os núcleos mantêm a orientação basal. Não se observa perda de polaridade celular nem pleomorfismo, ou mitoses atípicas. Manutenção da maturação celular (mucinas).	Núcleos aumentados, frequentemente com cromatina dispersa e nucléolos evidentes. Perda da polaridade celular e com estratificação nuclear, com os núcleos distribuídos ao longo de 1/3 da altura do epitélio. Mitoses atípicas. Apoptose/necrose proeminente, dando uma aparência "suja" à amostra. Diminuição de maturação celular (mucinas).

## 2.4. Processo de anotação

Cerca de 10% da base de dados final foi anotada por dois patologistas, com recurso ao programa *Sedeen Viewer*®. Todas as anotações foram revistas e correções foram feitas quando necessário. Para os casos mais complexos, em que um consenso não foi possível, à semelhança do ocorrido para as revisões de diagnóstico, as anotações foram revistas por um terceiro patologista.

Foi criado um mapa de anotação, em que cada área das imagens a serem anotadas foi marcada com uma cor diferente de acordo com a sua classificação. Um resumo do esquema de cores adotado pode ser consultado na **Tabela 3**.

**Tabela 3.** Mapa de anotação adotada para as imagens anotadas

<i>Cor</i>	<i>Representação</i>
	Fundo
	Normal/Benigno
	Displasia de baixo grau
	Displasia de alto grau/Carcinoma (inclui necrose)
	Áreas a ignorar/detritos
	Fulguração

## 2.5. Treino do algoritmo

Com base em estudos realizados previamente no diagnóstico assistido por computador para o Cancro Colorretal e devido à ausência de bases de dados de grandes dimensões anotadas, como já referido, foi decidido utilizar um método de treino semi-supervisionado. Este método de treino foi desenhado pelo grupo de trabalho do INESC TEC (*Institute for Systems and Computer Engineering, Technology and Science*) e, de forma resumida, consiste em três etapas:

- Pré-treino supervisionado: previamente à utilização da base de dados total disponível, foram primeiro utilizadas as WSIs anotadas. A utilização de um treino inicial com uma pequena percentagem da base de dados anotada tem a potencialidade de melhorar os resultados obtidos no treino de modelos fracamente supervisionado, sendo uma alternativa mais rápida à anotação extensiva de uma grande base de dados;
- Treino fracamente supervisionado: nesta fase, toda a base de dados disponível foi utilizada para o treino do modelo. O modelo foi usado para inferir sobre todas as classes de diagnóstico das várias secções das imagens. Baseado nessas predições, as áreas foram classificadas e selecionadas para o treino com base nas “etiquetas” de diagnóstico. Nesta fase as anotações são ignoradas a favor do diagnóstico final das amostras;

- Depois do treino do modelo são extraídas as características de cada área. Após a extração, são selecionadas áreas relevantes com base nas suas predições e vetores são criados a partir das características extraídas. Por último, é realizado o treino de um método de agregação que fará as predições com base nos vetores criados. Este passo permite uma predição de diagnóstico mais estável.

Por forma a avaliar a evolução da performance com o aumento do *set* de treino, foram realizados dois eventos de treino com *sets* de diferentes tamanhos. O primeiro evento de treino compreendeu cerca de 10% da base de dados total e foi, adicionalmente, subdividido em três fases de treino, com e sem pré-treino, de forma a perceber a verdadeira relevância dos dados anotados. O segundo evento de treino compreendeu cerca de 40% da base de dados total e foi testado apenas o modelo com pré-treino.

## **2.6. Estatística**

Para o cálculo das percentagens das amostras foi usado o método estatístico descritivo. Os restantes resultados estatísticos foram extraídos de valores fornecidos pelo algoritmo.

### 3. Resultados

#### 3.1. Processo de digitalização e verificação de imagem

Apesar das precauções tomadas para obter um processo digitalização rápido e eficaz, alguns problemas ocorreram durante esta fase.

Os digitalizadores usados têm a capacidade de avaliar a qualidade não só da lâmina histológica, mas da própria digitalização, assinalando com erro as lâminas que não preenchem os requisitos necessários a uma digitalização de qualidade, ou até mesmo, para que a digitalização possa ser feita. Dependendo do problema encontrado, uma diferente mensagem de erro é demonstrada no ecrã interativo. Na **Tabela 4** encontram-se registados os diferentes tipos de erros encontrados, as possíveis causas e as soluções encontradas para os resolver.

A frequência destes erros foi contabilizada de forma a perceber o seu impacto no processo de digitalização. No total, foram contabilizados 117 erros, o que corresponde a 0,77% do número total de lâminas digitalizadas (n=15147). Como é possível observar na **Figura 2** o erro mais frequente foi "Image quality", que ocorre quando o *software* do digitalizador considera que a imagem resultante da digitalização tem baixa qualidade, que ocorreu um total de 55 vezes, perfazendo cerca de 47% do total dos erros ocorridos. Pelo contrário, o erro que menos ocorreu foi o "Internal error", que pode ser considerado o erro mais grave uma vez que leva à paragem do processo de digitalização, causando atrasos, especialmente em instâncias em que a digitalização foi realizada durante a noite. Este ocorreu um total de 5 vezes, o que corresponde a 4,3% dos erros ocorridos.

#### 3.2. Base de dados final e anotações

Foi possível construir uma base de dados com um total de 10 000 WSIs, 1 000 das quais anotadas.

Abaixo, na Erro! A origem da referência não foi encontrada., é possível ver um exemplo de uma imagem anotada para locais com lesão de baixo grau e com lesão de baixo grau.

