

MODELOS PREDITIVOS DA MANUTENÇÃO DE ENDOSCÓPIOS

ANA MARGARIDA DA COSTA FERNANDES

novembro de 2025

MODELOS PREDITIVOS DA MANUTENÇÃO DE ENDOSCÓPIOS

Ana Margarida da Costa Fernandes

**Dissertação para obtenção do Grau de Mestre em
Engenharia Biomédica**

Orientador: Luís Coelho

Co-orientador: Eduardo Abreu

Júri:

Presidente:

Joaquim Alves, Professor Adjunto, ISEP

Vogais:

Sara Reis, Professor Adjunto, ISEP

Luís Coelho, Professor Coordenador, ISEP

Dedicatória

Quero dedicar esta tese aos meus pais, que fizeram tudo para me dar a oportunidade de estudar.

Resumo

A endoscopia consiste num procedimento médico efetuado com um instrumento designado "endoscópio", o qual permite a observação do interior de um órgão oco ou de uma cavidade do corpo. A ausência de um sistema centralizado e automático de gestão de tarefas pode resultar em falhas na sua programação, levando à omissão de manutenções e à realização de reparações tardias. Tal aumenta o risco de falhas nos equipamentos médicos e influencia a segurança dos pacientes.

A tese/dissertação desenvolveu-se em parceria com a empresa Fujifilm, para o desenvolvimento de algoritmos de *machine learning* (ML), para auxiliar na gestão da manutenção de endoscópios. O objetivo é identificar padrões que permitam determinar o tipo de intervenção seguinte (inspeção ou reparação), prever a quantidade de intervenções para os anos futuros, de 2025 a 2030; prever o tempo médio até à intervenção futura e otimizar a gestão do inventário de peças de reposição com base em dados históricos.

O estudo abrange o conjunto de dados de todos os equipamentos de clientes de Portugal submetidos a reparação ou inspeção no Centro de Serviços da Fujifilm entre 2020 e 2024. Foram aplicadas técnicas de pré-processamento de dados, análise estatística e métodos preditivos, incluindo regressão linear, regressão polinomial de segundo grau, modelos Random Forest e redes neuronais.

Os resultados demonstraram que o modelo de regressão linear aplicado à previsão do número de reparações anuais entre 2025 e 2030 obteve um coeficiente de determinação (R^2) de 0,9722, indicando que cerca de 97 % da variação nas reparações pode ser explicada pelo ano. O modelo de regressão polinomial de segundo grau apresentou um desempenho ligeiramente superior, com um R^2 ajustado de 0,976.

Na previsão do tempo entre reparações, os modelos de regressão linear apresentaram desempenho mais limitado, com R^2 ajustado de 0,1242 e, com validação cruzada, de aproximadamente 0,11. O modelo Random Forest obteve um erro médio absoluto (MAE) entre 122 e 123 dias, e erro quadrático médio (RMSE) entre 158 e 159 dias, o que equivale a uma margem de erro de cerca de cinco meses.

Relativamente à previsão do tipo de manutenção, o modelo Random Forest alcançou uma exatidão entre 77,37 % e 78,10 %, demonstrando uma boa capacidade de classificação. O modelo de redes neuronais também apresentou um desempenho positivo, com erros médios quadráticos (RMSE) de aproximadamente 0,58 tanto nos conjuntos de treino como de teste, o que indica ausência de sobreajuste (overfitting).

Palavras-chave: Endoscópio, Manutenção, *Machine Learning*, Gestão de Equipamentos Médicos

Abstract

Endoscopy is a medical procedure performed with an instrument called an “endoscope,” which allows observation of the interior of a hollow organ or body cavity. The absence of a centralized and automatic task management system can result in scheduling failures, leading to missed maintenance and delayed repairs. This increases the risk of medical equipment failures and affects patient safety.

The thesis/dissertation was developed in partnership with Fujifilm to develop machine learning (ML) algorithms to assist in the maintenance management of endoscopes. The goal is to identify patterns that allow determining the type of next intervention (inspection or repair), predict the number of interventions for future years, from 2025 to 2030; predict the average time until future intervention and optimize spare parts inventory management based on historical data.

The study covers the dataset of all equipment submitted for repair or inspection at the Fujifilm Service Center between 2020 and 2024. Data preprocessing techniques, statistical analysis, and predictive methods were applied, including linear regression, second-degree polynomial regression, Random Forest models, and neural networks.

The results showed that the linear regression model applied to predicting the number of annual repairs between 2025 and 2030 obtained a coefficient of determination (R^2) of 0,9722, indicating that approximately 97 % of the variation in repairs can be explained by the year. The second-degree polynomial regression model performed slightly better, with an adjusted R^2 of 0,976.

In forecasting weather between repairs, linear regression models showed more limited performance, with an adjusted R^2 of 0.1242 and, with cross-validation, approximately 0.11. The Random Forest model obtained better results, with a mean absolute error (MAE) between 122 and 123 days, and a root mean square error (RMSE) between 158 and 159 days, which is equivalent, on average, to a margin of error of about five months in predicting the time between interventions.

Regarding the prediction of the type of maintenance, the Random Forest model achieved an accuracy between 77.37 % and 78.10 %, demonstrating good classification ability. The Random Forest model obtained a mean absolute error (MAE) between 122 and 123 days and a root mean square error (RMSE) between 158 and 159 days, which is equivalent to a margin of error of about five months.

Keywords: Endoscope, Maintenance, Machine Learning, Medical Equipment Management

Agradecimentos

A conclusão do Mestrado em Engenharia Biomédica marca o fim de uma etapa importante, o que não poderia deixar de agradecer a todos os que me apoiaram durante este percurso académico.

Agradeço ao professor Luís Coelho pela orientação contínua e apoio incondicional ao longo da elaboração desta tese.

À equipa INNOVO da Fujifilm pela oportunidade de estagiar numa empresa de grande predominância em Portugal. Aos associados da INNOVO pelo auxílio e partilha de conhecimento, bem como pelo companheirismo. Um agradecimento especial ao Nuno, Miguel, Pedro Gonçalves, Eduardo, José, Ricardo, André, Tiago, João, Pedro Oliveira e Fábio.

Ao Eng. Eduardo Abreu agradeço pela orientação e auxílio ao longo do projeto e pelos conselhos valiosos.

À minha família agradeço pelo apoio, compreensão e paciência durante todo o meu percurso académico.

Aos meus colegas e amigos agradeço o acompanhamento e motivação ao longo dos dois anos do Mestrado. Gostaria de expressar o meu profundo agradecimento ao Rafael, Inês OP, Inês Magalhães, Rafaela, Vasco, Maria, AP, Gonçalo e Diogo.

Índice

1	<i>Introdução</i>	5
1.1	Enquadramento	5
1.2	Objetivos do Projeto	6
1.2.1	Objetivos Específicos	6
1.2.2	Contextualização da Abordagem	6
1.2.3	Requisitos para a Implementação do Sistema	7
1.3	Local de Desenvolvimento da Tese: Fujifilm	7
1.4	Estrutura do Relatório	10
2	<i>Manutenção de Equipamentos Médicos</i>	11
2.1	Tipos de Manutenção	11
2.2	Limpeza e Desinfecção de Endoscópios	13
2.3	Normalização	14
2.3.1	Decreto Lei nº 145/2009	15
2.3.2	Norma ISO 13485:2016	16
3	<i>Conceitos Teóricos</i>	19
3.1	Equipamento Médico: O Endoscópio	19
3.2	Machine Learning	21
3.3	A Endoscopia na Detecção de Doenças	24
4	<i>Endoscopia</i>	27
4.1	História da Endoscopia	27
4.2	Tecnologias da Endoscopia	29
4.3	Composição do Endoscópio	34
4.4	Tipos de Reparação	44
4.5	Inspeção do Endoscópio	45
5	<i>Estado da Arte</i>	51
6	<i>Materiais e Métodos</i>	57
6.1	Descrição do conjunto de dados	57
6.2	Preparação do dados	57
6.3	Pré-processamento do conjunto de dados	66
6.4	Seleção de características	67

6.5	Normalização dos dados: Z Score e Min-Max	68
7	Resultados e Discussão	71
7.1	Métodos estatísticos	71
7.1.1	Teste ANOVA e Kruskal-Wallis para a variável dependente Tempo_Entre_Reparacoes	72
7.1.2	Teste ANOVA e Kruskal-Wallis para a variável dependente Manutencao	73
7.1.3	Teste ANOVA e Kruskal-Wallis para a variável dependente Pecas	73
7.1.4	Método Mutual Information para as variáveis dependentes: Manutencao, Pecas e Tempo_Entre_Reparacoes	74
7.2	Análise Preditiva da quantidade de reparações anuais (2025 a 2030)	75
7.2.1	Regressão Linear	75
7.2.2	Regressão Polinomial de segundo grau	75
7.3	Análise Preditiva do tempo entre reparações	77
7.3.1	Modelo de regressão linear	77
7.3.2	Modelo de regressão linear com validação cruzada	78
7.3.3	Modelo Random Forest	78
7.4	Pecas mais substituídas por modelo endoscópico	80
7.5	Análise Preditiva do tipo de manutenção	81
7.5.1	Modelo Random Forest	81
7.5.2	Modelo de redes neuronais	82
7.6	Análise Preditiva das peças necessárias na futuras intervenções	83
7.6.1	Rede neuronal com a biblioteca neuralnet	83
7.6.2	Rede neuronal com a framework Keras	84
8	Considerações Finais	87
8.1	Conclusões	87
8.2	Trabalhos Futuros	88
	Bibliografia	90
	Anexos	94
	Anexo I	94
	Anexo II	97

Lista de Figuras

Figura 1 Linha do tempo da história da Fujifilm	9
Figura 2 Tipos de Manutenção	12
Figura 3 Etapas da limpeza e desinfecção de endoscópios	14
Figura 4 Órgãos que podem ser examinados em endoscopia	19
Figura 5 Tipos de endoscopia	20
Figura 6 Comparação entre as áreas cobertas pelo diagnóstico com endoscopia: (a) endoscopia superior; (b) endoscopia inferior; (c) área que não pode ser alcançada pela endoscopia superior convencional ou mesmo pela colonoscopia. Retirado de (Gounella, et al., 2023)	21
Figura 7 Aprendizagem supervisionada: (a) Regressão; (b) Classificação. Adaptado de (Yazici, Shaye, & Din, 2023)	22
Figura 8 (a) Clustering; (b) dimension reduction; (c) anomaly detection. Adaptado de (Yazici, Shaye, & Din, 2023)	23
Figura 9 Aprendizagem por reforço	24
Figura 10 Incidência e mortalidade por cancro em cada continente. Retirado de (All cancers – Fact sheet)	25
Figura 11 Incidência e mortalidade por tipo de cancro. Retirado de (All cancers – Fact sheet)	26
Figura 12 Processador da Fujifilm ELUXEO 7000	31
Figura 13 Gama de comprimentos de onda do modo LCI. Retirado de (Fujifilm)	31
Figura 14 (a) Sem a tecnologia LCI; (b) com a tecnologia LCI; Adaptado de (Fujifilm)	32
Figura 15 Gama de comprimentos de onda do modo BLI. Retirado de (Fujifilm)	32
Figura 16 (a) Modo White Light; (b) Modo BLI. Adaptado de (Fujifilm, s.d.)	33
Figura 17 Distribuição espectral dos modos de imagem endoscópica: (a) White Light; (b) LCI; (c) BLI	34
Figura 18 Ligação do endoscópio ao processador	35
Figura 19 Comportamento da Luz dentro do Endoscópio	36
Figura 20 Composição do Connetor	37
Figura 21 Connector Adaptado de (Fujifilm, s.d.)	37
Figura 22 Componentes da OSA	38
Figura 23 Intervenção terapêutica, retirado de (Fujifilm, s.d.)	39
Figura 24 Modos de Observação	39
Figura 25 Diâmetros de canal de trabalho, retirado de (Fujifilm, s.d.)	40
Figura 26 Configurações da funcionalidade de zoom, retirado de (Fujifilm, s.d.)	40
Figura 27 Mecanismo de ajuste de rigidez, retirado de (Fujifilm, s.d.)	41
Figura 28 Parte distal do endoscópio	41
Figura 29 Componentes presentes na CAP, retirado de (Fujifilm, s.d.)	42
Figura 30 Tecnologia CMOS	42
Figura 31 Endoscópio	43
Figura 32 Trolley com os processadores 7000 e um monitor	44
Figura 33 Acessórios descartáveis	47

Figura 34 Amostra do conjunto de dados.....	57
Figura 35 Pipeline pré-processamento de dados em Python	58
Figura 36 Exemplo de boxplot	63
Figura 37 Pipeline pré-processamento de dados em R	69

Lista de Tabelas

Tabela 1 Equipamento Série 700	58
Tabela 2 Quantidade de equipamentos por modelo no conjunto de dados.....	59
Tabela 3 Classificação de reparação e respetiva quantidade	61
Tabela 4 Quantidade de Reparções por ano	66
Tabela 5 Resultados da Análises Estatísticas para a variável <i>Tempo Entre Reparções</i> por Variáveis Explicativas.....	72
Tabela 6 Resultados da Análises Estatísticas para a variável Manutencao por Variáveis Explicativas	73
Tabela 7 Resultados da Análises Estatísticas para a variável Peças por Variáveis Explicativas ..	73
Tabela 8 Mutual Information para as variáveis alvo Manutencao, Pecas e Tempo_Entre_Reparacoes	74
Tabela 9 Resultados obtidos pelos modelos de regressão linear e de regressão polinomial de segundo grau	76
Tabela 10 Resultados obtidos RMSE e MAE para o estudo do tempo entre reparções	80
Tabela 11 Peças mais substituídas por modelo	80
Tabela 12 Resultados obtidos de accuracy para o estudo de previsão do tipo de reparação...	82
Tabela 13 Resultados obtidos RMSE para o estudo de previsão do tipo de reparação	83
Tabela 14 Resultados obtidos RMSE para o estudo de previsão de peças.....	84

Acrónimos e Símbolos

Lista de Acrónimos

AWT	<i>Air Water Tube</i>
BLI	<i>Blue Light Imaging</i>
BSA	<i>Bending Section Assembly</i>
CHA	<i>Camera Head Assembly</i>
DWA	<i>Drum Wire Assembly</i>
FCT	<i>Forceps Channel Tube</i>
FICE	<i>Flexible Spectral Imaging Color Enhancement</i>
FSA	<i>Flexible Section A</i>
FSB	<i>Flexible Section B</i>
ISA	<i>Insertion Section Assembly</i>
LCI	<i>Linked Color Imaging</i>
LGB	<i>Light Guide Bundle</i>
ML	<i>Machine Learning</i>
NBI	<i>Narrow Band Imaging</i>
NOZ	<i>Nozzle</i>
OSA	<i>Operation Section Assembly</i>
RBS	<i>Rubber Bending Section</i>
VE	<i>Virtual Endoscopy</i>
WJT	<i>Water Jet Tube</i>
WLI	<i>White Light Imaging</i>

1 Introdução

1.1 Enquadramento

Com o aprimoramento dos procedimentos clínicos, os equipamentos médicos hospitalares estão a tornar-se mais sofisticados, exigindo condições de funcionamento sem falhas, (Ucin, Nimpagaritse, Maru, Makhado, & Tebeka , 2024).

O ciclo de vida dos equipamentos médicos começa na conceção e no desenvolvimento, seguindo da sua fabricação, na embalagem e rotulagem, publicidade, venda, utilização e eliminação. A fase de desenvolvimento inclui o planeamento do desenvolvimento, a verificação/validação da conceção, o ensaio de protótipos e os ensaios clínicos (Jamshidi, Rahimi, & Ait-kadi, 2014).

Após o desenvolvimento e comercialização dos equipamentos médicos, como os endoscópios, é essencial assegurar a sua utilização segura e eficaz, o que implica a adoção de práticas de limpeza, desinfeção e esterilização. Estas etapas fazem parte do ciclo de vida funcional do equipamento, prolongando o tempo de vida do equipamento e garantir a segurança do paciente.

A endoscopia consiste num procedimento médico efetuado com um instrumento designado "endoscópio", o qual permite a observação do interior de um órgão oco ou de uma cavidade do corpo. A endoscopia é um procedimento médico que pode ser realizado por um médico ou por um cirurgião, dependendo da área do corpo e da natureza do procedimento. O doente pode permanecer consciente ou sob anestesia durante o procedimento, (Raghavendra & Pullaiah, 2018).

Contudo, apesar dos avanços tecnológicos e da crescente sofisticação dos equipamentos médicos, como os endoscópios, a gestão da manutenção destes equipamentos continua a ser um desafio significativo. As falhas operacionais imprevistas, frequentemente devido ao

facto de os equipamento apenas serem sujeitos a reparações quando já não estão adequados à sua utilização, resultando em interrupções nos serviços médicos. Quando tais falhas ocorrem em equipamentos críticos, como os endoscópios, os tempos de inatividade aumentam e a capacidade de realizar procedimentos médicos urgentes ou agendados fica comprometida. Além disso, essas falhas podem implicar elevados custos financeiros. A imprevisibilidade das necessidades de manutenção dos equipamentos médicos prejudica a eficiência dos hospitais, como também coloca em risco a continuidade dos cuidados de saúde e a qualidade dos mesmos. Por conseguinte, a implementação de soluções preditivas para o acompanhamento e a manutenção dos equipamentos visa minimizar os custos operacionais e os períodos de inatividade.

1.2 Objetivos do Projeto

O presente projeto visa desenvolver algoritmos de *machine learning*, para gerir a manutenção de equipamentos médicos, com foco particular em endoscópios.

1.2.1 Objetivos Específicos

- Aplicação de algoritmos de *machine learning* e a análise de dados para:
 - Prever a quantidade de intervenções para os anos futuros, de 2025 a 2030;
 - Identificar padrões que permitem a identificação do próximo tipo de intervenção, ou seja, inspeção ou reparação;
 - Prever o tempo médio até à intervenção futura;
 - Otimizar a gestão do inventário de peças de reposição com base em dados históricos.

1.2.2 Contextualização da Abordagem

- O sistema proposto pretende superar as limitações das abordagens convencionais de manutenção, nomeadamente:
 - Dependência de programação manual das intervenções de manutenção;
 - Aplicar uma gestão proativa, que visa:
 - Evitar e prever potenciais falhas;
 - Reduzir tempos de inatividade;

- Reduzir os custos operacionais;
- Ineficiência na gestão do inventário de peças sobresselentes, refletidas em excesso ou escassez, comprometendo as operações de manutenção.

1.2.3 Requisitos para a Implementação do Sistema

- Acesso a um conjunto de dados com informações detalhadas sobre os equipamentos médicos:
 - Identificação individual dos endoscópios;
 - A descrição do modelo do endoscópio;
 - Cliente responsável pelos endoscópios;
 - Registos históricos de intervenções;
 - Tipologia das intervenções realizadas;
 - Peças substituídas em cada intervenção;

1.3 Local de Desenvolvimento da Tese: Fujifilm

A empresa Fujifilm iniciou a sua atividade com a abertura de uma fábrica em Ashigara, na província de Kanagawa. A Fuji Photo Film Co. foi fundada a 20 de janeiro de 1934, tendo sido posteriormente renomeada para FUJIFILM Holdings Corporation, (90th Anniversary - History, 2024).

Em 1934, a empresa iniciou a comercialização das primeiras películas positivas para filmes e outros materiais fotossensíveis produzidos no Japão, incluindo películas para artes gráficas e papel para impressão fotográfica. Em 1936, a empresa iniciou o lançamento de produtos de forma independente, incluindo películas médicas de raios X, (90th Anniversary - History, 2024).

Desde o lançamento do produto de película de raios X em 1936, a empresa tem vindo a expandir os seus domínios de atuação através do aproveitamento das suas competências tecnológicas. A empresa oferece uma vasta gama de produtos neste domínio, incluindo informática médica, ressonância magnética e tomografia computadorizada, diagnósticos por imagem de raios X, endoscópios, diagnósticos por ultrassons e sistemas de diagnóstico *in-vitro*. A empresa também expandiu a sua atividade a nível mundial, com o objetivo de apoiar os doentes e os prestadores de serviços médicos em todo o mundo, através de diagnósticos rápidos e precisos, (Corporation, s.d.).

Em 1971, a empresa lançou endoscópios que aplicaram a sua experiência em lentes óticas ao domínio dos equipamentos médicos. Em 1972, a empresa inaugurou um novo estabelecimento destinado à produção integrada de películas de raios X para radiografia na fábrica de Fujinomiya, (90th Anniversary - History, 2024).

Durante a década de 1980, a empresa anunciou o lançamento do FCR (*Fuji Computed Radiography*), o primeiro sistema de diagnóstico por imagem de raios X digital a adotar um processamento informático. Em 1984, a Fujifilm introduziu um sistema endoscópico eletrónico, que incluiu os modelos GI superior, GI inferior e duodenal, (90th Anniversary - History, 2024) .

Em 2000, a Fujifilm introduziu os sistemas de ecografia *point-of-care*, como o Fujifilm Sonosite, que possibilitaram a realização de diagnósticos rápidos e precisos no próprio local de atendimento, (90th Anniversary - History, 2024).

Em outubro de 2006, a Fuji Photo Film Co., Ltd. procedeu à alteração da sua denominação social para FUJIFILM Holdings Corporation, adotando uma nova estrutura empresarial na qual passaria a deter duas empresas: a FUJIFILM Corporation, que sucederia à atividade da Fuji Photo Film Co., Ltd., e a Fuji Xerox Co., Ltd, (90th Anniversary - History, 2024).

Em 2008, a Fujifilm identificou três áreas de negócio: cuidados de saúde, materiais avançados e soluções documentais como os pilares da sua estratégia de crescimento no plano de gestão a médio prazo. No âmbito dos cuidados de saúde, a empresa implementou uma iniciativa com o objetivo de se tornar uma organização abrangente, contribuindo para a prevenção, diagnóstico e tratamento de doenças, (90th Anniversary - History, 2024).

Em 2010, a empresa desenvolveu e lançou endoscópios avançados, incluindo equipamentos de enteroscopia de duplo balão, uma tecnologia que revolucionou a visualização do trato gastrointestinal. Em 2015, a Fujifilm apresentou a linha Dri-Chem de analisadores automáticos de bioquímica seca, permitindo medições rápidas e precisas para diagnósticos laboratoriais, uma inovação significativa no setor de diagnóstico, (90th Anniversary - History, 2024).

Até 2014, a gestão do negócio da endoscopia em Portugal estava a cargo de distribuidores. Com o intuito de otimizar a qualidade do serviço e a declaração de lucros/perdas, foi instituído em 2015 o Centro de Serviço em Portugal da Fujifilm.

A Fuji Xerox Co., Ltd., que passou a ser totalmente detida pela FUJIFILM Holdings Corporation em 2019, procedeu à alteração da sua denominação social em abril de 2021 para FUJIFILM Business Innovation Corp, (90th Anniversary - History, 2024).

Em 2020, a FFPT (Fujifilm Portugal) criou o “Global Service Center”, um centro de equipas especializadas dedicadas à fotografia, impressão, endoscopia e diagnóstico *in vitro* (IVD). O serviço de manutenção de endoscopia denomina-se “INNOVO”, derivado da palavra “inovação”.

A FFPT adota uma abordagem proativa, em vez de implementar medidas apenas após a ocorrência de um problema, a empresa fornece formação aos clientes para mitigar a extensão do problema e minimizar as reparações substanciais. Desta forma, há uma monitorização

permanente da ocorrência de danos num curto período de tempo. Após a deteção do problema, o FSE (*Field Service Engineer*) é notificado, de modo que este forneça formação ao cliente, visando a prevenção de recorrências.

A FFPT aceita equipamentos provenientes de países europeus, designadamente França, Espanha e Itália.

Atualmente, a FFPT apresenta um mercado nacional diversificado, cerca de 48 % dos utilizadores são hospitais e clínicas do domínio público e os restantes 52 % ao setor privado. A rede de clientes da FFPT é geograficamente abrangente, com contratos com hospitais e clínicas situados em diversas regiões do país, incluindo o norte, centro, Alentejo, Algarve e Regiões Autónomas da Madeira e Açores.

Em fevereiro de 2025, a empresa Fujifilm foi reconhecida pela Liga Portuguesa Contra o Cancro (LPCC) com a distinção de "Empresa Solidária", em virtude do seu empenho e contributo relevante no apoio à luta contra o cancro. Esta distinção reflete o compromisso da Fujifilm em disponibilizar recursos tecnológicos e humanos direcionados para o rastreio de diversos tipos de cancro, nomeadamente da mama, do pulmão, do cólon e do reto. A atuação da empresa nestas áreas assenta em soluções tecnologicamente avançadas, suportadas por sistemas de Inteligência Artificial, que visam antecipar o diagnóstico oncológico, mas também otimizar o fluxo de trabalho dos profissionais de saúde e, consequentemente, promover o bem-estar da população, (Fujifilm distinguida como empresa solidária pela Liga Portuguesa Contra o Cancro, 2025).

Atualmente, a Fujifilm continua a investir substancialmente em investigação e desenvolvimento, integrando inteligência artificial e tecnologias nos seus sistemas médicos. Deste modo, a empresa visa disponibilizar soluções que atendam às necessidades médicas crescentes, expandindo o acesso a cuidados de saúde de qualidade. A Fujifilm engloba seis tipos de endoscopia: gastroscopia (endoscopia digestiva alta), colonoscopia (endoscopia digestiva baixa), enteroscopia, duodenoscopia, ultrassom e broncoscopia.

O esquema de linha do tempo apresentado a seguir sintetiza os marcos mais relevantes da história da Fujifilm, Figura 1.

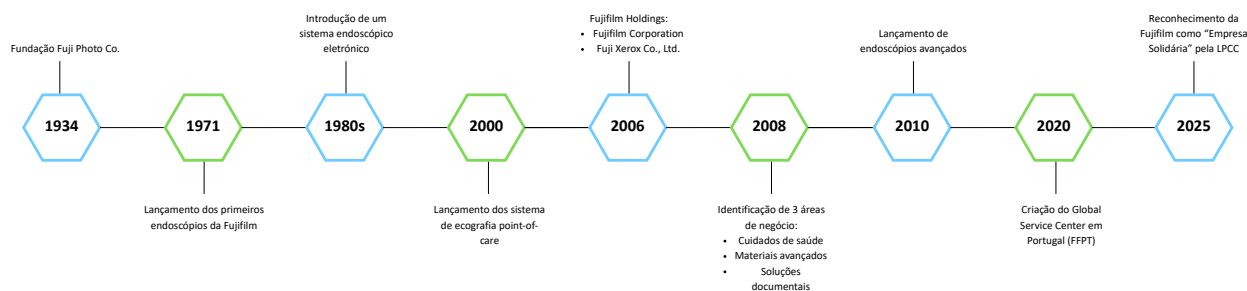


Figura 1 Linha do tempo da história da Fujifilm

1.4 Estrutura do Relatório

O relatório é dividido em oito capítulos e dois anexos. No primeiro capítulo é realizada um enquadramento da tese/dissertação, descrição dos objetivos a alcançar, contextualização da abordagem e uma breve introdução ao local de desenvolvimento (Fujifilm).

No segundo capítulo, é abordado o contexto da manutenção de equipamentos médicos, incluindo os procedimentos de limpeza e a importância da normalização no contexto de equipamentos médicos.

No terceiro capítulo, introduziu-se e explorou-se conceitos teóricos do endoscópio e fundamentos de *machine learning*.

No quarto capítulo, são descritos a evolução, tecnologia, tipos de reparação e inspeção de endoscópios.

No quinto capítulo, apresenta uma revisão do estado da arte, analisando as abordagens existentes na literatura.

No sexto capítulo, são descritos os materiais e métodos implementados, incluindo a preparação e normalização dos dados.

No sétimo capítulo, apresenta os resultados obtidos com as diferentes análises estatísticas e modelos preditivos.

No oitavo capítulo, são apresentadas as conclusões referentes ao trabalho desenvolvido e proposições de trabalhos futuros.

2 Manutenção de Equipamentos Médicos

Os equipamentos médicos são fundamentais para a prática clínica moderna, sendo imprescindíveis no diagnóstico, monitorização e tratamento de diversas condições de saúde. De acordo com o Infarmed, o termo "dispositivo médico" engloba um vasto conjunto de tecnologias e produtos que são destinados, pelo seu fabricante, a serem utilizados para fins comuns aos dos medicamentos, tais como prevenir, diagnosticar ou tratar uma doença humana.

A gestão de manutenção consiste numa abordagem estruturada e sistemática, que visa assegurar a planificação, organização, monitorização e avaliação dos processos da manutenção de modo a garantir a eficácia das operações, a otimização dos resultados e a gestão eficiente dos custos associados, (Jamshidi, Rahimi, & Ait-kadi, 2014). Uma manutenção adequada minimiza o risco de falhas inesperadas e contribui diretamente na qualidade de resultados clínicos.

Os profissionais responsáveis pela gestão dos equipamentos médicos desempenham um papel crucial, uma vez que asseguram que os equipamentos médicos se encontrem em condições de utilização e cumpram com os padrões de qualidade e segurança exigidos no setor da saúde. A gestão eficiente é fundamental para garantir a segurança dos pacientes e evitar desperdícios de recursos, prolonga a vida útil do equipamento e evita custos elevados associados a reparações não planeadas, (Khider & Hamza, 2023).

2.1 Tipos de Manutenção

No contexto dos equipamentos médicos, a manutenção pode ser de três tipos: corretiva, preventiva e preditiva, Figura 2.



Figura 2 Tipos de Manutenção

A manutenção corretiva só ocorre quando o equipamento deixa de funcionar. Em contrapartida, a manutenção preventiva é uma técnica realizada de forma periódica para prevenir falhas, com o intuito de prolongar a vida útil dos equipamentos através de calibração, substituição de peças, lubrificação, limpeza, entre outros procedimentos, (Painter, Baretich, & Velazquez-Berumen).

Ao recorrer à manutenção corretiva, as organizações evidenciam uma tendência para atrasar as ações de manutenção e assumem o risco de indisponibilidade dos seus recursos. Em contraste, a manutenção preventiva implica a antecipação das intervenções de manutenção, (Carvalho, et al., 2019).

A manutenção preditiva utiliza ferramentas de previsão para determinar o momento mais oportuno para a realização de intervenções de manutenção, com base na monitorização contínua. Esta abordagem possibilita a execução das ações de manutenção sejam executadas apenas quando necessárias, promovendo uma gestão mais eficiente dos recursos. Para tal, são utilizadas ferramentas preditivas que incluem a análise de dados históricos através de técnicas de *machine learning*, a avaliação de fatores de integridade (como sinais visuais, desgaste ou alterações de coloração), métodos de inferência estatística e abordagens de engenharia. A manutenção preditiva contribui para a redução dos tempos de inatividade e dos custos associados à substituição de peças. Além disso, ao prever falhas antes que estas ocorram, pode evitar a ocorrência de avarias dispendiosas e perdas significativas de tempo de produção provocadas por paragens imprevistas, (Carvalho, et al., 2019; Serradilla, Zugasti, Rodriguez, & Zurutuza, 2022).

Uma boa estratégia de manutenção deve melhorar o funcionamento do equipamento, reduzir as taxas de avaria e minimizar os custos, maximizando simultaneamente a vida útil do equipamento.

De acordo com a literatura sobre manutenção preditiva, é possível identificar três abordagens mais recorrentes: modelos físicos, modelos baseados em dados e modelos híbridos, (Liao & Köttig, 2016). Os métodos baseados em modelos físicos utilizam o conhecimento prévio do sistema para construir uma descrição matemática do processo de degradação. Embora estes modelos proporcionem uma interpretação clara do comportamento físico do sistema, a sua implementação em sistemas complexos revela-se difícil devido à elevada exigência de

modelação e de conhecimentos especializados, (Serradilla, Zugasti, Rodriguez, & Zurutuza, 2022).

Por outro lado, os métodos baseados em dados inferem o estado de funcionamento dos sistemas através da monitorização contínua das suas condições operacionais, recorrendo à aprendizagem obtida a partir de dados históricos. Estas abordagens incluem técnicas estatísticas (identificação de padrões), funções de fiabilidade (representação de probabilidade de um componente ou sistema continuar a funcionar sem falhas até um determinado tempo) e métodos de inteligência artificial. No entanto, a dificuldade em estabelecer uma correspondência direta entre os resultados obtidos e os fenómenos físicos subjacentes constitui uma limitação significativa, (Serradilla, Zugasti, Rodriguez, & Zurutuza, 2022).

As abordagens híbridas integram os dois paradigmas anteriormente mencionados, combinando o rigor interpretativo dos modelos físicos com a flexibilidade e capacidade preditiva dos modelos baseados em dados, (Liao & Köttig, 2016).

A inspeção pode ser realizada como uma atividade autónoma e em conjunto com a manutenção preventiva, de modo a garantir a funcionalidade dos equipamentos. Esta abordagem é importante, dado que a manutenção preventiva pode ser bastante invasiva, na medida em que envolve a remoção, limpeza ou substituição de componentes.

2.2 Limpeza e Desinfecção de Endoscópios

A limpeza inicial (pré-limpeza) dos endoscópios deve ser realizada imediatamente após cada utilização, sendo fundamental para remover quaisquer resíduos biológicos, como sangue, secreções ou tecidos, que possam ter ficado no equipamento. Com uma gaze, o exterior do equipamento é limpo e é realizada uma aspiração dos canais com água.

Antes de iniciar o processo de limpeza do equipamento, é necessário realizar um teste de fugas ou de hermeticidade, com o intuito de verificar a integridade do endoscópio. Para isso, é utilizado um insuflador e insuflar o equipamento a uma pressão interna de até $29,4 \pm 2$ kPa. No caso de a agulha do insuflador permanecer estável durante um período entre 30 e 45 segundos pode-se concluir que o equipamento não apresenta fugas. No entanto, para confirmar a existência de uma fuga, o equipamento deve ser mergulhado em água limpa enquanto ainda se encontra insuflado.

A criação contínua de bolhas de ar vindas do equipamento, enquanto submerso em água, indicam a possibilidade de existência de fuga. As fugas de ar podem provocar a falha do equipamento. A exposição de líquidos nas placas de circuito impresso pode resultar em anomalias no funcionamento do equipamento, incluindo a corrosão dos componentes metálicos, sendo possível danificar as ligações elétricas. Após a evaporação da água, é possível que esta deixe resíduos, o que implica um aumento de condutividade e, consequentemente, uma maior probabilidade de ocorrência de avarias. Na presença de fuga, o equipamento endoscópico deve ser encaminhado para as instalações da Fujifilm para fins de reparação.

Após a realização do teste de fugas, é iniciado o processo de limpeza com a utilização de um detergente sinérgico de enzima, ou seja, um detergente formulado com a presença de enzimas específicas que atuam em conjunto para a degradação de resíduos orgânicos e substâncias aderidas aos equipamentos. Para que o detergente sinérgico de enzima atinja a máxima eficácia, é essencial que o detergente seja adicionado a uma quantidade suficiente de água quente, o que favorece a ativação das enzimas e acelera o processo de limpeza. O detergente, uma vez diluído com água quente, deve ser mantido em ação por um período mínimo de quinze minutos, tempo necessário para garantir que as enzimas atuem de forma eficiente na quebra dos resíduos acumulados.

Após esse período de ativação, a solução é utilizada para proceder à limpeza dos canais do endoscópio, garantindo a remoção completa de contaminantes microscópicos. Antes de injetar o líquido, deve ser usado uma escova de limpeza para limpar as partes internas de todos os canais do endoscópio. A injeção da solução é realizada através do uso de tubos de injeção e uma seringa. É injetado, com uma seringa, 100 mL de solução detergente nos canais ar/água, no canal de sucção, no canal do jet e no canal de trabalho. De seguida, com uma gaze a superfície externa do equipamento é limpa.

Após a limpeza, os endoscópios devem ser submetidos a um processo de desinfecção, com o objetivo de eliminar os micro-organismos que podem ter sido deixados após a limpeza. Deste modo, o risco de infecção para o paciente é reduzido. A desinfecção é realizada em máquinas de desinfecção automáticas. Depois de limpar e desinfetar, todos os canais do equipamento devem ser secos para evitar a proliferação de bactérias.

A Figura 3 apresenta as etapas que devem ser seguidas após cada utilização de equipamentos endoscópicos.

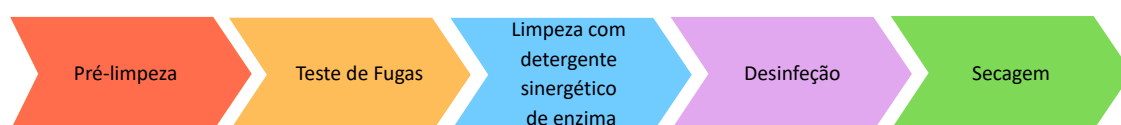


Figura 3 Etapas da limpeza e desinfecção de endoscópios

2.3 Normalização

A normalização é a atividade que, de forma organizada, viabiliza a elaboração das normas. Estabelece, face a problemas reais ou potenciais, disposições para a utilização comum e repetida, tendo em vista a otimização, num determinado contexto. A normalização abrange a elaboração, a edição e a implementação de normas e outros documentos normativos, (Qualidade, s.d.).

A Organização Internacional de Normalização (ISO, *International Organization for Standardization*) é uma organização internacional independente que fundou as normas ISO,

tendo como fim melhorar a qualidade de produtos e serviços, que são certificados quando seguem determinadas normas (ISO - International Organization for Standardization, s.d.).