#### 3.3. Treino do algoritmo

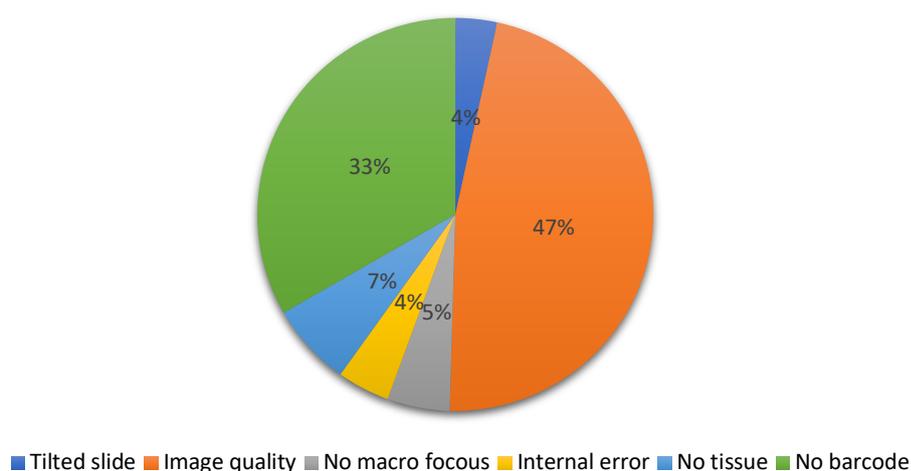
##### 3.3.1 Primeiro evento de treino

Para o primeiro evento de treino foi utilizado um *set* com um total de 1133 WSIs, das quais 100 se encontravam anotadas. Das WSIs utilizadas, 294 correspondem a amostras benignas, 517 de lesões de baixo grau e 222 de lesões de alto grau. Para as amostras anotadas 6 correspondem a amostras benignas, 35 de lesões de baixo grau e 59 de lesões de alto grau. A distribuição de cada categoria de diagnóstico no *set* de treino pode ser consultada na Erro! A origem da referência não foi encontrada..

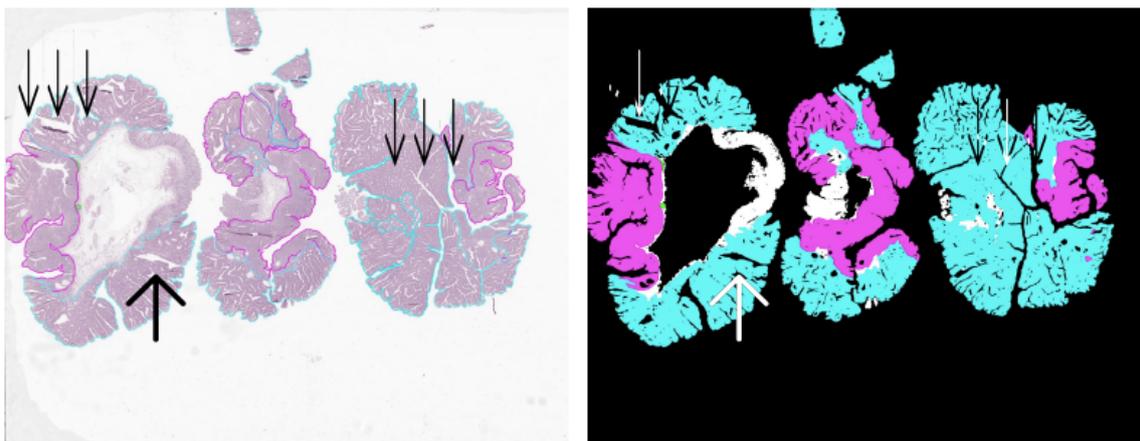
**Tabela 4.** Tipos de erros identificados pelos digitalizadores, com respetivas causas e resoluções.

Tipo de erro	Causa	Resolução
"No barcode"	Erro na leitura do código por este se encontrar rasurado ou com má qualidade de impressão	Impressão de nova etiqueta
"No macro focus"	Impossibilidade de focagem por retração do meio de montagem	Remontagem da lâmina e nova digitalização
"Image quality"	A qualidade de digitalização é considerada fraca pelo software do digitalizador	Nova digitalização. Em caso de persistência do erro procedeu-se à remontagem da lâmina
"Tilted slide"	Meio de montagem nas bordas da lâmina leva a que digitalizador a interprete como mal colocada	Remoção do meio em excesso, quando possível, ou remontagem da lâmina e nova digitalização
"No tissue"	Amostras de tamanho muito reduzido que não são identificadas pelo sistema <i>tissue finder</i> do digitalizador.	Tentativa de nova digitalização (maioritariamente mal sucedida)
"Internal error"	Lâmina presa no cesto ou no interior do digitalizador por lamela não corretamente alinhada ou por excesso de meio de montagem nas bordas.	Retirada do cesto internamente, remontagem e nova digitalização da lâmina.

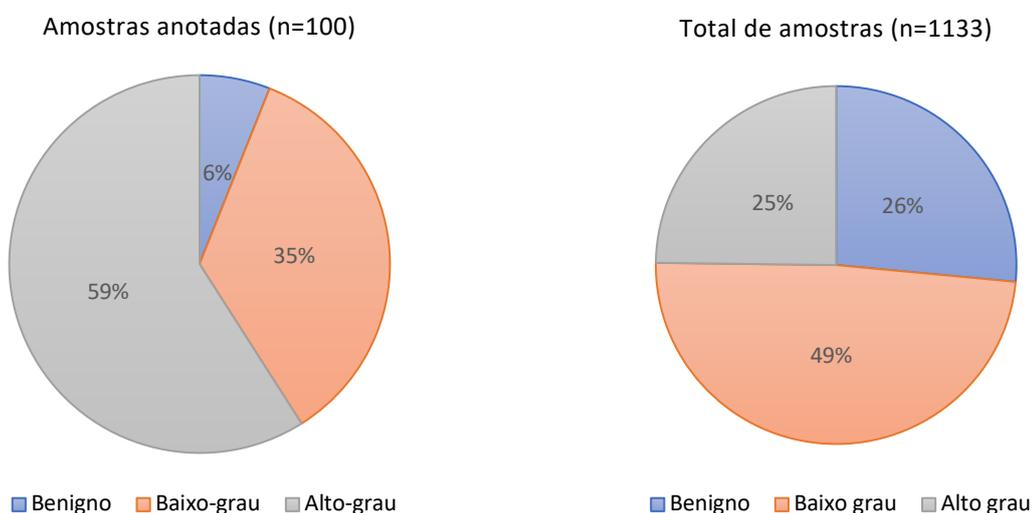
**Erros detetados pelos digitalizadores**



**Figura 2.** Tipos de erros detetados pelos digitalizadores durante o processo de digitalização. Das 15147 lâminas digitalizadas 117 deram origem a erros, ou seja, 0,77% das lâminas. Os erros mais frequentes foram "Image quality", correspondendo a 47% dos erros totais, e "No barcode", que corresponde a 33,3% dos erros.



**Figura 3.** Exemplo de imagem anotada. à esquerda, imagem antes do processamento de imagem para treino. À direita, imagem após processamento (remoção de fundo). É possível observar a azul marcadas as áreas de tecido com lesão de baixo grau e a rosa áreas com lesão de alto grau.



**Figura 4.** Distribuição de cada uma das categorias de diagnóstico no primeiro *set* de treino.

Esta distribuição é bastante desequilibrada quando se toma em conta apenas as *WSIs* não anotadas. No entanto, foi tido um especial cuidado em equilibrar estes números ao proceder-se à anotação de uma grande quantidade de casos de lesão de alto grau (59% dos casos anotados, como se pode verificar na **Figura 4**). Esperava-se, ainda, que com uma considerável quantidade destas amostras anotadas pudesse ser útil para um melhor treino do algoritmo na distinção de lesões mais graves e complexas.

Quando se converte esta informação na quantidade de *tiles* usados para o treino do algoritmo é possível perceber a real dimensão da informação dada a este. O *set* de treino é composto por um total de 1 322 596 de *tiles*. Destes 211 235 correspondem a *tiles* anotados e os restantes 1 111 361 a *tiles* não anotados. Um resumo da distribuição da informação anotada e não anotada usada no treino pode ser consultada abaixo, na **Tabela 5**.

**Tabela 5.** Sumarização da informação contida no primeiro set de treino.

	<i>Categorias de Diagnóstico</i>			Total
	Benigno	Lesão de baixo grau	Lesão de alto grau	
<i>Com Anotação</i>				
<i>WSIs</i>	6	35	59	100
<i>Tiles</i>	49 640	77 946	83 649	211 235
<i>Sem Anotação</i>				
<i>WSIs</i>	294	517	222	1033
<i>Tiles</i>	-	-	-	1 111 361

WSIs – Whole Slide Images

Por forma a perceber, adicionalmente, as diferenças de performance do modelo com e sem pré-treino, este evento de treino foi subdividido em três passos: treino apenas com as WSIs anotadas, treino com a totalidade do *set* sem pré-treino e treino com a totalidade do *set* após pré-treino.

A avaliação da performance foi realizada com recurso a um *set* de teste com 240 WSIs, das quais 54 correspondem a amostras benignas, 143 a amostras de lesão de baixo grau e 62 a amostras de lesão de alto grau.

Diferenças significativas foram observadas nas três diferentes fases de treino realizadas neste evento, com um aumento gradual da precisão por parte do modelo. É possível observar a evolução da performance na **Tabela 6**.

**Tabela 6.** Performance geral do modelo após o primeiro evento de treino, com e sem pré-treino.

Dataset	Pré-treino	Precisão (%)
Set anotado (n=100)	Não	75,00
Set total (n=1133)	Não	84,17
	Sim	<b>88,42</b>

No treino utilizando apenas as amostras anotadas foi atingida uma precisão de 75%. Foi possível aumentar esta precisão não só com o treino com a utilização do total das amostras apenas com a anotação fraca (“etiqueta” de diagnóstico), com a precisão a ser de 84,17%, mas também utilizando o pré-treino com as amostras anotadas, seguida do treino com o número total das amostras, onde a precisão atingida (e a mais alta desta fase) a ser de 88,42%.