2.3.1 Decreto Lei nº 145/2009

O Decreto-Lei n.º 145/2009, de 17 de junho, em Portugal, transpõe para a legislação nacional a Diretiva 93/42/CEE do Conselho da União Europeia, que regula os equipamentos médicos. Este Decreto-Lei estabelece as normas aplicáveis ao design, fabrico, comercialização, distribuição e utilização dos equipamentos médicos, com o objetivo de garantir a segurança e eficácia destes equipamentos, incluindo os endoscópios, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

Os equipamentos médicos são classificados em quatro categorias, de acordo com o nível de risco que representam para os pacientes e utilizadores: Classe I, Classe IIa, Classe IIb e Classe III. Esta classificação é determinada com base no risco potencial associado ao equipamento, sendo os equipamentos da Classe I os de menor risco e os da Classe III os de maior risco. À medida que o risco aumenta, as exigências de conformidade e avaliação tornam-se mais rigorosas. A regulamentação e os requisitos específicos para cada classe de equipamentos médicos estão definidos no Decreto-Lei n.º 145/2009, de 17 de junho, que estabelece as diretrizes para a sua certificação e monitorização, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

Os equipamentos da Classe I são classificados como de baixo risco, uma vez que, em geral, não entram em contacto invasivo com o corpo humano nem permanecem no organismo por períodos prolongados. Os termómetros, cadeiras de rodas e equipamentos de diagnóstico não invasivos são exemplos de equipamentos desta classe. A conformidade com os requisitos essenciais é verificada diretamente pelo fabricante, sem a necessidade de intervenção de um organismo notificado, à exceção dos casos de equipamentos de medição ou esterilizados. Todos estes equipamentos devem ser devidamente marcados com a marcação CE, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

A Classe IIa abrange equipamentos de risco moderado, com o potencial de causar danos significativos em caso de falha. Estes equipamentos podem entrar em contacto com o corpo de forma não invasiva ou ligeiramente invasiva, como é o caso de equipamentos de monitorização, lâmpadas de fototerapia ou equipamentos de diagnóstico. A conformidade destes equipamentos deve ser avaliada por um organismo notificado, sendo que o fabricante deve fornecer a documentação técnica que comprove o cumprimento dos requisitos de segurança e desempenho, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

Os equipamentos da Classe IIb são classificados como apresentando um risco elevado, uma vez que, geralmente, estão em contacto com o corpo de forma invasiva ou prolongada. Em caso de falha, estes equipamentos têm o potencial de causar danos graves à saúde do paciente. Exemplos incluem ventiladores mecânicos, equipamentos de infusão intravenosa e endoscópios rígidos. A avaliação de conformidade para estes equipamentos deve ser realizada por um organismo notificado, sendo exigido ao fabricante a apresentação de ensaios clínicos e informações técnicas detalhadas que comprovem a segurança, eficácia e desempenho do

equipamento, em conformidade com os requisitos regulamentares aplicáveis, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

Os equipamentos da Classe III são classificados como de alto risco, devido ao seu elevado potencial de causar danos graves ou fatais ao paciente, especialmente quando entram em contacto invasivo com o corpo de forma prolongada ou permanente. Exemplos típicos desta classe incluem equipamentos implantáveis, como pacemakers, stents coronários e próteses cardíacas. Estes equipamentos estão sujeitos a uma avaliação de conformidade mais rigorosa, que abrange ensaios clínicos extensivos e a apresentação de uma documentação técnica completa, incluindo auditorias regulares para garantir a sua conformidade contínua. A avaliação de conformidade envolve, ainda, uma inspeção clínica detalhada por um organismo notificado, conforme estipulado pelo Decreto-Lei n.º 145/2009, de 17 de junho, que estabelece os requisitos e processos regulamentares para a autorização e monitorização de equipamentos médicos de alto risco, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

A conformidade legal dos equipamentos médicos é garantida pela aplicação da marcação CE, que atesta o cumprimento das exigências legais da União Europeia. A obtenção da marcação CE é condicionada à realização de uma avaliação de conformidade, que pode incluir a execução de testes e auditorias por um organismo notificado. O procedimento de avaliação é diferenciado em função da classe do equipamento, sendo mais rigoroso para as classes superiores (IIb e III), (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

O Decreto-Lei n.º 145/2009 estabelece as obrigações de manutenção preventiva, visando assegurar a segurança e o desempenho adequado dos equipamentos médicos durante o seu ciclo de vida útil, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

A manutenção preventiva pode incluir inspeções periódicas, identificando qualquer sinal de desgaste ou falha iminente. Este procedimento permite evitar falhas inesperadas durante os procedimentos médicos. Adicionalmente, os testes de funcionalidade, que incluem a realização de testes de desempenho, são essenciais para garantir que o equipamento, como um endoscópio, continue a fornecer imagens nítidas e precisas e que todos os seus componentes funcionem corretamente, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

Os equipamentos médicos reutilizáveis, tais como os endoscópios, exigem uma atenção particular no que diz respeito ao reprocessamento. Este processo, definido como o conjunto de procedimentos destinados a limpar, desinfetar e, em certos casos, esterilizar os equipamentos para assegurar a sua descontaminação antes da reutilização, é regulado pelo Decreto-Lei, (Decreto-Lei n.º 145/2009, de 17 de junho, 2009).

2.3.2 Norma ISO 13485:2016

A segurança e o desempenho dos equipamentos médicos são regulamentados por requisitos essenciais, aplicáveis a todos os equipamentos colocados no mercado. Estes requisitos visam garantir que os equipamentos não apresentem riscos para a saúde e segurança dos pacientes

ou dos utilizadores e que cumpram a sua função com eficácia. Adicionalmente, a sua produção deve estar em conformidade com normas de qualidade, tais como a ISO 13485, que regula os sistemas de gestão da qualidade para equipamentos médicos, (International Organization for Standardization, 2016).

A norma internacional ISO 13485:2016 estabelece os requisitos para um Sistema de Gestão da Qualidade (SGQ) aplicável ao setor dos equipamentos médicos, incluindo as inspeções e manutenção necessárias para garantir que os equipamentos médicos atendam aos requisitos de segurança, qualidade e desempenho ao longo de todo o seu ciclo de vida. A norma em questão foi concebida com o intuito de auxiliar as organizações na implementação de um SGQ que contemple a segurança, a qualidade e a eficácia dos equipamentos médicos ao longo do seu ciclo de vida, abrangendo desde a conceção e desenvolvimento até à produção e atividades pós-comercialização, (International Organization for Standardization, 2016).

A norma ISO 13485:2016 estabelece que as organizações devem assegurar a realização de inspeções e testes adequados aos equipamentos médicos em diferentes etapas do seu ciclo de vida, abrangendo as fases de conceção, fabrico e pós-comercialização. Estas inspeções contribuem para a verificação do cumprimento dos requisitos de segurança, desempenho e qualidade estabelecidos, (International Organization for Standardization, 2016). A FUJIFILM Wako Chemicals Europe GmbH, fornecedor de produtos químicos e reagentes, também obteve a certificação pela mesma norma, (Quality management, s.d.).

Os equipamentos médicos devem ser submetidos a várias inspeções ao longo do processo de fabrico, com o objetivo de garantir a conformidade com as especificações estabelecidas, (International Organization for Standardization, 2016).

Durante as fases de conceção e desenvolvimento, os fabricantes devem estabelecer procedimentos controlados para verificar e validar a conceção do equipamento médico, garantindo assim que esta cumpre os requisitos regulamentares, as normas de segurança e as especificações de desempenho antes de avançar para a produção, (International Organization for Standardization, 2016).

As atividades pós-comercialização, incluindo inspeções, são necessárias para garantir que os equipamentos continuam a cumprir as normas de segurança e desempenho após a sua distribuição. As inspeções podem basear-se em relatórios de utilizadores, reclamações ou resultados da vigilância de produtos. As inspeções devem ser efetuadas em conformidade com as normas predefinidas, a fim de garantir que o produto cumpre os requisitos regulamentares e de segurança, (International Organization for Standardization, 2016).

A norma ISO 13485:2016 destaca a necessidade de estabelecer critérios claros para a realização de calibrações e manutenção para garantir que os equipamentos médicos mantenham um desempenho adequado. As organizações devem implementar programas de manutenção preventiva para garantir que os equipamentos médicos permaneçam em condições ideais de funcionamento durante a vida útil. A manutenção preventiva pode incluir inspeções regulares para detetar desgaste, falhas potenciais ou necessidade de ajustes para evitar falhas no

momento do uso. A frequência e os procedimentos de manutenção devem ser documentados com base nas recomendações do fabricante e na experiência operacional do equipamento, (International Organization for Standardization, 2016).

Equipamentos médicos que exigem precisão nas medições, como termômetros, monitores de pressão arterial e equipamentos de imagem, devem ser submetidos a calibrações regulares. A calibração assegura que o equipamento esteja a operar de acordo com os padrões de precisão exigidos. A norma exige que as calibrações sejam realizadas de acordo com procedimentos documentados e que as evidências de calibração sejam registradas para futura auditoria ou inspeção, (International Organization for Standardization, 2016).

A norma ISO 13485:2016 exige que todos os procedimentos de inspeção e manutenção sejam adequadamente documentados. Essa documentação é fundamental para garantir a rastreabilidade e a conformidade com as normas regulatórias. Quando as inspeções identificam não conformidades, a norma ISO 13485:2016 exige que o fabricante tome ações corretivas para resolver os problemas identificados, (International Organization for Standardization, 2016).

Apesar de a norma ISO 13485:2016 não especificar frequências precisas para as inspeções, destaca a importância de implementar um sistema robusto de inspeções, testes e calibrações ao longo do ciclo de vida dos equipamentos médicos. Os fabricantes são obrigados a estabelecer critérios de inspeção claros, documentar os resultados e realizar inspeções contínuas, incluindo a vigilância pós-comercialização. Ademais, a calibração, a manutenção e as ações corretivas regulares são fundamentais para assegurar que os equipamentos médicos estejam em conformidade com as normas de segurança, qualidade e regulamentação. Ao aderirem a estas diretrizes, os fabricantes podem garantir que os seus equipamentos médicos permanecem seguros e eficazes para os utilizadores, simultaneamente ao cumprimento das expectativas regulamentares internacionais, (International Organization for Standardization, 2016).

3 Conceitos Teóricos

3.1 Equipamento Médico: O Endoscópio

O endoscópio é um instrumento médico que pode ser introduzido no corpo humano para permitir a visualização das partes internas, tais como o esófago, estômago, duodeno, intestino delgado, cólon, bexiga, pulmão e brônquios, Figura 4.

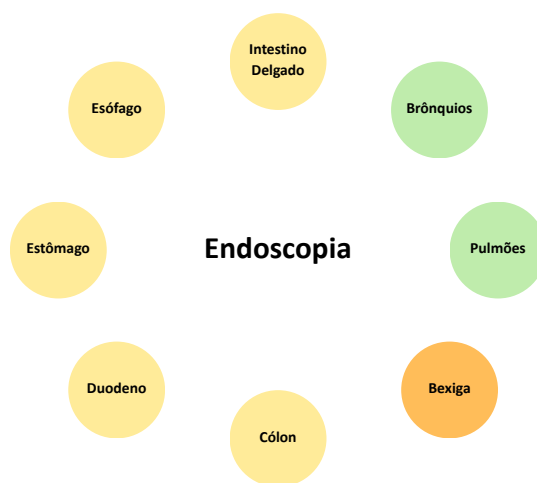


Figura 4 Órgãos que podem ser examinados em endoscopia

Um sistema endoscópico pode ser projetado como um equipamento exclusivamente ótico, contendo lentes ou fibras transparentes, ou ainda como uma combinação desses elementos com câmaras de vídeo integradas ou adicionais. A extremidade do endoscópio é introduzida através de uma pequena incisão feita na pele ou por meio de aberturas naturais do corpo, como a boca ou o reto, dependendo da área a ser examinada. Para garantir uma visualização adequada, é usada uma fonte de luz para iluminar a cavidade interna. O endoscópio, quando equipado com câmaras de vídeo, capta imagens que são posteriormente transmitidas para um monitor, onde podem ser visualizadas em tempo real e também gravadas para efeitos de documentação do procedimento ou de análise posterior, (Boese, et al., 2022).

A visualização interna proporcionada pela endoscopia permite aos médicos diagnosticar condições como infeções, inflamações e lesões, facilitando a observação direta dos órgãos e tecidos afetados. Adicionalmente, esta técnica permite a obtenção de amostras de tecido para subsequente análise patológica, nomeadamente através de biópsias, contribuindo para a identificação de patologias como o cancro ou para a avaliação de outras condições clínicas. A capacidade de examinar visualmente o interior do corpo de forma minimamente invasiva também contribui para diagnósticos mais rápidos e precisos, além de permitir a realização de tratamentos simultâneos, como a remoção de pólipos ou a cauterização de áreas danificadas, (Browning, Cloutier, Rich, & Leavesley, 2022).

Os endoscópios estão ligados a sistemas de imagem digital que permitem a visualização em ecrãs e a documentação fotográfica ou em vídeo, (Boese, et al., 2022).

Os endoscópios médicos são normalmente designados de acordo com o domínio da sua aplicação, órgão ou procedimento, (Boese, et al., 2022).

Existem vários tipos de endoscopia, cada um direcionado a áreas específicas do organismo, de modo a proporcionar uma avaliação detalhada. Entre os procedimentos mais comuns destacam-se a gastroscopia, colonoscopia, enteroscopia, duodenoscopia, endoscopia por ultrassons e broncoscopia, Figura 5.



Figura 5 Tipos de endoscopia

A Esofagoscopia, a Gastroscopia e a Duodenoscopia constituem um subgrupo da endoscopia gastrointestinal. Os procedimentos descritos são realizados com o objetivo de diagnosticar e tratar doenças a nível do sistema digestivo superior. Os gastroscópios são introduzidos pela cavidade oral permitindo visualizar o esófago, o estômago e o duodeno, (Boese, et al., 2022).

A colonoscopia é um procedimento médico utilizado para a observação do cólon. O colonoscópio consiste numa sonda tubular alargada e flexível. A sua introdução ocorre através do reto, podendo avançar até ao final do intestino grosso e parcialmente no intestino delgado. Além da imagiologia, os pólipos ou outras lesões podem ser removidos durante o exame, através da realização de ressecções endoscópicas que não estão limitadas pelo diâmetro do canal de trabalho, (Boese, et al., 2022).

A ecografia endoscópica (EUS, *Endoscopic ultrasound*) é um procedimento que combina a endoscopia e a ecografia para criar imagens do trato digestivo e dos órgãos e tecidos adjacentes. Um aparelho de ultrassons na ponta do tubo utiliza ondas sonoras de alta frequência para criar imagens detalhadas do tubo digestivo e de outros órgãos e tecidos. Estes incluem os pulmões, o pâncreas, a vesícula biliar, o fígado e os gânglios linfáticos.

A broncoscopia é um procedimento que permite aos médicos observar os pulmões e as vias respiratórias.

A Figura 6 ilustra os dois tipos de procedimentos endoscópicos, endoscopia superior e inferior. É possível observar que ainda existem algumas áreas onde a endoscopia convencional não é capaz de alcançar. Esta limitação impede a detecção de anomalias comuns, tais como hemorragias, úlceras e tumores, quando estes se encontram em estágios iniciais, (Gounella, et al., 2023).

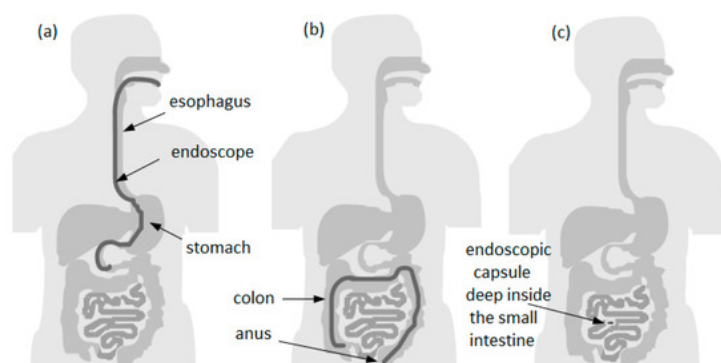


Figura 6 Comparação entre as áreas cobertas pelo diagnóstico com endoscopia: (a) endoscopia superior; (b) endoscopia inferior; (c) área que não pode ser alcançada pela endoscopia superior convencional ou mesmo pela colonoscopia. Retirado de (Gounella, et al., 2023)

3.2 Machine Learning

A *Machine learning* é uma subárea da Inteligência Artificial. Pode ser definida como um conjunto de algoritmos capazes de aprender de forma autónoma, com pouca ou nenhuma intervenção humana adicional. Os algoritmos de *Machine learning* desempenham um papel fundamental na resolução de diversos problemas complexos, como a análise de grandes volumes de dados (*big data*), o reconhecimento de imagens, o reconhecimento de voz e o desenvolvimento de sistemas robóticos.

A *Machine learning* pode ser classificada em três tipos principais: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço.

Na aprendizagem supervisionada, o conjunto de dados de entrada e saída são conhecidos. A aprendizagem supervisionada é uma técnica de *machine learning* orientada por tarefas, cujo objetivo é estabelecer uma correspondência entre entradas e saídas com base em dados rotulados. Este paradigma requer um conjunto de dados de treino previamente anotado, que permite ao modelo aprender a realizar previsões. O processo consiste na análise de pares input-

output para ajustar os parâmetros do modelo, de modo que este seja capaz de generalizar para novos dados, (Yazici, Shayea, & Din, 2023).

As duas principais tarefas neste contexto são a regressão, Figura 7(a) que lida com previsões de valores contínuos, e a classificação, Figura 7(b), que se foca na atribuição de categorias discretas a dados de entrada. Ao longo do processo de treino, o modelo é iterativamente ajustado até se obter uma função que represente a relação entre as variáveis de entrada e de saída. Uma vez treinado, o modelo pode ser utilizado para prever a saída correspondente a novas entradas. Entre os algoritmos mais utilizados neste tipo de aprendizagem incluem-se *Artificial neuronal networks*, (ANN), *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Random Forest* (RF), a regressão linear (LR, *linear regression*) e a regressão logística (LGR, *logistic regression*), (Yazici, Shayea, & Din, 2023).

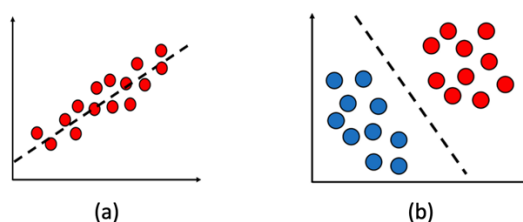


Figura 7 Aprendizagem supervisionada: (a) Regressão; (b) Classificação. Adaptado de (Yazici, Shayea, & Din, 2023)

Na aprendizagem não supervisionada, apenas as variáveis de entrada são conhecidas, sendo o objetivo identificar padrões sem qualquer supervisão externa. Ao contrário da aprendizagem supervisionada, não existe uma correspondência direta entre entradas e saídas, nem qualquer supervisão externa durante o processo de treino. O modelo deve, por isso, inferir autonomamente relações entre os dados. Entre os principais tipos de problemas tratados neste domínio destacam-se a *clustering* (agrupamento), Figura 8 (a), a redução de dimensionalidade, Figura 8 (b), e a deteção de anomalias, Figura 8 (c). O agrupamento consiste em organizar instâncias não classificadas em subconjuntos com base em medidas de similaridade, que podem variar consoante o domínio de aplicação e os objetivos do estudo. No entanto, a interpretação dos grupos formados pode ser complexa, exigindo frequentemente conhecimento prévio sobre o contexto dos dados. A avaliação do desempenho da aprendizagem não supervisionada tende a ser mais subjetiva e dependente do domínio, em comparação com a aprendizagem supervisionada, (Yazici, Shayea, & Din, 2023).