Por forma a melhor compreender a performance obtida, foram extraídos os resultados (precisão e médias de confiança) para cada categoria de diagnóstico, individualmente. Foi possível perceber que o modelo teve mais sucesso na interpretação de lesões de baixo grau, com uma precisão de 90,9% e uma média de

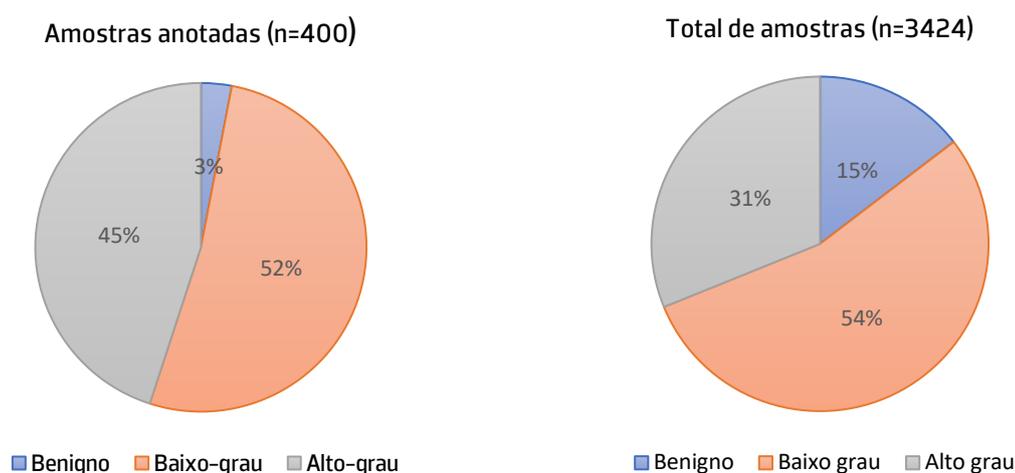
confiança de 93% nos diagnósticos realizados corretamente. No entanto, a média de confiança quando errado foi também a mais alta (88,3%). No sentido oposto, a menor precisão pertence às lesões de alto grau, sendo esta apenas de 82,3%, mas foi também para estas lesões que se observou uma maior distância entre as médias de confiança, com 92,5% de média de confiança quando o diagnóstico foi realizado corretamente e 74,9% quando o resultado estava errado (Tabela 7).

**Tabela 7.** Precisão e médias de confiança obtidos para cada uma das categorias de diagnóstico no primeiro evento de treino, com pré-treino.

Diagnóstico	Precisão (%)	Média de confiança quando correto (%)	Média de confiança quando errado (%)
Benigno	88,8	92,1	80,6
Lesão de baixo grau	<b>90,9</b>	93	88,3
Lesão de alto grau	82,3	92,5	74,9

### 3.3.2 Segundo evento de treino

No segundo evento de treino foi utilizado um *set* com um total de 3424 WSIs, das quais 400 anotadas. Das WSIs utilizadas, 495 correspondem a amostras benignas, 1857 de lesões de baixo grau e 1072 de lesões de alto grau. Para as amostras anotadas 12 correspondem a amostras benignas, 207 de lesões de baixo grau e 181 de lesões de alto grau. A distribuição de cada categoria de diagnóstico no *set* de treino pode ser consultada na **Figura 5**. Quando esta informação é passada para *tiles*, o *set* de treino é composto por um total de 5 5770 979 *tiles*, dos quais 505 617 são anotados. Um resumo da distribuição da informação anotada e não anotada usada no treino pode ser consultada abaixo, na **Tabela 8**.



**Figura 5.** Distribuição de cada uma das categorias de diagnóstico para o segundo *set* de treino.

**Tabela 8.** Sumarização da informação contida no segundo *set* de treino.

<i>Categorias de Diagnóstico</i>				
	Benigno	Lesão de baixo grau	Lesão de alto grau	Total
<b>Com Anotação</b>				
WSIs	12	207	181	400
<i>Tiles</i>	145 898	196 116	163 603	505 617
<b>Sem Anotação</b>				
WSIs	483	1650	891	3024
<i>Tiles</i>	-	-	-	5 265 362

WSIs – *Whole Slide Images*

A avaliação da performance foi realizada com recurso a um *set* de teste com 1009 WSIs, das quais 168 correspondem a amostras benignas, 537 a amostras de lesão de baixo grau e 304 a amostras de lesão de alto grau.

Para este evento de treino a precisão de diagnóstico alcançada pelo algoritmo foi de **89,89%**. Neste evento de treino tentou-se, ainda, perceber qual a capacidade do algoritmo para distinguir entre amostras benignas e com lesão. A precisão para o caso binário (benigno versus lesão) foi de **96%** (Tabela 9).

**Tabela 9.** Precisões obtidas no segundo evento de treino.

	Precisão (%)
Geral	89,89
Caso Binário	96

À semelhança do que foi feito no evento de treino anterior, foram extraídos os resultados (precisão e médias de confiança) para cada categoria de diagnóstico, individualmente. A melhor precisão obtida ocorreu, novamente, nas lesões de baixo grau, 92,4%. Ainda nesta categoria, observa-se uma subida na média de confiança nos resultados corretos, com uma média e confiança de 95,6%, e uma descida na média de confiança nos diagnósticos realizados incorretamente, 81,9%. Ao contrário do ocorrido anteriormente, a precisão mais baixa ocorre no diagnóstico dos casos benignos, com 81,5%, uma descida quando comparado com o treino anterior. No entanto, a maior diferença entre as médias de confiança continua a ocorrer nas lesões de alto grau, com 97,6% de média de confiança quando o diagnóstico é realizado corretamente e 78,2% quando o diagnóstico dado é errado. Esta é também a média de confiança com maior aumento entre os dois eventos de treino para os resultados corretos.

Estes resultados podem ser consultados na **Tabela 10**.

**Tabela 10.** Precisão de médias de confiança obtidos para cada uma das categorias de diagnóstico no segundo evento de treino.

Diagnóstico	Precisão (%)	Média de confiança quando correto (%)	Média de confiança quando errado (%)
Benigno	81,5	92,1	87,7
Lesão de baixo grau	<b><u>92,4</u></b>	95,6	81,9
Lesão de alto grau	90,1	97,6	78,2

#### 4 Discussão

Os resultados obtidos no presente trabalho mostram-se promissores, com uma performance muito semelhante à de outros trabalhos, apesar de o modelo utilizado ser distinto (9,40,41).

Em trabalhos anteriormente feitos sobre diagnóstico assistido por computador na patologia é colocado uma grande ênfase no tempo necessário para realizar as anotações, no entanto raramente é descrito o processo de anotação ou, ainda, o passo que o precede: o processo de digitalização e os erros que podem ocorrer durante este processo (9,40,41,45,46). Na rotina da patologia digital a percentagem de erros no processo de digitalização situa-se entre os 0,5% e os 0,1%, após otimização do procedimento (47,48). Neste trabalho, a quantidade de erros registada, apesar de um pouco mais alta do que encontrado na rotina (0,77%), é baixa, reduzindo o seu impacto no tempo despendido na criação da base de dados. Este aumento relativamente à percentagem observada na digitalização das lâminas de rotina pode ser explicado pelo facto de terem sido utilizadas lâminas de arquivo, que vão perdendo alguma qualidade com o tempo, especialmente a retração do meio de montagem, causando possíveis pontos de desfocagem na digitalização – tendo sido o erro “Image quality” o mais comum ao longo do processo de digitalização. A instabilidade causada pelo levantamento da lamela, devido à retração do meio nas bordas, é uma das principais causas de desfocagem durante a digitalização, como apontado em um trabalho prévio sobre a qualidade das imagens histológicas na Patologia Digital (49).

Se a fase de inspeção das lâminas parece ter sido essencial para minimizar a quantidade de erros encontrados, a qualidade das lâminas também afeta a qualidade do processo de digitalização. Se, para a visualização de uma lâmina ao microscópio ótico, alguns artefactos técnicos não têm um grande impacto, na patologia digital, devido à impossibilidade de ajustar a focagem, estes mesmos artefactos podem tornar impossível a visualização adequada da imagem histológica. Artefactos como dobras, bolhas ou detritos levam a desfocagens que podem afetar toda a imagem e que levam à impossibilidade da sua utilização (48). Este problema é amplificado quando falamos no diagnóstico assistido por computador, onde a precisão e a capacidade de realizar um diagnóstico pelo algoritmo pode ser afetada pelos artefactos técnicos e/ou pela fraca qualidade de digitalização. Assim, destaca-se a importância do trabalho dos técnicos de anatomia patológica e, particularmente, da forma como abordam todos os processos desde a receção da amostra no laboratório até à obtenção da preparação histológica, por forma a assegurar a obtenção de cortes de qualidade que irão minimizar os erros encontrados no processo de digitalização, bem como assegurar que a imagem que chega para diagnóstico ao patologista é passível de ser avaliada.

No treino do algoritmo, a realização de dois eventos de treino distintos permitiu retirar algumas informações importantes. Com a realização do primeiro evento de treino foi possível perceber a potencialidade do algoritmo e do seu método de treino. Apesar da base de dados utilizada ser bastante reduzida, a boa precisão alcançada (88,42%) permitiu avançar para o segundo evento com uma base de

dados de maiores dimensões. A precisão alcançada de 89,89% com o segundo evento de treino permitiu perceber que o aumento do *set* de treino pode, efetivamente, ter um impacto positivo na performance do algoritmo. Estas precisões encontram-se dentro das já alcançadas por outros trabalhos feitos no diagnóstico assistido por computador para o cancro colorretal (41,45).

Dois pontos relevantes podem, ainda, ser extraídos dos resultados. O primeiro é a importância das amostras anotadas no treino do algoritmo. Como foi possível ver no primeiro evento de treino, as anotações, e consequentemente o pré-treino, tiveram um impacto positivo não só na performance geral do algoritmo (com esta a aumentar com a realização do pré-treino), mas também nas médias de confiança. Este impacto é visto, particularmente nas médias de confiança para as lesões de alto grau, onde 59% das amostras anotadas correspondem a esta categoria. Aqui, além da média de confiança quando o diagnóstico é correto ser a mais alta, é também onde se verificou a maior distância entre as médias de confiança. No segundo evento de treino, onde existe um maior equilíbrio entre a quantidade de amostras de lesões de baixo e de alto grau é também possível ver uma melhoria na média de confiança quando correto para as lesões de baixo grau, bem como um afastamento entre as médias de confiança. É, assim, possível dizer que as anotações, mesmo que apenas de uma pequena percentagem da base de dados de treino total, leva a uma melhoria da performance do algoritmo, tornando-o mais robusto do que seria caso fosse utilizado apenas um método de treino fracamente supervisionado.

O segundo ponto é que o número de amostras de cada categoria de diagnóstico presente na base de dados parece ter um impacto direto na performance do algoritmo. Em ambos os eventos de treino as bases de dados utilizadas possuem uma maior quantidade de amostras de lesões de baixo grau, quando comparado com as restantes duas categorias de diagnóstico. Nos resultados, isto traduziu-se numa melhor performance para esta categoria, sendo aquela em que se vê a maior precisão diagnóstica. A importância do número de amostras é observada, particularmente, no segundo evento de treino. Neste, observa-se um aumento da precisão para as lesões de alto grau, acompanhando o aumento do número de amostras desta categoria no *set* de treino. Pelo contrário, para os casos benignos verifica-se uma descida na precisão, de 88,8% para 81,5%, uma vez que existem poucas amostras desta categoria no *set*. Pode-se assim dizer que a performance individual para cada categoria de diagnóstico depende da quantidade de exemplos dados ao algoritmo durante o treino, podendo assim a performance ser melhorada com um maior equilíbrio na quantidade de amostras de cada categoria.