A redução de dimensionalidade é a etapa de pré-processamento para remover características redundantes, ruídos e dados irrelevantes, a fim de melhorar a precisão das características de aprendizagem e reduzir o tempo de treino, (Velliangiri, Alagumuthukrishnan, & Joseph, 2019).

A deteção de anomalias consiste na identificação de observações, eventos ou pontos de dados que se desviam do padrão ou do esperado, tornando-os inconsistentes com o resto do conjunto de dados.

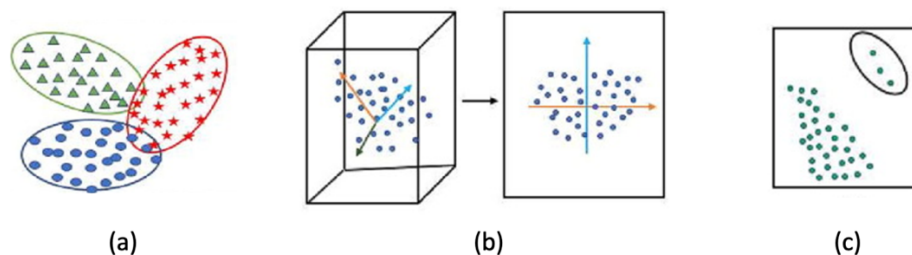


Figura 8 (a) Clustering; (b) dimension reduction; (c) anomaly detection. Adaptado de (Yazici, Shaye, & Din, 2023)

A aprendizagem por reforço difere dos outros tipos de aprendizagem pelo seu modo de aprendizagem. Em vez de utilizar dados rotulados ou não rotulados, o agente aprende através da interação com o ambiente por um mecanismo de tentativa e erro, com o objetivo de maximizar a recompensa acumulada esperada, Figura 9. Em várias situações, os algoritmos de aprendizagem por reforço são modelados como Processos de Decisão de Markov (MDP), que correspondem a processos estocásticos de controlo em tempo discreto. Existem cinco termos fundamentais na aprendizagem por reforço: agente, estado, ambiente, ação e recompensa. O agente recebe um estado inicial S_0 e realiza uma ação A_0 no ambiente com base nesse estado. De seguida, o ambiente transita para um novo estado S_1 e atribui uma recompensa R_1 ao agente. O objetivo do agente é maximizar a recompensa acumulada, (Yazici, Shaye, & Din, 2023).

Um algoritmo de aprendizagem é composto por quatro componentes essenciais: política, função de recompensa, função de valor e ambiente. As políticas são responsáveis por mapear estados para as ações executadas pelo agente. A função de recompensa avalia os estados atuais e atribui penalizações ou recompensas em função do resultado da ação. A função de valor, que pode ser do tipo valor-estado ou valor-ação, estima a recompensa esperada a longo prazo a partir de um determinado estado ou ação. O ambiente corresponde à tarefa ou simulação onde o agente tenta maximizar a recompensa acumulada recorrendo ao mecanismo de tentativa e erro, (Yazici, Shaye, & Din, 2023).

O agente interage com o ambiente, deteta o seu estado atual e o estado do ambiente, aprendendo e recolhendo informação para realizar ações adequadas. Assim, o agente percebe o estado exato do ambiente em cada instante e escolhe uma ação que provoca a transição para um novo estado. Existe um sistema de recompensa-punição consoante a ação selecionada e o seu resultado: se a ação for considerada boa, o agente recebe uma recompensa; caso contrário, é penalizado. Este algoritmo permite que o agente descubra as ações corretas apenas através do processo de tentativa e erro, sem qualquer feedback explícito sobre o seu desempenho, (Yazici, Shaye, & Din, 2023).

Em cada passo t , o agente observa o estado S_t e escolhe uma ação A_t do espaço de ações A . Em seguida, recebe uma recompensa escalar r_t que indica a qualidade da ação escolhida, e transita para o próximo estado S_{t+1} . O algoritmo utiliza a experiência composta pelo conjunto (S_t, A_t, R_{t+1}) .

S_{t+1}) para aprender a mapear estados a uma medida do valor a longo prazo de se encontrar naquele estado, designada função de valor ótimo, (Yazici, Shaya, & Din, 2023).

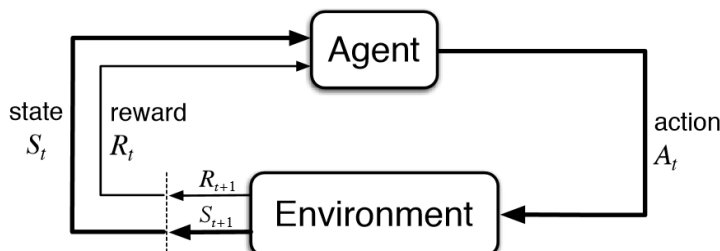


Figura 9 Aprendizagem por reforço

3.3 A Endoscopia na Detecção de Doenças

De acordo com (Rutter, et al., 2015), anualmente, dezenas de milhões de pessoas são submetidas a procedimentos endoscópicos na Europa. O mercado global de procedimentos endoscópicos foi estimado de 191,1 milhões em 2024 e deve atingir os 207,5 milhões de procedimentos até 2030. Fatores como a crescente adoção de procedimentos endoscópicos, o aumento dos gastos com saúde e o envelhecimento populacional podem impulsionar o crescimento do mercado, (Endoscopy Procedures Estimates Market (2025 - 2030)). Em 2024, verificou-se um predomínio do mercado da endoscopia gastrointestinal, representando 69,8 % do volume. O mercado é impulsionado por fatores como o envelhecimento populacional, o aumento das doenças crônicas e do número de lesões traumáticas que afetam o sistema nervoso central e a preferência por procedimentos minimamente invasivos.

O cancro constitui um dos principais desafios atuais em termos sociais, de saúde pública e económicos. Atualmente, é responsável por aproximadamente uma em cada seis mortes a nível global (16,8 %) e por cerca de uma em cada quatro mortes (22,8 %) atribuídas a doenças não transmissíveis, (Bray , Laversanne, Weiderpass, & Soerjomataram, 2021).

Segundo a Organização Mundial de Saúde (OMS), em 2022, foram diagnosticados 69.567 novos casos de cancro e registadas 33.762 de mortes associadas a esta doença, em Portugal. Do total de novos casos, cerca de 38 mil ocorreram em homens, sendo os tipos mais comuns o cancro da próstata, colorretal e do pulmão. Entre as mulheres, os tipos mais frequentes foram o cancro da mama, colorretal e do pulmão. Já a nível mundial, foram diagnosticados cerca de 19 milhões de novos casos e 9 milhões de mortes associadas a cancro. As estimativas sugerem que aproximadamente um em cada cinco homens ou mulheres desenvolvem cancro ao longo da vida, enquanto cerca de um em cada nove homens e uma em cada 12 mulheres morrem da doença, (All cancers – Fact sheet).

A incidência de cancro e a mortalidade resultante têm vindo a aumentar a nível global, refletindo o envelhecimento populacional e o impacto de diversos fatores de risco, incluindo o

consumo de tabaco, álcool, uma alimentação pouco saudável, a inatividade física e a poluição atmosférica, (Chen, Cao, & Prettner, 2023).

Para além de constituir um obstáculo significativo ao aumento da esperança média de vida, o cancro acarreta custos sociais e económicos consideráveis. Estes custos variam de forma substancial consoante o tipo de cancro, a região geográfica e o sexo dos indivíduos afetados, refletindo desigualdades estruturais no acesso à prevenção, diagnóstico precoce e tratamento adequado. O impacto económico inclui gastos diretos com cuidados de saúde, as perdas de produtividade e os encargos para as famílias e os sistemas de proteção social, (Chen, Cao, & Prettner, 2023).

A incidência e mortalidade por cancro não são distribuídas de forma homogênea entre as regiões, de acordo com a Figura 10. Observa-se que a Ásia concentra aproximadamente metade dos casos (49,2 %) e mais da metade das mortes (56,1 %), o que pode refletir fatores como acesso limitado a serviços de saúde, diagnóstico tardio e menor cobertura de programas de prevenção. As disparidades regionais evidenciam as lacunas existentes entre países e regiões no que respeita aos determinantes sociais da saúde, infraestruturas de cuidados oncológicos e políticas de saúde pública, (All cancers – Fact sheet).

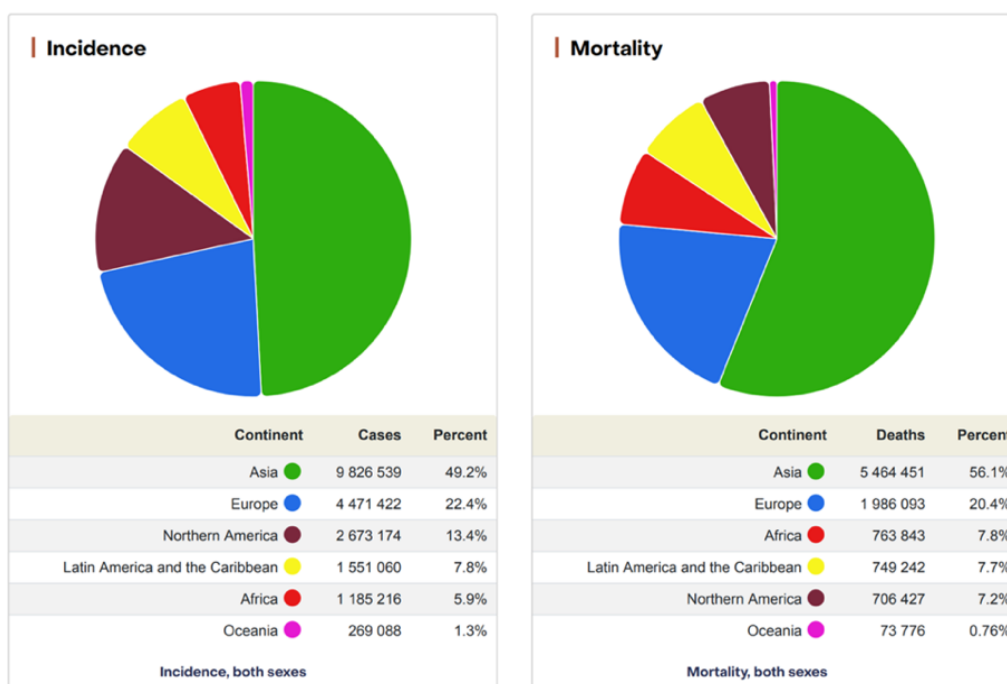


Figura 10 Incidência e mortalidade por cancro em cada continente. Retirado de (All cancers – Fact sheet)

Os cancros gastrointestinais constituem um dos principais encargos da saúde pública na Europa, representando aproximadamente um quarto de todos os tumores malignos diagnosticados no continente, Figura 11. Este impacto é intensificado pelo envelhecimento progressivo da população europeia e pela exposição a diversos fatores de risco ambientais. De acordo com estimativas recentes, registam-se anualmente cerca de 600 mil novos casos de cancro

gastrointestinal na União Europeia, resultando em aproximadamente 360 mil mortes atribuídas a esta patologia. Estes números evidenciam a elevada carga associada aos câncros GI, tanto em termos de morbilidade como de mortalidade, sublinhando a necessidade de estratégias eficazes de prevenção, diagnóstico precoce e tratamento, (Săftoiu, et al., 2020).

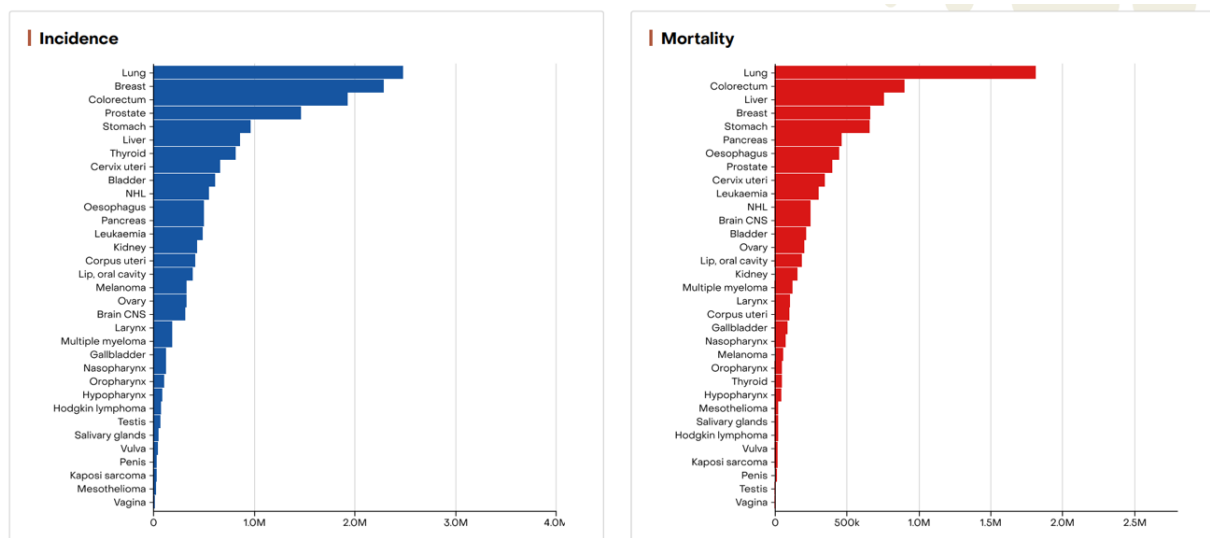


Figura 11 Incidência e mortalidade por tipo de cancro. Retirado de (All cancers – Fact sheet)

4 Endoscopia

A endoscopia é uma técnica essencial na medicina moderna, permitindo a visualização das cavidades internas do corpo para diagnóstico e tratamento.

4.1 História da Endoscopia

A endoscopia moderna começou no início do século XIX. Hipócrates foi o primeiro a descrever a utilização da endoscopia na época do Império Romano. Na prática médica, Asclepius e Hipócrates centraram-se na abordagem natural e no tratamento de doenças, destacando a importância de compreender a saúde do paciente, a independência da mente e a necessidade de harmonia entre o ambiente individual, social e natural, (Kleisiaris , Sfakianakis, & Papathanasiou , 2014).

Os espéculos eram frequentemente utilizados para a exploração do interior do corpo humano. No entanto, a utilização da luz ambiente apresentava frequentemente limitações no que diz respeito à capacidade de observação. A primeira grande melhoria foi implementada por Philipp Bozzini, um médico alemão. Este desenvolveu, em 1806, um instrumento em forma de funil, constituído por uma fonte de luz externa, que consistia numa vela e num espelho (Lichtleiter), (Marcelissen & Vijgen, 2023). A iluminação era mantida constante, colocando a vela numa mola dentro da caixa, de modo que, à medida que a vela ardia, a mola mantém a chama na mesma posição. O reflexo do espelho era utilizado para iluminar orifícios pouco profundos para exame e manipulação. Como a uretra era a cavidade mais facilmente acessível desta forma, foi a primeira a ser examinada, (Browning, Cloutier, Rich, & Leavesley, 2022).

Desormeaux concluiu que, quando aplicada corretamente, uma lente tinha a capacidade de condensar a luz num ponto mais estreito e mais brilhante. Este avanço representou uma mudança significativa na abordagem da endoscopia, uma vez que anteriormente se acreditava que o tubo endoscópico deveria ser preenchido com algum tipo de prisma, espelho ou refletor

para ultrapassar o ângulo formado pelo eixo horizontal da boca e o eixo vertical da faringe e do esôfago. Assim, em 1855, Antonin Desormeaux, propôs a utilização de um sistema de espelhos côncavos a 45 graus e lentes com uma mistura de terebintina alcoólica como fonte de luz em substituição da vela. No entanto, estas fontes de luz só eram práticas para iluminar tecidos internos a pequenas profundidades, (Marcelissen & Vijgen, 2023).

Diversas inovações foram desenvolvidas para resolver o problema da baixa intensidade luminosa em endoscopia. Uma das soluções iniciais foi o uso de um fio de magnésio em combustão, que gerava uma luz intensa e brilhante. No entanto, esse método apresentava inconvenientes, como o excesso de calor e fumo, o que tornava a sua utilização impraticável. Em 1860, foi obtida a primeira imagem endoscópica por Nepomuk Czermak, tendo sido capturada a partir da sua própria laringe, (Marcelissen & Vijgen, 2023).

O subsequente progresso foi realizado pelo urologista alemão Maximilian Nitze, que substituiu a técnica de iluminação baseada em querosene, por um fio de platina aquecido eletricamente, nomeado de cistoscópio Nitze-Leiter. Apesar da iluminação ser adequada, foram encontradas grandes dificuldades com o considerável calor gerado, sendo necessário um sistema de arrefecimento por água e a utilização de pilhas como fonte de energia. O fio de platina foi mais tarde substituído por uma lâmpada elétrica após a invenção de Thomas Edison, em 1880, resultando numa melhor iluminação e redução das lesões térmicas, (Marcelissen & Vijgen, 2023).

No período compreendido entre 1900 e 1950, o desenvolvimento das técnicas endoscópicas foi limitado. No entanto, houve alguns progressos notáveis, devido às significativas contribuições de Harold Hopkins para o campo da ótica, que resultaram na inovação da evolução da endoscopia. Em 1948, Hopkins criou a lente de zoom, um avanço ótico que mais tarde seria incorporado aos endoscópios, permitindo um ajuste mais preciso da ampliação. Em 1959, criou um sistema de lentes de haste, que representava uma inversão do design original de Nitze. Esse novo sistema utilizava os pequenos espaços entre as varetas de vidro para criar "lentes de ar", o que auxiliou na redução na perda de luz por refração, aprimorando a qualidade das imagens. Essa inovação foi crucial para a melhoria da iluminação e visualização em endoscopia, (Browning, Cloutier, Rich, & Leavesley, 2022).

Em 1952, Fourstier, Bladu e Valmier tornaram a fonte de luz de um endoscópio novamente externa devido ao calor gerado pela iluminação distal iniciada por Bruck e Nitze. Tal foi conseguido através da transmissão de luz por uma haste de quartzo no endoscópio rígido, (Browning, Cloutier, Rich, & Leavesley, 2022).

Devido às trajetórias de luz mais longas, tornou-se essencial que a transmissão de luz alcançasse o nível máximo de eficiência. Assim, surgiu a terceira conquista endoscópica de Hopkins, o feixe de fibras. Originalmente concebido por Heinrich Lamm, o feixe de fibras consistia num conjunto de fibras de sílica flexíveis que permitiam a transmissão de luz de uma extremidade à outra. Em 1954, Harold Hopkins, em conjunto com Narinder Kapany, aplicou a técnica de feixes de fibras para a iluminação em endoscopia, marcando um avanço significativo no campo. Foram desenvolvidos dois tipos de feixes: feixes incoerentes, nos quais a orientação das fibras não era

relevante, e os feixes coerentes, onde a orientação das fibras era crucial para a transmissão de imagens, (Browning, Cloutier, Rich, & Leavesley, 2022).

4.2 Tecnologias da Endoscopia

Atualmente, as quatro técnicas endoscópicas principais mais amplamente utilizadas são: a endoscopia de luz branca (WLI, *White Light Imaging*), a imagiologia de banda estreita (NBI, *Narrow-Band Imaging*), o realce de cor da imagiologia espectral flexível da Fujifilm (FICE, *Flexible Spectral Imaging Color Enhancement*) e a endoscopia virtual (VE, *Virtual Endoscopy*). A NBI e o FICE constituem duas modalidades complementares à WLI que podem proporcionar um contraste melhorado das estruturas dos tecidos entre a mucosa e as lesões, (Browning, Cloutier, Rich, & Leavesley, 2022).

A WLI é a técnica de referência utilizada para captar uma imagem RGB (*Red-Green-Blue*) padrão que fornece imagens baseadas na refletância da parede luminal (o espaço interior de um órgão tubular, como um vaso sanguíneo ou um intestino). Contudo, esta técnica pode não detetar, de forma eficaz, cancro gastrintestinal.

As lesões mucosas malignas apresentam um aumento da angiogénese, o que se traduz frequentemente num aumento significativo do número de vasos sanguíneos. Dado que o propósito principal da realização de um procedimento endoscópico é a deteção de lesões malignas, a técnica WLI pode apresentar dificuldades a identificar esses padrões, (Lin, Karakasheva, Hicks, Bass, & Rustgi, 2016).

A tecnologia NBI utiliza luz azul e verde para iluminar as cavidades corporais para aproveitar a absorção de luz vascular nestes comprimentos de onda, fornecendo um contraste adicional à imagem. Esta tecnologia utiliza comprimentos de onda de luz azul (415 nm) e verde (540 nm). Uma vez que a faixa de absorção da hemoglobina se encontra nesses valores de comprimentos de onda, a luz é absorvida pelos tecidos que contêm sangue, enquanto se reflete no tecido mucoso circundante. A luz azul, ao apresentar um comprimento de onda mais curto, permite uma penetração mais superficial da mucosa. Quando a luz azul e verde entra em contacto com a superfície do tecido, essa luz é absorvida pelos vasos sanguíneos, mas refletida pelo tecido circundante. Pequenos tumores são normalmente envoltos por um número elevado de vasos sanguíneos. Contudo, esta técnica apresenta certas limitações, como a geração de imagens de baixa luminosidade para lesões localizadas em regiões distantes de órgãos luminiais de grande dimensão, como o estômago, e uma visualização insuficiente da microestrutura da mucosa na superfície do tumor. Essas restrições comprometem a eficácia da NBI no contexto da endoscopia de rastreio, (Boeriu, et al., 2015).

O FICE produz imagens de alto contraste de lesões gastrointestinais. É um processo de pós-aquisição que divide a imagem RGB nas três cores respetivas e altera digitalmente os comprimentos de onda para aumentar o contraste.

O VE recorre à tomografia computadorizada (TAC) ou à ressonância magnética (RM) para apresentar um modelo bidimensional (2D) ou tridimensional (3D) das cavidades ocas, (Browning, Cloutier, Rich, & Leavesley, 2022) (Osawa, et al., 2018).

Os atuais sistemas de endoscópios recorrem a uma combinação de fontes de luz de banda larga (lâmpadas de arco de xénon, halogéneos ou LEDs), filtros de largura de banda e análises digitais para produzir WLI, NBI ou FICE.

Embora as técnicas atuais disponibilizem diversos modos complementares de visualização das cavidades internas do corpo, o contraste e a definição entre o tecido saudável e o tecido afetado são limitados, especialmente no cancro em fase inicial. Por conseguinte, é imperativo desenvolver tecnologias que proporcionem um elevado contraste e a capacidade de visualizar a neoplasia ou o cancro em fase inicial durante a visualização endoscópica dos órgãos ocas.

A obtenção de resultados diagnósticos e terapêuticos em procedimentos endoscópicos depende fortemente da qualidade das imagens geradas. Nesse contexto, foi desenvolvida a tecnologia *Multi Light*, com o objetivo de superar as limitações das técnicas FICE e NBI, proporcionando uma visualização aprimorada e facilitando o rastreio do trato gastrointestinal. Esta tecnologia integra os modos de observação *Linked Color Imaging* (LCI) e *Blue Light Imaging* (BLI), que se distinguem pela capacidade de aumentar o contraste, especialmente na visualização de lesões malignas, (Fujifilm) (Osawa, et al., 2018).

Os modos LCI e BLI utilizam intensidades de emissão elevadas em comprimentos de onda curtos, especialmente na faixa de 410 nm, o que resulta num contraste de cor elevado ao amplificar as diferenças na refletância espectral entre os tecidos com diversas alterações histológicas. O LCI é gerado a partir do processamento computacional de imagens do modo BLI-*bright*. Essa técnica favorece a visualização de pequenas variações nas cores, facilitando o reconhecimento de lesões e inflamações, particularmente em áreas com alta densidade de vasos sanguíneos, como o trato gastrointestinal, (Fujifilm; Osawa, et al., 2018).

A tecnologia *Multi Light* está integrada no processador da Fujifilm ELUXEO 7000, Figura 12, que utiliza uma fonte de luz composta por quatro LEDs independentes (azul-violeta, azul, verde e vermelho) com comprimentos de onda específicos e de alta intensidade. A luz de comprimento de onda curto, especialmente no espectro azul-violeta e azul, penetra facilmente na mucosa e dispersa-se pouco e destaca a hemoglobina nos vasos sanguíneos, melhorando a visualização vascular e contribuindo para um diagnóstico mais preciso, (Fujifilm) (Osawa, et al., 2018).



Figura 12 Processador da Fujifilm ELUXEO 7000

O processamento LCI utiliza uma combinação avançada de espectros de luz que diferencia os tons de vermelho, oferecendo uma diferenciação mais precisa em comparação com a tecnologia de imagem WLI. Esse processo envolve o pré-processamento de radiação de banda estreita, que otimiza a captura das imagens, e a aplicação de tecnologia de cores específicas no pós-processamento, separando as cores nas componentes de azul, verde e vermelho. Essa separação aprimora o contraste das cores visíveis, permitindo uma distinção mais clara entre as lesões colorretais e a mucosa normal, (Neumann, Kreft, Sivanathan, Rahman, & Galle, 2021).

As lesões malignas gástricas, sob o modo WLI, apresentam uma coloração similar à da mucosa circundante, mas a refletância em comprimentos de onda curtos é significativamente distinta, (Fujifilm) (Osawa, et al., 2018).

O modo LCI revela alterações na cor da mucosa antes do aparecimento de alterações morfológicas em várias lesões gástricas. Faz uso de uma luz otimizada em comprimentos de onda específicos, geralmente na faixa de 410 a 450 nm (Figura 13), que é eficaz na absorção pela hemoglobina presente nos vasos sanguíneos, (Fujifilm; Osawa, et al., 2018).

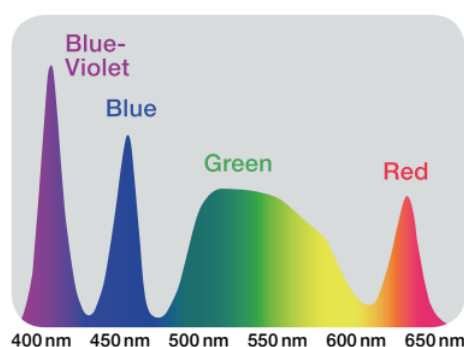


Figura 13 Gama de comprimentos de onda do modo LCI. Retirado de (Fujifilm)

O processamento digital do sinal LCI aumenta o contraste de cor, expandindo a cor próxima à vermelhidão da mucosa, Figura 14, (Fujifilm; Osawa, et al., 2018). O contraste de cor é

aprimorado na gama de cores vermelhas, isto é, o vermelho parece mais vermelho e o branco parece mais branco.

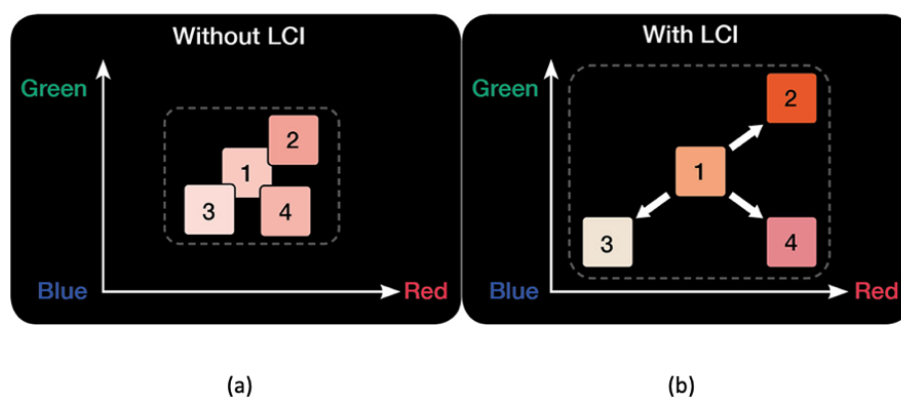


Figura 14 (a) Sem a tecnologia LCI; (b) com a tecnologia LCI; Adaptado de (Fujifilm)

A imagiologia com BLI permite uma visualização superior dos padrões vasculares e mucosos superficiais. Este modo é utilizado para identificar lesões superficiais, como pólipos, e alterações nas camadas mucosas, possibilitando a detecção precoce de lesões. As imagens de BLI são geradas a partir de uma combinação de imagens espectrais e de luz branca, (Osawa, et al., 2018).

Esta tecnologia baseia-se na iluminação num campo estreito do espectro luminoso, com recurso a dois lasers monocromáticos de comprimento de onda de 410 e 450 nm, Figura 15. Resultando no realce de alterações muito pequenas na superfície da mucosa, Figura 16, (Osawa, et al., 2018).

O modo BLI, com dois tipos de lasers a 410 nm e 450 nm, é útil para o diagnóstico de pólipos colorrectais, mas apresenta uma desvantagem na detecção de pólipos devido à sua escuridão. O modo BLI-bright, que utiliza um comprimento de onda de 450 nm de maior potência em comparação com o modo BLI, foi desenvolvido para superar esta limitação.

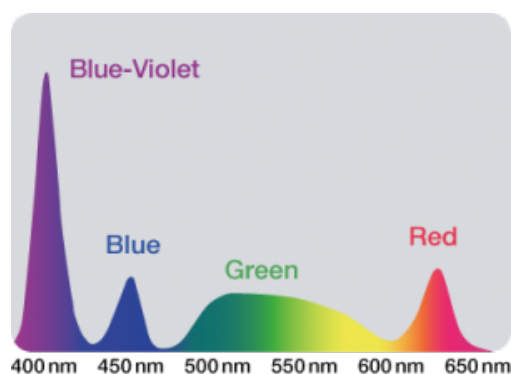


Figura 15 Gama de comprimentos de onda do modo BLI. Retirado de (Fujifilm)

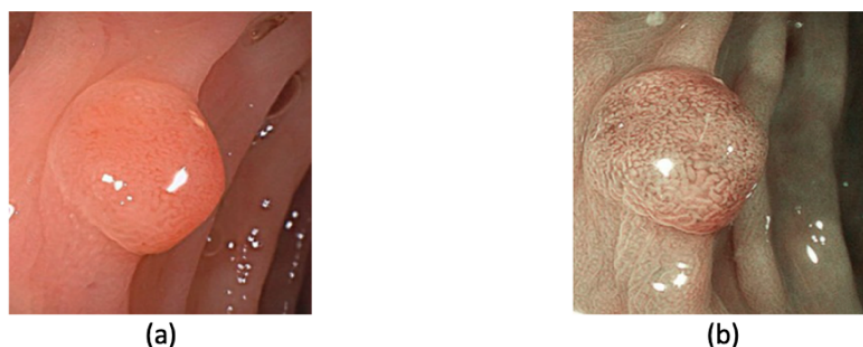


Figura 16 (a) Modo White Light; (b) Modo BLI. Adaptado de (Fujifilm, s.d.)

O modo *White Light* apresenta uma distribuição espectral ampla, com picos nas regiões do azul-violeta, verde e vermelho, que resulta numa imagem com aparência natural, semelhante à iluminação ambiente. Este modo é eficaz para a observação geral da mucosa, permitindo a identificação de alterações morfológicas. No entanto, apresenta limitações na diferenciação de estruturas vasculares superficiais, devido ao baixo contraste gerado pela absorção inespecífica da luz pela hemoglobina.

A hemoglobina (Hb) possui máximos de absorção em comprimentos de onda específicos, o que influencia diretamente a visibilidade das estruturas vasculares. A oxi-hemoglobina (HbO_2) apresenta um pico de absorção primário a 415 nm (banda de Soret) e picos secundários em 540 nm e 576 nm. A desoxi-hemoglobina (Hb) apresenta os seus picos de absorção a 430 nm e 555 nm. Quando o tecido é iluminado com luz contendo apenas os comprimentos de onda correspondentes aos picos de absorção da hemoglobina, os vasos absorvem essa luz de forma intensa, aparecendo mais escuros em relação à mucosa circundante. Uma região muito escura indica alta absorção de luz pela hemoglobina, o que corresponde geralmente a um vaso sanguíneo ou uma área com elevada concentração de sangue.

Uma vez que o modo *White Light* utiliza uma combinação ampla de comprimentos de onda, esse efeito de contraste é atenuado, pois a luz refletida por outras estruturas interfere na distinção clara dos vasos sanguíneos. Por outro lado, o modo LCI intensifica a emissão de luz nas faixas azul e verde, o que permite uma melhor diferenciação das tonalidades da mucosa.

O modo BLI, Figura 17 (c), utiliza predominantemente luz azul-violeta, com menor intensidade nas faixas verde e vermelha. Essa característica é particularmente relevante do ponto de vista fisiológico, uma vez que a hemoglobina — tanto na sua forma oxigenada (HbO_2) como desoxigenada (Hb) — apresenta picos de absorção em comprimentos de onda próximos de 415 nm, 540 nm e 576 nm. Assim, o BLI permite um realce eficaz das estruturas vasculares, melhorando o contraste entre os vasos sanguíneos e a mucosa adjacente.

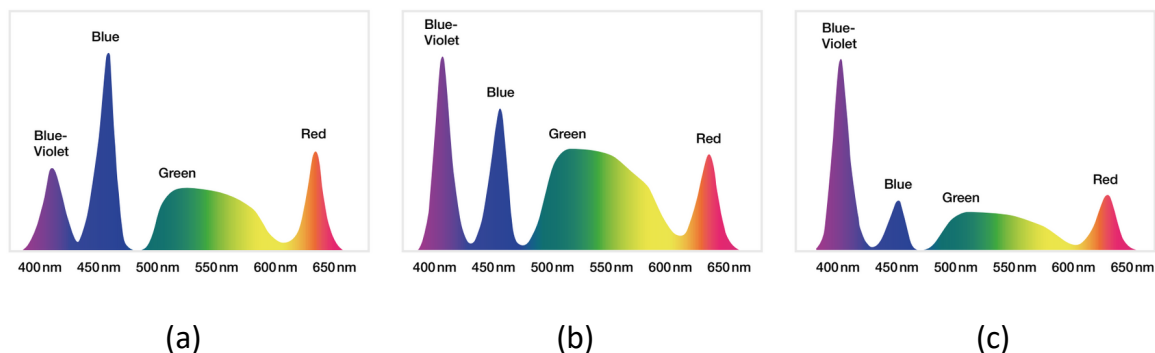


Figura 17 Distribuição espectral dos modos de imagem endoscópica: (a) White Light; (b) LCI; (c) BLI

Estudos como o de (Ikematsu, et al., 2017) e outros ensaios controlados aleatórios avaliaram a capacidade do BLI na detecção de pólipos colorretais. Ikematsu et al. utilizaram uma fonte de luz de xénon e obtiveram um número significativamente superior de adenomas por paciente no modo BLI em comparação com o modo WLI.

O estudo clínico realizado por (Ang, et al., 2019), em Singapura, comparou a taxa de detecção de pólipos entre os modos BLI-bright e WLI com uma fonte de luz a laser. Os pacientes foram aleatoriamente distribuídos para os grupos BLI-bright ou WLI. As taxas de detecção de pólipos para o BLI e o WLI foram, respetivamente, 61,1% e 48,3% ($p=0,002$), enquanto as taxas de detecção de adenomas foram de 48,3% e 34,4% ($p=0,010$). Assim, os autores concluíram que o modo BLI tem uma maior precisão na detecção de pólipos e de adenomas.

No estudo realizado em 2018 por (Santos, et al., 2018) foi utilizada a técnica BLI para a diferenciação entre tecidos benignos e malignos, o conjunto de dados incluiu 920 lesões superficiais diagnosticadas em 457 pacientes. Os autores obtiveram uma precisão de 95,5 % na diferenciação de lesões neoplásicas de não neoplásicas.

4.3 Composição do Endoscópio

O endoscópio pode ser dividido em três grandes partes: *Connector*, OSA (*Operation Section Assembly*) e ISA (*Insertion Section Assembly*).

O *connector* desempenha a função de fornecer energia elétrica aos componentes internos do endoscópio, como a câmara de vídeo, as placas de circuito impresso (PCB, *Printed Circuit Board*) e iluminação. As PCBs são concebidas para encaminhar sinais elétricos através de componentes eletrónicos. A câmara de vídeo requer energia elétrica constante para funcionar de forma eficaz e fornecer imagens claras e nítidas, necessárias para o diagnóstico médico.

O *connector* é a parte responsável por permitir a entrada de luz de uma fonte externa através da sua conexão com o sistema de vídeo, ou processador. O *connector* liga-se ao processador de fonte de luz ELUXEO 7000 (BL-7000), como ilustra a Figura 18.



Figura 18 Ligação do endoscópio ao processador

A luz gerada pela fonte externa percorre os cabos de fibra ótica até à extremidade distal do endoscópio, proporcionando a iluminação necessária para examinar adequadamente a área de interesse, Figura 19.

A fibra ótica é uma estrutura fina, com formato cilíndrico, produzida com materiais como vidro ou plástico. Esta fibra possibilita a transmissão de luz através de reflexão interna total, permitindo que a luz percorra a distância entre dois pontos.

A reflexão interna total é um fenómeno ótico que ocorre quando um raio de luz, ao tentar passar de um meio mais denso (com maior índice de refração) para um meio menos denso, atinge a interface entre os dois meios com um ângulo de incidência superior ao ângulo crítico. Em vez de se refratar, o raio é completamente refletido de volta para o meio mais denso.

A trajetória da luz ao atravessar a fronteira entre dois meios com diferentes densidades óticas é determinada, essencialmente, pelos índices de refração desses meios e pela relação entre o ângulo de incidência e o ângulo crítico. O ângulo crítico é definido como o maior ângulo de incidência para o qual a luz ainda se refrata na fronteira entre os meios. Quando o ângulo de incidência é igual ao ângulo crítico, a luz refrata-se ao longo da interface, formando um ângulo de refração de 90° em relação à normal.