As categorias de diagnóstico escolhidas para este trabalho não são as mesmas que as utilizadas em trabalhos anteriores. No presente trabalho foi escolhido distinguir amostras normais de amostras com lesões de baixo grau e de amostras com lesões de alto grau, outros trabalhos optam por avaliar apenas a distinção entre benigno e maligno (35), ou por abordar um ponto de grande variabilidade interobservador que é a distinção entre tipos de adenomas (45,46). Nos casos em que categorias semelhantes são utilizadas não é referido o desempenho individual para cada categoria (41) ou é apenas feita uma distinção

entre displasia e tumor (9). Não é, assim, possível fazer uma comparação direta entre estes resultados neste trabalho e nos realizados anteriormente. No entanto, como já foi referido, a precisão geral entre os diferentes trabalhos é em tudo semelhante à obtida neste, podendo assim, ser dito, que as categorias de diagnóstico escolhidas não têm um impacto significativo na performance deste. As diferenças nas escolhas das categorias de diagnóstico parecem residir no objetivo traçado para cada trabalho, não sendo possível afirmar que uma determinada divisão de diagnóstico é mais correta do que outra.

Apesar de todos os resultados do presente trabalho serem importantes, dá-se um enfoque especial à precisão obtida para o caso binário, com um valor de 96%. Este resultado demonstra que este algoritmo tem a capacidade de distinguir amostras benignas das amostras com lesão. Esta precisão é também semelhante à alcançada por Masud *et al*, que alcançou uma precisão de 96,33% na distinção entre mucosa colorretal benigna e a maligna com recurso a um algoritmo de *machine learning* (35); e superior à alcançada anteriormente por Kalkan *et al*, em 2012, em que a precisão atingida foi apenas de 87,69% (41). Assim, o poder discriminativo desta ferramenta informática entre amostras benignas e amostras com lesão poderá ajudar os patologistas na seleção dos casos com lesão, casos prioritários, para observação microscópica e diagnóstico, ajudando a reduzir o tempo de resposta no diagnóstico e iniciar possíveis terapias. Desta forma, os Patologistas conseguem organizar a sua carga laboral mais eficientemente, que se reflete com um impacto positivo e significativo no rastreio do Cancro Colorretal.

De referir que o algoritmo utilizado no presente trabalho não foi validado com uma base de dados externa, sendo os *sets* de treino e de validação provenientes da mesma base de dados, que foi construída com amostras de um único laboratório. A semelhança/uniformização encontrada no processamento, coloração e processo de digitalização das amostras pode influenciar positivamente os resultados obtidos, isto é a performance do algoritmo poderá apresentar uma precisão mais baixa no diagnóstico de amostras provenientes de outra fonte externa.

## 5 Conclusão

A histologia, hoje, tem ainda um papel fundamental no diagnóstico do cancro, sendo a base e o *gold standard* da patologia. Continua, assim, a fazer sentido investir no desenvolvimento desta área, contribuindo para a melhoria da resposta dada aos doentes oncológicos. Este desenvolvimento tornar-se ainda mais urgente quando falámos no tempo de resposta no diagnóstico destes pacientes, crítico para os pedidos de técnicas complementares ou para o início de tratamentos, que pode ser comprometido pela crescente carga laboral para os laboratórios de Anatomia Patológica e, conseqüentemente, para os Patologistas.

Com este trabalho, foi pretendido a construção de uma base de dados com amostras colorretais e avaliar o desempenho de uma ferramenta informática capaz de ajudar os patologistas na sua rotina laboral diária, especificamente, na gestão das amostras colorretais. Apesar da base de dados construída ser de grandes dimensões e dos resultados obtidos durante os testes ao algoritmo serem promissores, algumas questões continuam sem a devida resposta e nem todos os obstáculos foram ultrapassados.

O maior *set* de treino avaliado neste trabalho contem um total de 3424 amostras. Este é um número superior ao da grande maioria dos trabalhos efetuados anteriormente, no entanto é ainda um número muito aquém das 10 000 WSIs que compõem a base de dados construída, e que melhor representam a rotina de um laboratório na patologia colorretal. É assim, ainda, necessário continuar o treino e avaliação da performance do algoritmo com *sets* de treino de número crescente (até atingir as 10 000), por forma a perceber se esta performance pode ainda ser melhorada e se existe um tamanho ótimo para o *set*. Nestes novos eventos de treino terá de ser tomada em conta o número e a proporcionalidade entre as amostras anotadas e não anotadas de cada categoria, por forma a tentar uniformizar as performances para as três categorias de treino e, possivelmente, aumentar a precisão global do algoritmo.

A isto, acrescenta-se a necessidade de testar o algoritmo numa base de dados externa, por forma a avaliar o seu comportamento em amostras que podem ter variações no processamento e coloração e testar a sua verdadeira aplicabilidade na rotina para diferentes laboratórios, uma vez que os *sets* de teste utilizados podem estar a influenciar positivamente a performance alcançada.

Como foi referido no presente trabalho, a necessidade de uma base de dados de dimensões consideráveis, os problemas levantados para a sua construção e o investimento monetário necessário para o fazer, podem continuar a ser um entrave ao CAD no Cancro Colorretal, havendo a necessidade da construção de uma base de dados pública extensa que permita uma maior acessibilidade para a construção e teste dos algoritmos de *deep learning*. Acrescenta-se ainda a resistência de alguns patologistas na utilização destas ferramentas informáticas. O facto de ainda não ser possível explicar claramente como é que estes algoritmos aprendem com a informação que lhes é fornecida, ou quais os parâmetros da imagem utilizam para fazer o diagnóstico ajuda a que alguns profissionais tenham um certo grau de desconfiança, limitando o uso do diagnóstico assistido por computador na Anatomia Patológica.