O cálculo do ângulo crítico é realizado com base na Lei de Snell, expressa pela Equação 1:

$$n_1 \sin \theta_c = n_2 \sin 90^\circ \quad (1)$$

Onde n_1 é o índice de refração do meio mais denso (meio de incidência); θ_c é o ângulo crítico e n_2 o índice de refração do meio menos denso.

Dado que $\sin 90^\circ = 1$, a equação pode ser simplificada pela Equação 2:

$$\sin \theta_c = \frac{n_2}{n_1} \quad (2)$$

O procedimento padrão de endoscopia consiste na orientação da luz através de uma sonda composta por fibras óticas. Para garantir a propagação eficiente da luz ao longo da sonda, os

raios luminosos são introduzidos com um ângulo de incidência superior ao ângulo crítico das fibras, assegurando assim a ocorrência de reflexão interna total, (Abd-Almonem).

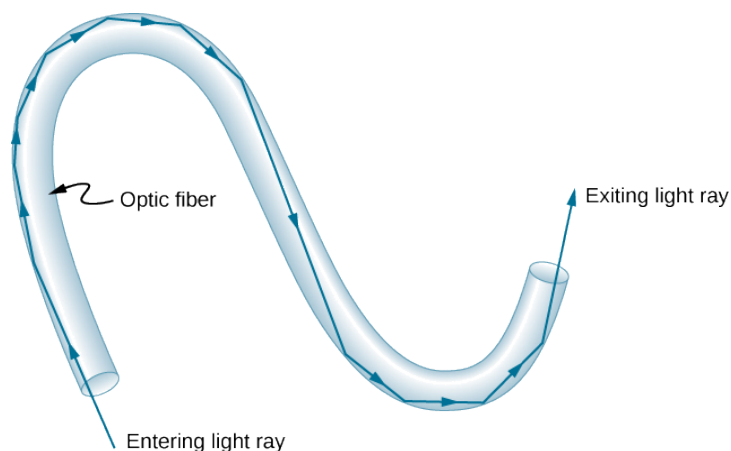


Figura 19 Comportamento da Luz dentro do Endoscópio

O *connector* permite o fornecimento de ar e água até à extremidade distal do endoscópio, de modo a limpar a lente da câmara de vídeo quando esta se encontra suja, garantindo imagens nítidas durante a sua utilização. Além disso, o *connector* está equipado com uma porta de sucção, que permite a absorção de fluidos provenientes da extremidade do endoscópio, Figura 20 e Figura 21.

Em todos os equipamentos endoscópicos da série 700, há uma placa de identificação, onde está presente informações relativamente ao modelo, número de série e ano de fabricação do equipamento.

Um dos aspetos críticos do *connector* é a sua capacidade de controlar a pressão interna do sistema endoscópico. Durante o procedimento, é frequentemente aplicada uma pressão de aproximadamente 30 kPa. Esta pressão é devidamente ajustada para criar o espaço necessário para a expansão das paredes dos órgãos, facilitando a progressão do endoscópio.

Na Figura 20, estão representadas os componentes que são ligadas ao processador. O conetor de abastecimento de água liga-se ao reservatório de água, o conetor de ventilação liga-se ao adaptador de ventilação, a guia de luz capta a luz irradiada pela fonte de luz, o conetor de sucção liga-se ao tubo do equipamento de sucção e o conetor de jato de água é sustentado por uma seringa ou à bomba de abastecimento de água.



Figura 20 Composição do Connetor

Na Figura 21 está representada a porção do *connector* que liga ao processador, nesta figura estão identificadas as componentes que são necessárias para estabelecer a comunicação entre o endoscópio e o processador, bem como para permitir a passagem de luz e de ar. A alimentação do equipamento é realizada por indução, a janela de comunicação utiliza radiação de infravermelhos para a passagem de sinal remoto.

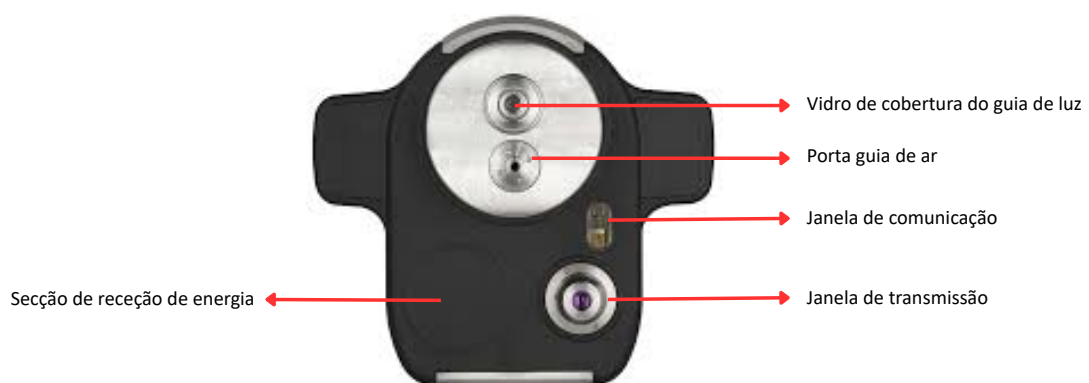


Figura 21 Connector Adaptado de (Fujifilm, s.d.)

A seção de operação do endoscópio, Figura 22, designada por OSA, constitui a unidade de controlo responsável pela manipulação da extremidade distal do instrumento. Esta seção integra um conjunto de manípulos (ou comandos), travões, botões e válvulas. A OSA é tipicamente equipada com dois manípulos independentes, os quais possibilitam o movimento direcional da extremidade distal nos planos vertical (cima/baixo) e horizontal (esquerda/direita).

Os travões permitem controlar a posição e estabilidade do tubo de inserção. Esses travões funcionam como mecanismos de bloqueio mecânico, que impedem o movimento involuntário do tubo quando se deseja manter uma posição específica.

A OSA incorpora um conjunto de válvulas de controlo, uma válvula de ar/água e uma válvula de sucção. Estas válvulas desempenham um papel fundamental nas funções de insuflação de ar, irrigação de água e aspiração. A válvula de ar/água permite alternar entre a insuflação, através da introdução controlada de ar ou dióxido de carbono (CO_2), e a lavagem da lente ótica com um jato de água, garantindo uma visibilidade adequada durante o exame. O CO_2 é mais comumente utilizado por ser rapidamente absorvido e excretado pelo organismo.

A válvula de sucção, por sua vez, possibilita a remoção eficiente de líquidos, secreções ou resíduos do campo de visão. Estas funções só podem ser possíveis quando são colocadas umas válvulas específicas, usam o princípio pressão-retorno, a ativação ocorre mediante a aplicação de força manual, e o retorno à posição de repouso assegurado por um sistema de mola incorporado.

A OSA integra um pórtico que constitui a interface física entre o meio externo e o canal de instrumentação do endoscópio. Este canal operativo permite a introdução de instrumentos auxiliares, incluindo fórceps de biópsia, sondas terapêuticas e pinças de coagulação, que são conduzidos até à extremidade distal do endoscópio com o objetivo de realizar intervenções diagnósticas ou terapêuticas, Figura 23.

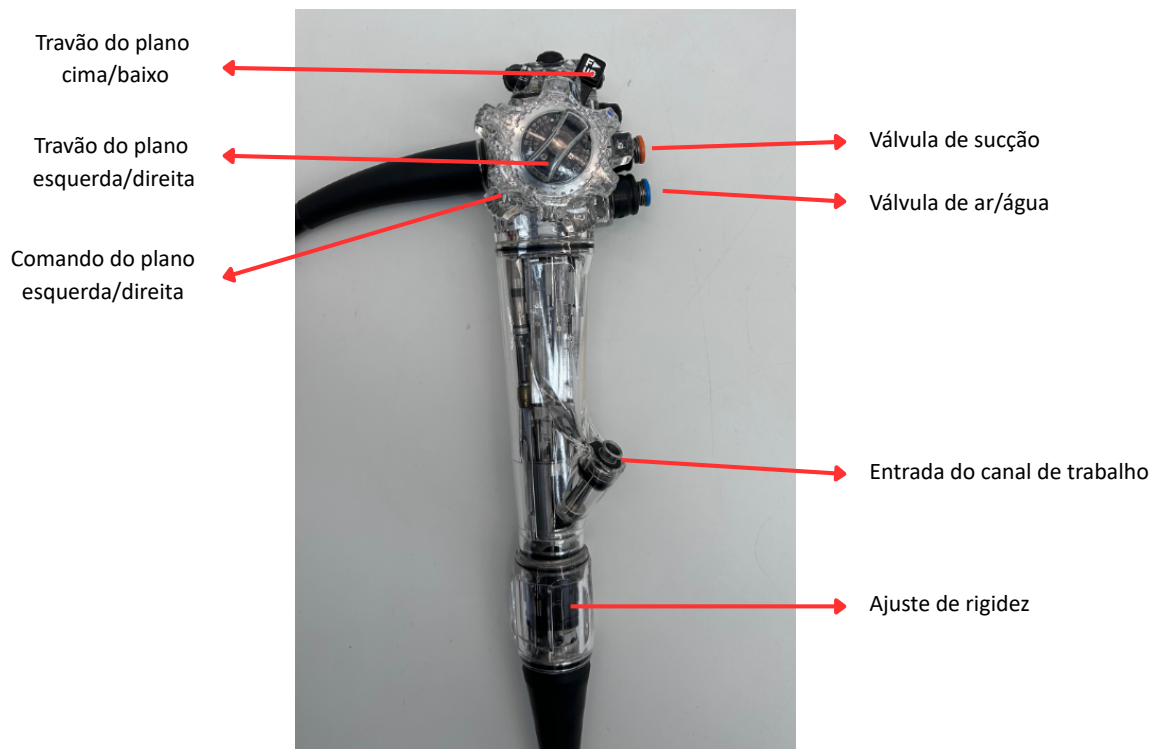


Figura 22 Componentes da OSA

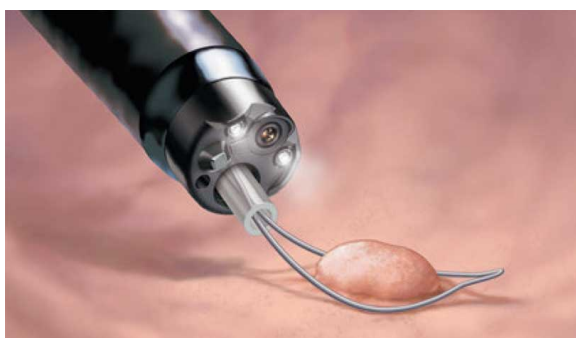


Figura 23 Intervenção terapêutica, retirado de (Fujifilm, s.d.)

Adicionalmente, a OSA incorpora quatro a cinco botões funcionais, cuja configuração depende do modelo específico do endoscópio. Os botões funcionais da OSA podem ser configurados diretamente no processador, permitindo a personalização de diversas funções, tais como a captura de imagem, o controlo da intensidade luminosa e a ativação de filtros digitais de realce de imagem, como o FICE e o BLI/LCI.

Ao acionar o botão configurado para ativação dos filtros BLI ou LCI, por norma o botão 2, a imagem exibida sofre uma alteração na sua aparência. Ao pressionar uma vez o botão, o filtro LCI é ativado, ao pressionar novamente o botão, o sistema alterna para o modo BLI. O modo FICE funciona de forma semelhante, pressionando o botão 4, Figura 24.

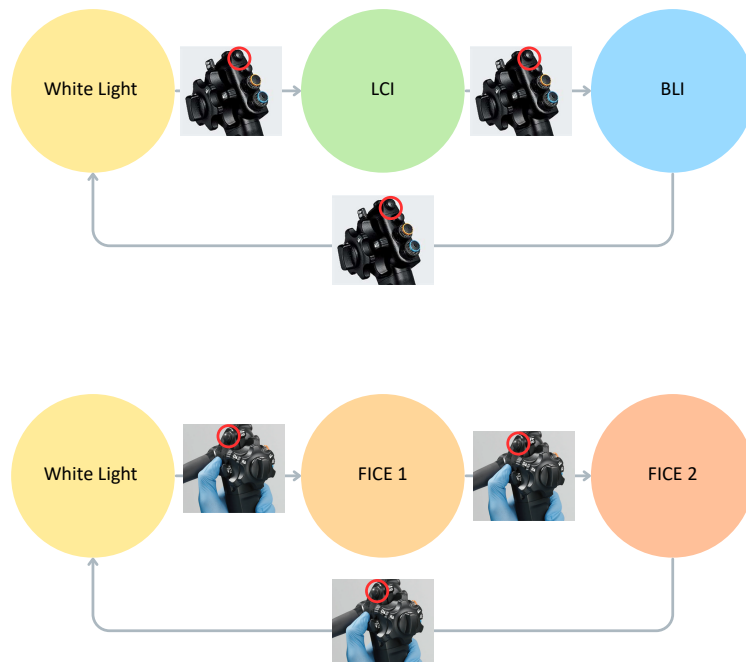


Figura 24 Modos de Observação

O canal de trabalho apresenta especificações que variam consoante o tipo e modelo do endoscópio, Figura 25. No caso dos gastroscópios, os diâmetros disponíveis incluem 2,4 mm, 2,8 mm e 3,8 mm. Já os colonoscópios apresentam apenas dois diâmetros: 3,2 mm e 3,8 mm.

Esta informação encontra-se na OSA, juntamente com a identificação do respetivo modelo, permitindo uma consulta rápida por parte do utilizador e facilitando a seleção de acessórios compatíveis com o diâmetro do canal.

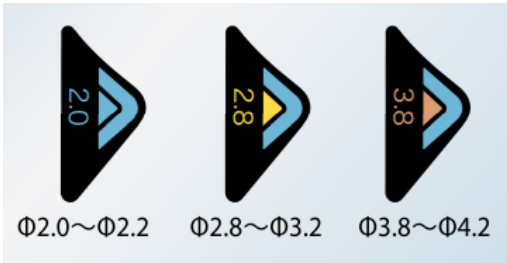


Figura 25 Diâmetros de canal de trabalho, retirado de (Fujifilm, s.d.)

Nos modelos que integram a funcionalidade de zoom ótico, um quinto botão é adicionado, permitindo ao utilizador realizar ajustes na ampliação da imagem. Estes modelos dispõem de dois botões dedicados ao controlo da ampliação: um para aumentar o nível de zoom e outro para o reduzir. A ampliação ótica máxima é de 135 x, Figura 26, que permite uma visualização detalhada da superfície da mucosa e dos padrões vasculares, (Fujifilm).

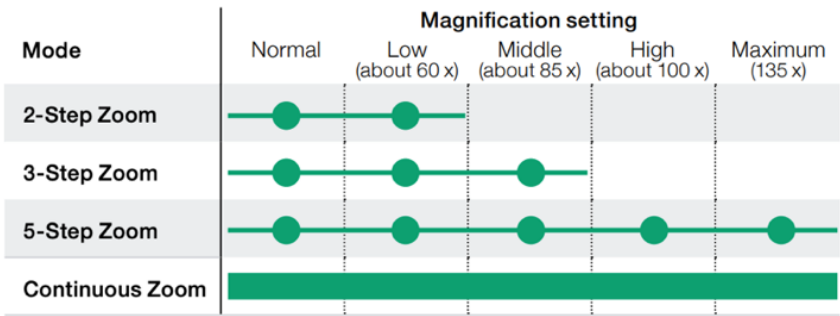


Figura 26 Configurações da funcionalidade de zoom, retirado de (Fujifilm, s.d.)

Alguns modelos de colonoscópios estão equipados com um mecanismo de ajuste variável de rigidez, Figura 27, que permite a modificação da rigidez do tubo de inserção. Essa funcionalidade facilita a movimentação do endoscópio ao longo do trato intestinal, proporcionando uma inserção mais fluida e uma maior facilidade no avanço do equipamento.

O sistema de ajuste de rigidez apresenta quatro posições distintas, cada uma associada a um nível específico de tensão aplicado ao tubo de inserção. Na posição neutra, o tubo apresenta uma flexibilidade máxima, uma vez que não é exercida qualquer força de tensão. À medida que o manípulo é ajustado para as posições 1 a 3, o grau de rigidez aumenta progressivamente. Quando o manípulo está na posição 3, o tubo atinge o estado de máxima tensão, resultando numa estrutura mais rígida.

No estudo realizado em 2021 por (Gerges, et al., 2021), foi avaliado a eficácia e a usabilidade do sistema de rigidez dos equipamentos do modelo EC-760R-V/I durante procedimentos médicos. A amostra do estudo foi composta por 180 participantes, dos quais 81 eram do sexo

masculino e 99 do sexo feminino. Segundo os resultados obtidos no estudo, o sistema de ajuste de rigidez foi utilizado em 130 dos 180 casos observados. Adicionalmente, 97,7% dos utilizadores avaliou o sistema como útil na prevenção da formação de *loops*.

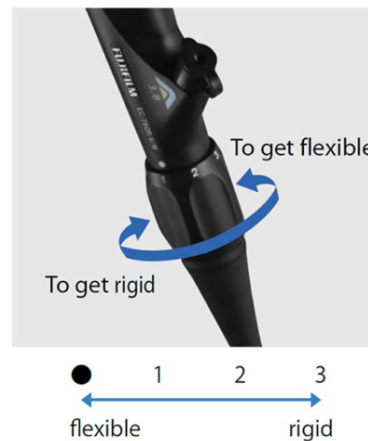


Figura 27 Mecanismo de ajuste de rigidez, retirado de (Fujifilm, s.d.)

O conjunto da secção de flexão (BSA, *Bending Section Assembly*) pertence à secção ISA, Figura 28, do endoscópio médico é caracterizada por apresentar uma flexibilidade que permite ao médico orientar a extremidade distal. Esta secção é composta por um conjunto de anéis de aço inoxidável ou nitinol interligados. Uma manga de malha fina de aço inoxidável ou polímero envolve o conjunto do anel, proporcionando resistência à abrasão e protegendo os componentes internos. A BSA é envolta por um material de borracha flexível e resistente (RBS, *Rubber Bending Section*) cuja função é vedar o conjunto do anel contra a entrada de fluidos, sendo uma superfície capaz de estabelecer contacto com o interior do corpo humano.

O FSA (*Flexible Section Assembly*) é o tubo de inserção, isto é, a parte que é inserida no corpo. No seu interior encontram-se o canal de trabalho, o tubo que fornece ar e água para limpar a extremidade distal, o tubo jet, a câmara de vídeo e as guias de luz.

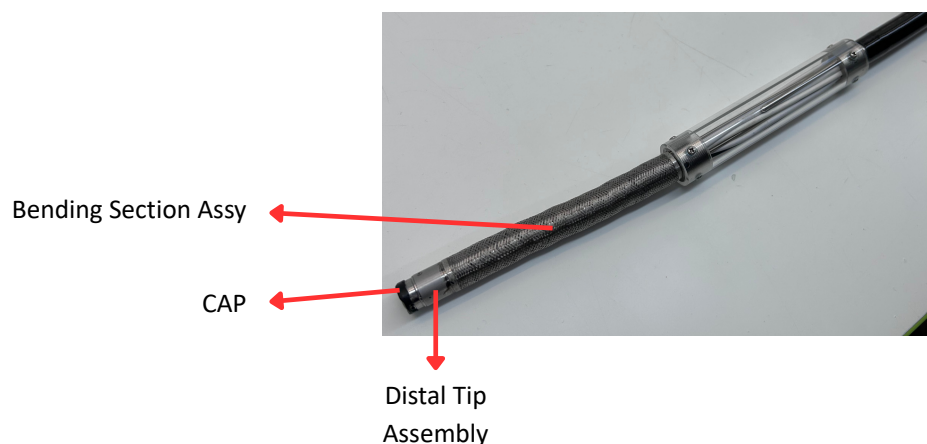


Figura 28 Parte distal do endoscópio

A extremidade distal de um endoscópio integra diversos componentes fundamentais que asseguram a eficácia do exame endoscópico. A Figura 29 ilustra a extremidade distal (CAP) de um modelo de endoscópio, cuja disposição dos componentes tende a variar de modelo para modelo. A extremidade distal integra duas guias de luz, o canal de trabalho, a câmara de vídeo, o nozzle e o canal *jet*. O propósito da CAP é proteger os componentes delicados, como a câmara, de impactos. Para além disso, por ser fabricada com um material biocompatível, a CAP impede o contacto directo das partes metálicas do endoscópio com o interior do corpo humano.

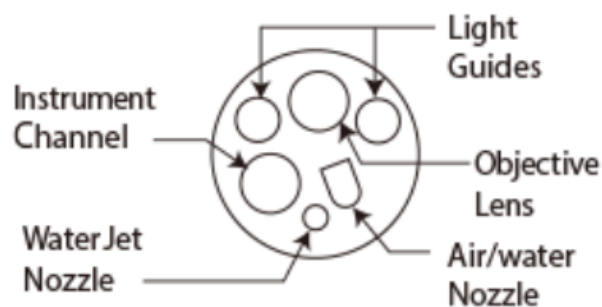


Figura 29 Componentes presentes na CAP, retirado de (Fujifilm, s.d.)

A câmara de vídeo capta as imagens da cavidade examinada e transmite ao processador, que por sua vez processa os sinais recebidos do endoscópio e transmite como sinal de vídeo. O semicondutor de alta precisão, *CMOS Image Sensor*, Figura 30, integra amplificadores, circuitos de processamento de sinal e conversores analógico-digitais. Este semicondutor encontra-se integrado diretamente na extremidade do endoscópio, convertendo de forma rápida o sinal analógico em sinal digital no local da observação. A tecnologia CMOS suporta a tecnologia de varredura progressiva de 60 *frames*, na qual imagens completas são processadas, em vez de meios *frames* processados quando se utiliza o método de varredura entrelaçada. Resulta numa imagem de alta resolução e imagens em movimento suaves com redução do efeito de desfocagem. Esta funcionalidade garante uma transmissão de imagem com elevado brilho e ruído reduzido, capturada por uma lente e convertendo-a em sinais elétricos para cada pixel.

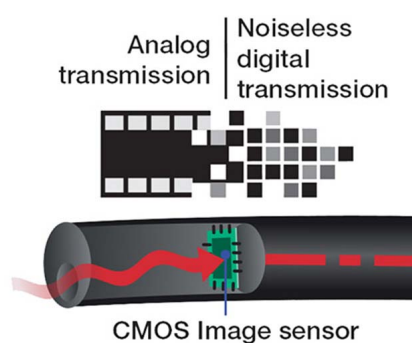


Figura 30 Tecnologia CMOS

Os jatos de água dianteiros, *nozzle* e *jet*, encontram-se em todos os equipamento endoscópios, em exceção dos broncoscópios. O modelo EG-720R também não possui *jet*. O *nozzle* é utilizado para limpar a lente da câmara de vídeo de detritos ou outros componentes. O *jet* é utilizado para manter a área interna do organismo limpa.

As guias de luz consistem em feixes de fibras óticas responsáveis pela condução da luz proveniente do processador até a região de interesse, permitindo a adequada iluminação. Estas guias percorrem o equipamento desde o *conector* até à extremidade distal, posicionando-se nos *barrel's* integrados na CAP.

As guias de luz e a câmara de vídeo percorrem todo o equipamento, estes elementos iniciam-se no *connector*, passam pelo tubo FSB, ligam-se à OSA e estendem-se até à parte distal. Por sua vez, o tubo de ar/água, o tubo de aspiração e o *jet* são interrompidos na OSA por válvulas. O trajeto entre a OSA e a parte distal é feito por tubos distintos, uma vez que a quantidade de ar, água e sucção libertada na parte distal é regulada pelas válvulas de ar/água e de aspiração, respetivamente. Já o canal de trabalho inicia-se na OSA, percorre o tubo FSA e termina na parte distal, Figura 31.

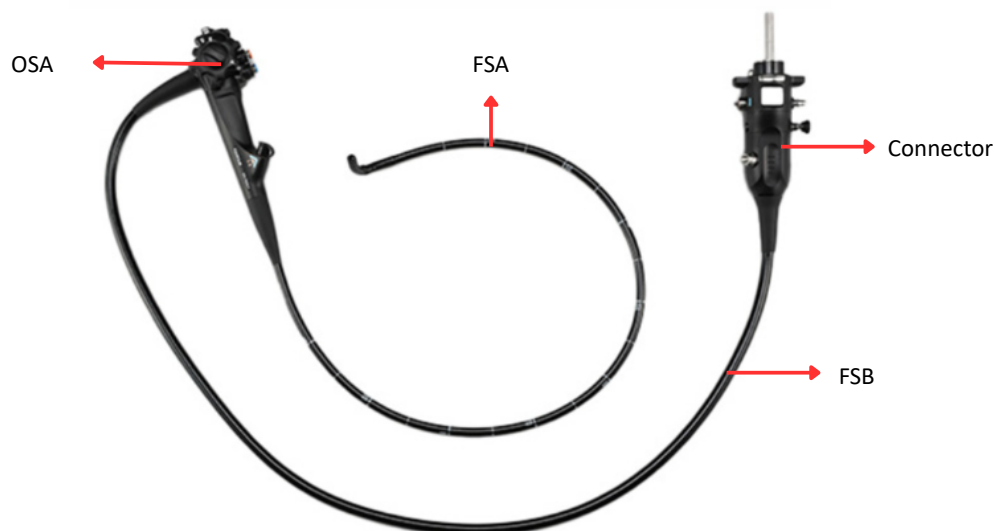


Figura 31 Endoscópio

A Figura 32 apresenta os elementos necessários para a realização de um exame endoscópico: um monitor que recebe o sinal do processador e exibe a informação e um *trolley* para transportar e mover o equipamento de endoscopia.

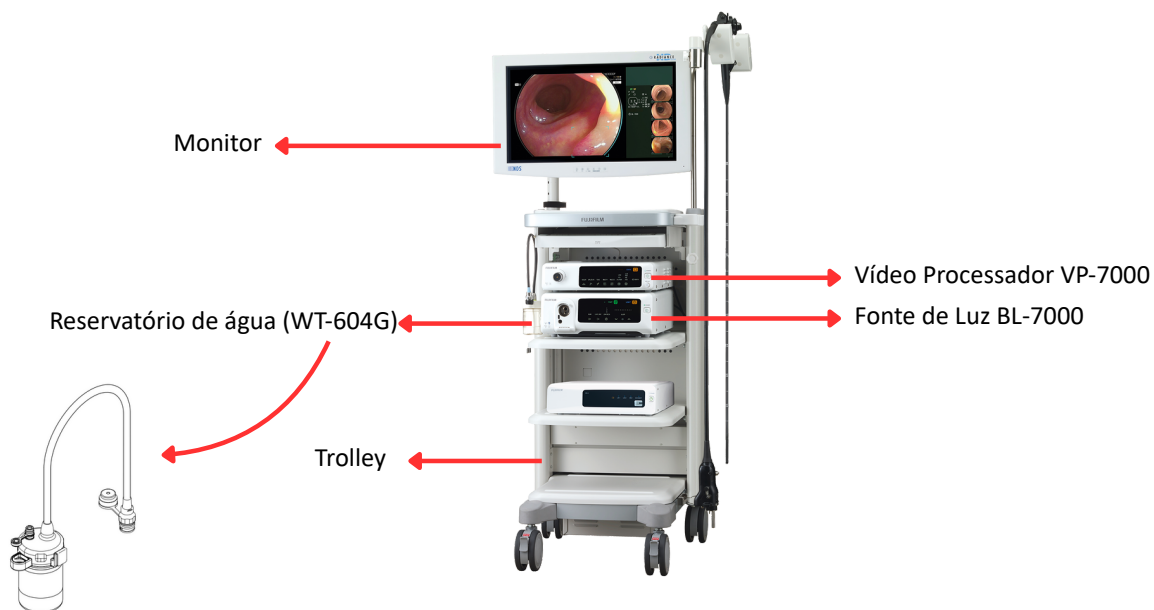


Figura 32 Trolley com os processadores 7000 e um monitor

4.4 Tipos de Reparação

Os endoscópios são instrumentos delicados que, devido à sua complexidade e ao ambiente em que operam, estão sujeitos a diversos tipos de deterioração e avarias, que podem comprometer tanto a funcionalidade do equipamento quanto a segurança e o bem-estar dos pacientes.

A utilização regular dos endoscópios resulta no desgaste gradual dos seus componentes, como os canais de instrumentação, os componentes da extremidade distal, tubo de inserção e os comandos de operação são particularmente suscetíveis ao uso contínuo. A exposição frequente a fluidos biológicos, bem como aos agentes químicos utilizados nos processos de desinfecção e esterilização, antecipa a degradação dos equipamentos. Além disso, o incorreto manuseamento durante e após a utilização causa uma proporção significativa de danos acidentais. Entre os incidentes mais comuns destacam-se os causados pelo contacto com instrumentos cortantes, como pinças e fórceps de biópsia, ou pela aplicação excessiva de força durante a introdução do endoscópio. A lente da objetiva rachada ou riscada, a deformação da BSA ou os canais de ar/água obstruídos são exemplos de danos que podem ser causados pelo uso incorreto dos equipamentos.

Embora o desgaste seja um processo inevitável no ciclo de vida de qualquer equipamento, é possível mitigar os seus efeitos através da implementação de manutenções preventivas e formação adequada. A realização periódica de inspeções, a limpeza adequada conforme os protocolos estabelecidos e a reparação atempada de danos são práticas fundamentais para a preservação das condições de funcionamento dos equipamentos. Tais ações não só garantem

a eficiência operacional, mas também prolongam a vida útil dos equipamentos e reduzem a probabilidade de falhas imprevistas.

As intervenções de reparação podem ser classificadas, de maneira geral, em duas grandes categorias: reparações de pequena dimensão (*minor repair*) e reparações de grande dimensão (*major repair*).

As reparações de pequena dimensão referem-se, sobretudo, a danos superficiais ou a ajustes funcionais menores, que não exigem a desmontagem completa do equipamento nem a substituição de componentes estruturais significativos, como o canal de trabalho. Essas reparações têm como objetivo restaurar ou otimizar a funcionalidade dos equipamentos sem comprometer a integridade estrutural.

Dentro da categoria de reparações de pequena dimensão podem ser distinguidos dois tipos: *small repairs* e reparações mais complexas.

As *small repairs* envolvem a substituição de peças pequenas, ajustes simples, como ajustes de angulações e de folgas dos comandos, ou limpeza de componentes que entram em contacto com meio biológico. As restantes reparações *minor*, podem envolver a substituição de peças e o uso de cola epóxi para selar ou fixar componentes. Nas reparações em que é utilizada cola, o tempo necessário para a sua cura é um fator crucial, uma vez que é necessário um período de espera para garantir que os materiais estejam devidamente fixados antes de o equipamento poder ser utilizado.

Para a fixação e vedação de componentes do endoscópio é utilizada a cola epóxi devido às propriedades específicas como a biocompatibilidade e resistência elevada ao impacto, aos fluidos corporais e aos químicos de desinfecção. Nas reparações em que não é necessário a utilização desta cola, o tempo de reparação pode ser relativamente mais rápido, por não ser necessário o tempo de cura da cola.

Por outro lado, as reparações de grande dimensão envolvem um maior grau de complexidade técnica. Ao contrário das *minor repairs*, as *major repairs* passam pela linha de reparação, uma vez que exigem a desmontagem completa do equipamento para a substituição de componentes internos. Qualquer reparação que envolva a substituição da peça metálica DTA (*Distal Tip Assembly*) que reúne os tubos de ar, água, câmara de vídeo e guias de luz, de forma a facilitar o seu posicionamento adequado na CAP; câmara de vídeo (CHA); o tubo flexível A (FSA); o conjunto da secção de flexão (BSA) ou as guias de luz (LGB) é classificada como uma *major repair*.

4.5 Inspeção do Endoscópio

A inspeção dos endoscópios deve ser minuciosa e conduzida com rigor técnico, a fim de garantir a conformidade do equipamento. A inspeção pretende identificar as condições de funcionamento do equipamento e, caso seja identificada alguma anomalia, deve ser realizado

um relatório técnico e um orçamento com o intuito de documentar os problemas detectados e estimar os custos da reparação.

A inspeção dos endoscópios deve iniciar-se, preferencialmente, com uma avaliação visual minuciosa de toda a extensão do equipamento. Nessa fase, é fundamental verificar se há sinais de desgaste físico, como marcas de impacto, deformações, vincos, descoloração ou qualquer outra alteração que possa comprometer o desempenho do equipamento.

O *connector* deve estabelecer uma ligação estável ao processador de imagem. Uma conexão inadequada pode comprometer a iluminação do campo visual, prejudicando a qualidade do exame. A imagem endoscópica não deve apresentar ruído, desfocagem, embaciamento ou outras irregularidades.

A ponta de prova de luz, localizada no *connector*, não deve estar solta ou apresentar folgas. Caso isso se verifique, deve ser devidamente apertada ou ajustada. Além disso, as guias de luz devem estar corretamente posicionadas junto à ponta de prova de luz, assegurando a condução eficiente da iluminação. A janela da lente de guia de luz também deve ser verificada, não deve apresentar fissuras ou sinais de condensação (humidade interna). É no *conector*, nas PCBs, que se encontram as informações específicas do equipamento, tais como o número de série, o modelo e o número de exames do equipamento. Por conseguinte, as placas de identificação devem estar legíveis, de forma a assegurar a rastreabilidade e a correta identificação do endoscópio.

O *connector* está integrado ao tubo FSB (*Flexible Section B*), o qual deve ser cuidadosamente inspecionado quanto à presença de manchas significativas, deformações, vincos e amolecimento ou endurecimento. Estas alterações podem ser identificadas visualmente ou ao aplicar tensão ao tubo, sendo indicativas de possíveis danos estruturais internos.

É no interior dos tubos FSA e FSB que os componentes essenciais como a câmara de vídeo, guias de luz, canal de trabalho, canal de ar/água e canal de jet se encontram. Desta forma, o FSA e o FSB devem ser submetidos a uma inspeção visual minuciosa. A presença de porções amassadas pode ser indicador de danos internos que comprometam o funcionamento do endoscópio, como obstrução dos canais, comprometimento da imagem ou dificuldade na inserção de acessórios pelo canal de trabalho.

Na OSA, os comandos de angulação devem ser verificados quanto à presença de fricção excessiva. Adicionalmente, deve-se garantir que o comportamento dos comandos de angulação seja independente entre si, ou seja, a manipulação de um comando não deve interferir no funcionamento do outro. Outro aspeto crucial da inspeção é a verificação das folgas nos comandos de angulação. A presença de folgas pode resultar em movimentos imprecisos ou instáveis. Além disso, deve ser realizada a verificação das angulações em todas as direções operacionais, nomeadamente para cima, baixo, direita e para a esquerda, assegurando que a ponta do endoscópio se mova com precisão em todas essas direções.

A funcionalidade do sistema de insuflação de ar deve ser igualmente testado, com o equipamento conectado ao processador. Para este teste, deve-se ligar a aspiração no

processador e colocar a válvula descartável de ar/água no conector de abastecimento de água, localizado no *connector*. De seguida, o orifício da válvula de ar/água, Figura 33, deve ser tapado com o dedo. Deve-se observar a quantidade de ar e/ou água que é expelida pela ponta distal do endoscópio. A emissão de ar pode ser verificada submergindo a ponta distal num recipiente com água, observando-se a formação de bolhas. Após a remoção do dedo do orifício, não deve ocorrer qualquer saída de ar ou água, o que indicará o funcionamento adequado do sistema.

A insuflação de água deve ser avaliada de forma semelhante. Neste caso, deve-se pressionar a válvula de ar/água com o dedo e verificar se há saída de água através do NOZ. A água deve ser direcionada à objetiva da câmara, com o objetivo de remover possíveis detritos acumulados na superfície. Após a descompressão da válvula, o endoscópio não deve continuar a irrigar. Caso a pulverização de água seja insuficiente ou persista após a liberação da válvula, o equipamento deve ser submetido a reparação.



Figura 33 Acessórios descartáveis

Cada modelo de endoscópio apresenta características específicas em relação às suas capacidades de angulação. No entanto, é importante destacar que, para a avaliação precisa das angulações de cada modelo, existem fichas técnicas específicas que contêm informações detalhadas sobre os valores exatos de angulação e outras especificidades técnicas de cada equipamento.

Nos modelos de gastroscópios, as angulações são, geralmente, de 210° na direção superior, 90° na direção inferior, e 110° nas direções esquerda e direita. A angulação superior de 210° foi projetada com o objetivo de facilitar o acesso e o tratamento de lesões localizadas em áreas de difícil alcance.

Nos colonoscópios, as angulações são de 180° nas direções superior e inferior e 160° nas direções esquerda e direita. Além da verificação das angulações, é fundamental realizar também a inspeção dos travões do endoscópio, que têm como função bloquear a angulação do tubo de inserção, conforme aplicada pelos comandos. Além da sua função de impedir o movimento da parte distal do equipamento, os travões não devem apresentar um comportamento duro ou rígido.

A verificação do funcionamento do zoom ótico realiza-se em duas fases. Na primeira fase, avalia-se o funcionamento progressivo do sistema: o botão de ampliação é pressionado sucessivamente quatro vezes, observando-se o aumento gradual da imagem, seguido da utilização do botão de retorno para regressar à posição inicial. Na segunda fase, testa-se o funcionamento contínuo do zoom, mantendo o botão de ampliação pressionado até atingir o nível máximo de zoom. De seguida, o botão de retorno é mantido pressionado até o sistema regressar novamente à posição inicial. Este procedimento permite validar a resposta progressiva e contínua do mecanismo de zoom, assegurando a sua estabilidade e precisão ao longo de todo o intervalo de ampliação. Se, durante o processo de ampliação, a imagem se apresentar desfocada, apesar da correta execução do procedimento de verificação, o equipamento deve ser submetido a intervenção técnica especializada para assegurar o seu correto funcionamento.