As possíveis vantagens trazidas pelo CAD à Patologia são inúmeras, no entanto, mais estudos e investimento são necessários para que esta seja uma tecnologia passível de ser implementada na rotina. O algoritmo de *deep learning* aqui apresentado ainda não está pronto para ser utilizado. Assim, pretende-se continuar a testar o algoritmo por forma a que o treino seja um reflexo das amostras colorretais observadas na rotina de Anatomia Patológica, bem como testar amostras externas, por forma a tornar este algoritmo uma ferramenta possível de utilizar na rotina laboral dos patologistas no âmbito do diagnóstico de biópsias colorretais. Durante estes futuros treinos e avaliações é de ressaltar a importância em equilibrar o número das amostras de cada categoria, por forma a tentar uniformizar a performance e melhorar a performance no global e, assim, ser uma ferramenta informática que irá melhorar a gestão da carga laboral dos patologistas, proporcionando um melhor rastreio e diagnóstico no cancro colorretal.

## Referências Bibliográficas

1. International Agency for Research on Cancer (IARC). Global Cancer Observatory.
2. Colorectal cancer statistics – Colorectal cancer is the third most common cancer worldwide [Internet]. [cited 2020 Aug 3]. Available from: <https://www.wcrf.org/dietandcancer/cancer-trends/colorectal-cancer-statistics>
3. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*. 2020;70(3):145–64.
4. Serviço Nacional de Saúde. Rastreio do cancro colorretal [Internet]. [cited 2020 Jul 30]. Available from: <https://www.dgs.pt/em-destaque/rastreio-do-cancro-colo-rectal.aspx>
5. Hassan C, Antonelli G, Dumonceau JM, Regula J, Bretthauer M, Chaussade S, et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2020. *Endoscopy*. 2020;52(8):687–700.
6. Catarina G, Ribeiro P. Carcinoma colo-rectal : diagnóstico e tratamento. 2010.
7. de Jonge L, Worthington J, van Wifferen F, Iragorri N, Peterse EFP, Lew J bin, et al. Impact of the COVID-19 pandemic on faecal immunochemical test-based colorectal cancer screening programmes in Australia, Canada, and the Netherlands: a comparative modelling study. *The Lancet Gastroenterology and Hepatology*. 2021;6(4):304–14.
8. Ricciardiello L, Ferrari C, Cameletti M, Gaianill F, Buttitta F, Bazzoli F, et al. Impact of SARS-CoV-2 Pandemic on Colorectal Cancer Screening Delay: Effect on Stage Shift and Increased Mortality. *Clinical Gastroenterology and Hepatology* [Internet]. 2021; Available from: <https://doi.org/10.1016/j.cgh.2020.09.008>
9. Yoshida H, Yamashita Y, Shimazu T, Cosatto E, Kiyuna T, Taniguchi H, et al. Automated histological classification of whole slide images of colorectal biopsy specimens. *Oncotarget*. 2017;8(53):90719–29.
10. Griffin J, Treanor D. Digital pathology in clinical use: Where are we now and what is holding us back? *Histopathology*. 2017;70(1):134–45.
11. Pallua JD, Brunner A, Zelger B, Schirmer M, Haybaeck J. The future of pathology is digital. *Pathology Research and Practice* [Internet]. 2020;216(9):153040. Available from: <https://doi.org/10.1016/j.prp.2020.153040>
12. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *Journal of Pathology Informatics*. 2018;9(1):1–11.
13. Centro Hospitalar Universitário da Cova da Beira [Internet]. [cited 2020 Jul 6]. Available from: <http://www.chcbeira.pt/index.php/informacao-util/2012-04-26-10-10->

27/localizacao-do-centro-hospitalar-cova-da-beira-epe/46-informacao-util/?cix=noticia18821&lang=4

14. Swamy M, Pradesh Pashu-Chikitsa M, Vishwavidyalaya V, Dwivedi S, Dubey A, Verma Y. The advent of digital pathology: A depth review. Article in JOURNAL OF ENTOMOLOGY AND ZOOLOGY STUDIES [Internet]. 2019;7(2):43–9. Available from: <https://www.researchgate.net/publication/331587775>
15. Santos CG. Telepatologia: Análise Comparativa de Gastos. Faculdade de Economia da Faculdade do Porto. 2018;
16. Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *Journal of Pathology*. 2019;249(3):286–94.
17. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman D, et al. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. *Journal of Pathology Informatics* [Internet]. 2019;9(1):1–33. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6437786/?report=printable>
18. Pell R, Oien K, Robinson M, Pitman H, Rajpoot N, Rittscher J, et al. The use of digital pathology and image analysis in clinical trials. *Journal of Pathology: Clinical Research*. 2019;5(2):81–90.
19. Deng S, Zhang X, Yan W, Chang EIC, Fan Y, Lai M, et al. Deep learning in digital pathology image analysis: a survey. *Frontiers of Medicine*. 2020;14(4):470–87.
20. Mahajan D, Downs-Kelly E, Liu X, Pai RK, Patil DT, Rybicki L, et al. Reproducibility of the villous component and high-grade dysplasia in colorectal adenomas <1 cm: Implications for endoscopic surveillance. *American Journal of Surgical Pathology*. 2013;37(3):427–33.
21. Osmond A, Li-Chang H, Kirsch R, Divaris D, Falck V, Liu DF, et al. Interobserver variability in assessing dysplasia and architecture in colorectal adenomas: A multicentre Canadian study. *Journal of Clinical Pathology*. 2014;67(9):781–6.
22. Turner JK, Williams GT, Morgan M, Wright M, Dolwani S. Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales. *Histopathology*. 2013;62(6):916–24.
23. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*. 2016;33:170–5.
24. Dimitriou N, Arandjelović O, Caie PD. Deep Learning for Whole Slide Image Analysis: An Overview. *Frontiers in Medicine*. 2019;6(November):1–7.

25. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*. 2020;288(1):62–81.
26. Rakha EA, Toss M, Shiino S, Gamble P, Jaroensri R, Mermel C, et al. Current and future applications of artificial intelligence in pathology: A clinical perspective. *Journal of Clinical Pathology*. 2020;(figure 1):1–6.
27. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Communications*. 2020;40(4):154–66.
28. Sakamoto T, Furukawa T, Lami K, Pham HHN, Uegami W, Kuroda K, et al. A narrative review of digital pathology and artificial intelligence: Focusing on lung cancer. *Translational Lung Cancer Research*. 2020;9(5):2255–76.
29. Jahn SW, Plass M, Moifar F. Digital Pathology: Advantages, Limitations and Emerging Perspectives. *Journal of Clinical Medicine*. 2020;9(11):3697.
30. He Y, Zhao H, Wong STC. Deep learning powers cancer diagnosis in digital pathology. *Computerized Medical Imaging and Graphics [Internet]*. 2021;88(June 2020):101820. Available from: <https://doi.org/10.1016/j.compmedimag.2020.101820>
31. Kwak JT, Hewitt SM. Multiview boosting digital pathology analysis of prostate cancer. *Computer Methods and Programs in Biomedicine [Internet]*. 2017;142:91–9. Available from: <http://dx.doi.org/10.1016/j.cmpb.2017.02.023>
32. Nagpal K, Foote D, Liu Y, Chen PHC, Wulczyn E, Tan F, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digital Medicine [Internet]*. 2019;2(1). Available from: <http://dx.doi.org/10.1038/s41746-019-0112-2>
33. Rathore S, Niazi T, Iftikhar MA, Chaddad A. Glioma grading via analysis of digital pathology images using machine learning. *Cancers*. 2020;12(3):1–13.
34. Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, et al. Artificial intelligence in lung cancer pathology image analysis. *Cancers*. 2019;11(11):1–16.
35. Masud M, Sikder N, Nahid A al, Bairagi AK, Alzain MA. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors (Switzerland)*. 2021;21(3):1–21.
36. Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*. 2015;20(1):237–48.
37. Vo DM, Nguyen NQ, Lee SW. Classification of breast cancer histology images using incremental boosting convolution networks. *Information Sciences [Internet]*. 2019;482:123–38. Available from: <https://doi.org/10.1016/j.ins.2018.12.089>

38. Wan T, Cao J, Chen J, Qin Z. Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. *Neurocomputing* [Internet]. 2017;229:34–44. Available from: <http://dx.doi.org/10.1016/j.neucom.2016.05.084>
39. Bejnordi BE, Veta M, van Diest PJ, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - Journal of the American Medical Association*. 2017;318(22):2199–210.
40. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Scientific Reports* [Internet]. 2020;10(1):1–11. Available from: <http://dx.doi.org/10.1038/s41598-020-58467-9>
41. Kalkan H, Nap M, Duin RPW, Loog M. Automated Colorectal Cancer Diagnosis for Whole-slide Histopathology. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*. Springer Berlin Heidelberg; 2012. p. 550–7.
42. Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine* [Internet]. 2019;25(8):1301–9. Available from: <http://dx.doi.org/10.1038/s41591-019-0508-1>
43. Campanella G, Silva VWK, Fuchs TJ. Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology. 2018; Available from: <http://arxiv.org/abs/1805.06983>
44. Thakur N, Yoon H, Chong Y. Current trends of artificial intelligence for colorectal cancer pathology image analysis: A systematic review. *Cancers*. 2020;12(7):1–19.
45. Korbar B, Olofson AM, Mirafior A, Nicka CM, Suriawinata MA, Torresani L, et al. Deep learning for Classification of Colorectal Polyps on Whole-Slide Images. *Journal of Pathology Informatics*. 2017;
46. Wei JW, Suriawinata AA, Vaickus LJ, Ren B, Liu X, Lisovsky M, et al. Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on Histopathologic Slides. *JAMA Network Open*. 2020;3(4):1–11.
47. Frassetta F, Caputo A, Guglielmino R, Pellegrino MG, Runza G, L'Imperio V. A Survival Guide for the Rapid Transition to a Fully Digital Workflow: The "Caltagirone Example." *Diagnostics*. 2021;11(10):1916.

48. Ratamero JA, Aneiros-Fernandez J, del Moral RG. Complete Digital Pathology for Routine Histopathology Diagnosis in a Multicenter Hospital Network. *Archives of Pathology and Laboratory Medicine*. 2020;(144):221–8.
49. Hosseini MS, Brawley-Hayes JAZ, Zhang Y, Chan L, Plataniotis K, Damaskinos S. Focus Quality Assessment of High-Throughput Whole Slide Imaging in Digital Pathology. *IEEE Transactions on Medical Imaging*. 2020;39(1):62–74.