Na extremidade distal do equipamento localiza-se a BSA, estrutura composta por uma malha que permite a mobilidade da ponta distal do endoscópio. Esta deve ser submetida a uma avaliação quanto à presença de falhas na sua estrutura ou de deformações. A BSA está envolta por um material de borracha flexível e resistente (RBS), fixa por um fio (fio de costura) coberto por cola epóxi. Tanto a cola epóxi como a RBS que envolvem a BSA devem ser cuidadosamente avaliados, de forma a garantir que não apresentem cortes, desgaste ou deformações. Estas falhas podem comprometer a integridade do equipamento e, consequentemente, a segurança do paciente, uma vez que ambas as substâncias estão em contacto direto com o interior do corpo humano.

A CAP, sendo a primeira parte do equipamento a estabelecer contacto com o interior do corpo humano e a que irá manter o maior contacto durante o procedimento, não deve apresentar cortes, fraturas ou sinais de queimaduras. A inspeção deve ser realizada de forma metódica, utilizando um microscópio ótico, que permite uma avaliação detalhada e precisa, garantindo que eventuais danos ou imperfeições sejam devidamente identificados.

A lente da câmara de vídeo deve ser minuciosamente inspecionada no microscópio ótico para a possível presença de riscos, arranhões, impactos, invasão de fluídos ou selos de cola de epóxi rachados.

As *barrel's* que envolvem as guias de luz não devem apresentar fissuras, apresentar sujidade interior ou ter os selos de cola epóxi rachados. A sujidade no interior do barril pode obstruir a passagem da luz ou criar reflexos na imagem. Assim, o estado físico dos barris impacta, de forma direta, a integridade e imagem durante o procedimento médico.

Caso sejam detetados detritos nas *barrel's* ou na lente da câmara, deve ser utilizada uma gaze embebida numa solução detergente apropriada para limpar de forma suave. Se os detritos não forem completamente removidos ou se a imagem endoscópica não retornar uma imagem clara, o equipamento deverá ser encaminhado para reparação.

O *nozzle*, presente na CAP, deve ser submetido a uma inspeção minuciosa utilizando um microscópio ótico, com o objetivo de identificar a presença de qualquer tipo de sujidade ou

anomalia no formato. A presença de contaminantes ou a deformação do nozzle pode obstruir ou restringir o fluxo adequado de ar ou água.

5 Estado da Arte

A gestão de manutenção de equipamentos médicos tem sido amplamente realizada de forma manual, o que, apesar de ser uma prática comum, apresenta várias limitações. As limitações de uma gestão manual engloba erros humanos, que podem comprometer a eficácia dos processos de manutenção e o funcionamento dos equipamentos médicos, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

A falta de um sistema centralizado e automático de gestão resulta em falhas na programação de tarefas, o que pode levar à omissão de manutenções e à realização de reparações tardias, aumentando o risco de falhas nos equipamentos médicos e influenciando a segurança dos pacientes. A gestão ineficaz das tarefas de manutenção também pode resultar no uso inadequado dos recursos, como a duração da reparação e utilização de peças que não são necessárias para a reparação do equipamento. Além disso, a gestão manual de peças de reposição pode levar a um excesso ou falta de *stock*, aumentando os custos e o tempo de inatividade. Por conseguinte, um sistema inteligente que possa automatizar a gestão da manutenção do equipamento médico e otimizar o processo pode ser necessário, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

O estudo de (Izadi, Bakhshali, & Sarrafzadeh, 2023) pretende otimizar a disponibilidade e a fiabilidade de equipamentos utilizados em meio hospitalar, bem como reduzir os custos de manutenção dos mesmos. Assim, é proposto um sistema inteligente que recorre a algoritmos de *machine learning*. A metodologia é composta por quatro fases principais: recolha de dados, pré-processamento, desenvolvimento de modelos e implementação do sistema. A informação recolhida abrange o desempenho do equipamento, as atividades de manutenção e o inventário de peças sobresselentes. A recolha de dados foi efetuada no sistema de gestão da manutenção do hospital e em outras fontes relevantes da Universidade de Ciências Médicas de Mashhad, Mashhad, Irão.

As etapas de pré-processamento deste estudo incluíram a limpeza de dados, a transformação de dados e a seleção de características. A limpeza de dados consistiu na remoção de valores ausentes e *outliers* do conjunto de dados. A transformação de dados envolveu a conversão de dados categóricos em dados numéricos, utilizando o processo de codificação one-hot. A seleção de características envolve a seleção das características mais relevantes para os modelos de *machine learning*.

Os modelos de *machine learning* utilizados neste estudo incluem regressão logística, árvores de decisão e redes neurais artificiais. Os modelos foram treinados com os dados pré-processados e avaliados recorrendo a técnicas de validação cruzada.

A previsão de falhas no equipamento é efetuada através da utilização de modelos de *machine learning*, que recorrem a características como a idade do equipamento, a frequência de utilização e o histórico de manutenção, com vista a estimar a probabilidade de ocorrência de falhas. O módulo emite um alerta aos utilizadores sempre que a probabilidade de falha do equipamento excede um valor limite, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

O módulo de programação de manutenção recorre a modelos de *machine learning* para planear tarefas de manutenção com base no desempenho do equipamento e na disponibilidade de peças sobresselentes, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

O módulo de gestão de peças sobresselentes recorre a modelos de *machine learning* para gerir o inventário com base na previsão de falhas do equipamento e no agendamento de manutenção. Os modelos utilizam características como a probabilidade de falha do equipamento e os cronogramas das tarefas de manutenção para prever as peças sobresselentes necessárias e as suas quantidades, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

O módulo de previsão de falhas de equipamentos foi avaliado com dados históricos, com cem amostras de equipamentos médicos. Os dados foram divididos em treino e teste, 70 % e 30 %, respetivamente. Foram usados três modelos de *machine learning*: regressão logística, árvores de decisão e redes neurais artificiais. O modelo de redes neurais artificiais obteve uma exatidão de 87 %, a regressão logística de 82 % e o modelo de árvore de decisão de 79 %, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

Para o módulo de programação de manutenção foram utilizados dados simulados para testar a eficácia dos modelos de *machine learning* na priorização de tarefas de manutenção. O conjunto de dados contém informações sobre as tarefas de manutenção programadas com base nas métricas de desempenho do equipamento e na disponibilidade de peças sobresselentes. Foram usados os mesmos modelos de *machine learning*. Os resultados obtidos demonstraram que o modelo de rede neuronal artificial superou os outros modelos na minimização do tempo de inatividade do equipamento, apresentando uma redução média de 20% no tempo de inatividade, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

O módulo de gestão de peças sobresselentes utilizou dados simulados. O conjunto de dados inclui informações das probabilidades previstas de avaria do equipamento e nos calendários de tarefas de manutenção. Foram usados os mesmos modelos de *machine learning*. O modelo de

rede neuronal artificial superou os outros modelos na minimização dos custos de inventário de peças sobresselentes, com uma redução média de custos de 15 %, (Izadi, Bakhshali, & Sarrafzadeh, 2023).

No âmbito da modernização dos processos de manutenção de equipamento médicos, (Ma, Cao, & Lu, 2024) apresenta uma proposta relevante para a modernização da manutenção de equipamentos médicos através da aplicação de tecnologias inteligentes. O estudo propõe o desenvolvimento e implementação de um sistema de manutenção inteligente que visa, simultaneamente, melhorar a eficiência operacional e reduzir os custos associados à gestão técnica dos equipamentos utilizados em ambientes clínicos.

O sistema proposto baseia-se na integração de tecnologias emergentes, nomeadamente a Internet das Coisas (*Internet of Things* – IoT), a análise de *big data* e algoritmos de inteligência artificial, com o objetivo de aperfeiçoar a precisão na previsão de falhas e garantir intervenções de manutenção mais assertivas e oportunas. A recolha permanente de dados operacionais permite detetar anomalias precoces e identificar padrões de comportamento que possam indicar falhas iminentes. Este processo é essencial para a implementação de estratégias de manutenção preditiva, nas quais as intervenções são realizadas com base na condição real do equipamento, antecipando falhas e minimizando interrupções não planeadas, (Ma, Cao, & Lu, 2024).

No desenvolvimento do sistema inteligente de manutenção de equipamentos médicos, foi adotado a linguagem Java. Foi utilizado o *framework* Spring Boot para acelerar o desenvolvimento, simplificando a configuração e a integração de bibliotecas, (Ma, Cao, & Lu, 2024).

A IoT possibilita a conexão dos equipamentos médicos, permitindo a recolha e transmissão de dados em tempo real. Por sua vez, a análise de grandes volumes de dados permite identificar padrões e potenciais riscos no funcionamento dos equipamentos, realizando uma extração detalhada de informações sobre o desempenho dos equipamentos. Os algoritmos de inteligência artificial fornecem uma base de decisão científica para a previsão de falhas e uma manutenção precisa.

Os dados recolhidos pelos sensores são armazenados e processados numa plataforma de *Big Data*, concebida para gerir grandes volumes de informação provenientes de múltiplos equipamentos em simultâneo. A análise dos dados obtidos permite extrair padrões de funcionamento dos equipamentos e identificar quaisquer anomalias que sejam indicadoras de falhas iminentes, (Ma, Cao, & Lu, 2024).

O sistema incorpora mecanismos de agendamento inteligente das atividades de manutenção, otimizando a alocação de recursos humanos e materiais. Esta funcionalidade contribui para uma gestão mais racional dos custos operacionais e para a maximização da disponibilidade dos equipamentos. Adicionalmente, o sistema fornece suporte à tomada de decisão através da disponibilização de relatórios analíticos e indicadores baseados em dados históricos e em

tempo real, permitindo aos gestores hospitalares uma visão mais informada e estratégica da gestão de ativos, (Ma, Cao, & Lu, 2024).

Com base na análise preditiva, que recorre a algoritmos de inteligência artificial para antecipar falhas antes que estas ocorram, torna-se possível realizar intervenções de manutenção de forma proativa e eficiente. Através da aprendizagem de padrões em dados históricos, a inteligência artificial é capaz de identificar desvios comportamentais e sinais precoces de anomalia, permitindo uma resposta atempada que minimiza o risco de avarias e otimiza a disponibilidade dos equipamentos, (Ma, Cao, & Lu, 2024).

Com base nos dados obtidos, o sistema é capaz de propor ações de manutenção automatizadas, tais como a programação de manutenções preventivas ou a realização de reparações em equipamentos que exibem comportamentos inconsistentes. Deste modo, o sistema reduz a dependência de manutenções corretivas urgentes, que são mais dispendiosas, e otimiza a alocação de recursos, tornando o processo de manutenção mais eficiente, (Ma, Cao, & Lu, 2024).

A implementação do sistema inteligente de manutenção resultou em melhorias significativas na eficiência operacional e na redução de custos associados à manutenção de equipamentos médicos. A monitorização em tempo real permitiu a deteção precoce de falhas, reduzindo o tempo de inatividade não planeado e assegurando a continuidade dos serviços médicos. A utilização de algoritmos de *machine learning* possibilitou a previsão de falhas e a otimização dos processos de manutenção, contribuindo para a redução do consumo de recursos humanos e materiais. Além disso, o suporte à decisão baseado em dados forneceu insights valiosos para os gestores hospitalares, promovendo uma gestão mais refinada e inteligente dos equipamentos médicos. Estas melhorias traduziram-se em benefícios económicos e sociais significativos para as instituições de saúde, (Ma, Cao, & Lu, 2024).

O estudo realizado por Shamayleh, Awad e Farhat, (Shamayleh, Awad, & Farhat, 2020), apresenta uma abordagem inovadora para a gestão da manutenção de equipamentos médicos, integrando tecnologias emergentes como a IoT e algoritmos de *machine learning*. Assim, foi desenvolvido um sistema de manutenção preditiva para equipamentos médicos para prever e classificar o estado do equipamento como conforme ou não conforme.

O estudo visa melhorar a gestão da manutenção desses equipamentos críticos, permitindo a previsão de falhas antes que ocorram, o que pode reduzir significativamente os custos e o tempo de inatividade dos equipamentos. Por outro lado, é relevante aumentar a confiabilidade e a disponibilidade dos equipamentos médicos, garantindo que estes permaneçam operacionais por mais tempo, essencial para o bom funcionamento dos serviços de saúde e minimizará os custos de detenção de inventário de peças de reposição, (Shamayleh, Awad, & Farhat, 2020).

Os autores recorreram a dados provenientes de um hospital nos Emirados Árabes Unidos, relativos a um analisador Vitros-Immunoassay (equipamento para realizar testes imunológicos de amostras biológicas, como sangue e urina) com base em eventos de manutenção e avaliação de criticidade como um bom candidato para a transição da manutenção corretiva para a

preditiva. Foi definido como objeto de estudo por ser o equipamento com um histórico de avarias elevada, (Shamayleh, Awad, & Farhat, 2020).

O estudo identificou como modo de falha principal o deslizamento da correia do braço dosador, causado pelo desgaste da correia e pelo movimento das polias (componentes mecânicos utilizados para transmitir força e movimento por meio de cordas, correias ou cabos). Para prever esse tipo de falha, foram utilizados sinais de vibração captados por acelerômetros sem fios. Esses dados foram processados por um analisador de sinais e analisados com o algoritmo Support Vector Machine (SVM), que, com base em dados históricos rotulados, classificou o estado do equipamento (normal ou defeituoso) com alta precisão, (Shamayleh, Awad, & Farhat, 2020).

A implementação da abordagem de manutenção preditiva demonstrou ser economicamente viável, proporcionando poupanças significativas nos custos de diagnóstico e reparação, que podem atingir até 25 %, e um período de retorno do investimento de um ano, (Shamayleh, Awad, & Farhat, 2020).

O estudo de Rahman et al., (Rahman, et al., 2023), propõe um modelo preditivo baseado em técnicas de inteligência artificial com o intuito de otimizar a gestão da manutenção de equipamentos médicos, contribuindo para a redução de custos operacionais e para o aumento da eficiência na prestação de cuidados de saúde. A investigação parte do pressuposto de que as estratégias convencionais de manutenção corretiva e preventiva não são, por si só, suficientemente eficazes para lidar com a crescente complexidade e criticidade dos equipamentos médicos modernos.

Para o desenvolvimento do modelo, os autores utilizaram um conjunto de dados real composto por 8.294 equipamentos médicos considerados críticos, representando 44 tipos distintos de equipamentos, recolhidos em 15 instituições de saúde na Malásia. Estes dados incluíam informações relevantes sobre o histórico de falhas, tempo de operação, condições de utilização e características técnicas dos equipamentos. Os equipamentos foram classificados em diferentes categorias de risco, de acordo com a sua probabilidade de falha ao longo do ciclo de vida, o que permite uma gestão mais eficaz dos recursos técnicos, logísticos e financeiros, (Rahman, et al., 2023).

O estudo comparou abordagens de *machine learning* e *deep learning* para a construção do modelo preditivo. Os resultados indicaram que os classificadores baseados em métodos *ensemble* — que combinam múltiplos modelos preditivos — superaram os algoritmos de *deep learning* no que respeita ao desempenho global. Os valores de exatidão, especificidade e precisão obtidos com o modelo de *machine learning* foram de 77,90 %, 87,60 % e 75,39 %, respetivamente. Após a realização de um processo de seleção de atributos (*feature selection*), que visou eliminar variáveis irrelevantes e melhorar a robustez do modelo, esses indicadores aumentaram para 79,50 %, 88,36 % e 77,43 %, (Rahman, et al., 2023).

A análise dos resultados demonstrou que o modelo preditivo tem potencial para antecipar falhas de forma fiável, permitindo a intervenção atempada das equipas de manutenção e a

prevenção de interrupções inesperadas dos serviços clínicos. De forma concreta, estima-se que a aplicação prática do modelo possa representar uma poupança anual de aproximadamente 326.330,88 ringgits malaios (MYR), o que reforça o valor económico da proposta, (Rahman, et al., 2023).

Os autores salientam que a integração de dados não estruturados, como registos textuais de manutenção, poderá potenciar a exatidão do modelo. Contudo, reconhecem que o processamento deste tipo de informação apresenta desafios significativos, nomeadamente em termos de complexidade e exigência computacional, (Rahman, et al., 2023).

6 Materiais e Métodos

6.1 Descrição do conjunto de dados

O conjunto de dados utilizado no presente estudo abrange todos os equipamentos de clientes, de Portugal, que foram submetidos a processos de reparação ou inspeção nas instalações do centro de serviços da Fujifilm, entre janeiro de 2020 e dezembro de 2024. O conjunto de dados inicial é composto por nove variáveis, incluindo o *Cliente* e o *Serial Number* (número de série do equipamento), que foram devidamente anonimizada através da atribuição de um identificador numérico aleatório, garantindo, assim, a proteção da privacidade e a confidencialidade, *Material Description* (a descrição do equipamento), *Priority text* (o tipo de reparação), *SrvOrdCost*, *Contract*, *Plant Section* (especifica se o equipamento é de um cliente ou de empréstimo), *Short text* (especifica o tipo de reparação) e *Column1*. A Figura 34 representa um pequeno extrato do conjunto de dados que será utilizado para a análise estatística.

Cliente	Material Description	Serial Number	Priority text	Short text	Column1	Modelo	Classification
6	EC-760R-V/L ROUT.COLONO SCOPE 700 SERIES	84	Major repair	CHA	30/01/20	EC-760R-V/L	MAJOR
117	EC-740T/L	211	Major repair	CHA	23/01/20	EC-740T/L	MAJOR
76	EG-760R ROUTINE GASTRO SCOPE 700 SERIES	258	Major repair	FCT	09/01/20	EG-760R	MAJOR
76	EG-760R ROUTINE GASTRO SCOPE 700 SERIES	257	Minor repair	CAP+RBS + NOZ	09/01/20	EG-760R	MINOR
46	EC-760R-V/L ROUT.COLONO SCOPE 700 SERIES	367	Minor repair	Small repair	06/01/20	EC-760R-V/L	MINOR
117	EG-760R ROUTINE GASTRO SCOPE 700 SERIES	646	Minor repair	RBS	10/01/20	EG-760R	MINOR

Figura 34 Amostra do conjunto de dados

6.2 Preparação do dados

Para a preparação dos dados utilizados na análise estatística, foi utilizada linguagem de programação *Python* com o objetivo de realizar o pré-processamento e a estruturação inicial do conjunto de dados, código apresentado no Anexo I. As etapas deste processo incluíram a

remoção de variáveis irrelevantes, a exclusão de registos referentes a equipamentos descontinuados e a extração dos modelos. As variáveis removidas foram *Contract*, *Plant Section* e *SrvOrdCost*.

Como referido anteriormente, procedeu-se à filtragem dos dados, restringindo o conjunto de análise aos modelos da série 700 (nomeadamente os modelos 720, 740 e 760). Adicionalmente, as datas de reparação foram organizadas de forma sequencial, permitindo o cálculo do intervalo médio entre intervenções para cada modelo, expresso em número de dias. O conjunto de dados resultante, devidamente tratado e estruturado, foi posteriormente exportado em ficheiros .csv para subsequente análise estatística, a qual foi desenvolvida em linguagem R.

O pipeline de pré-processamento e tratamento dos dados, que contempla estas etapas, está esquematizado na Figura 35.

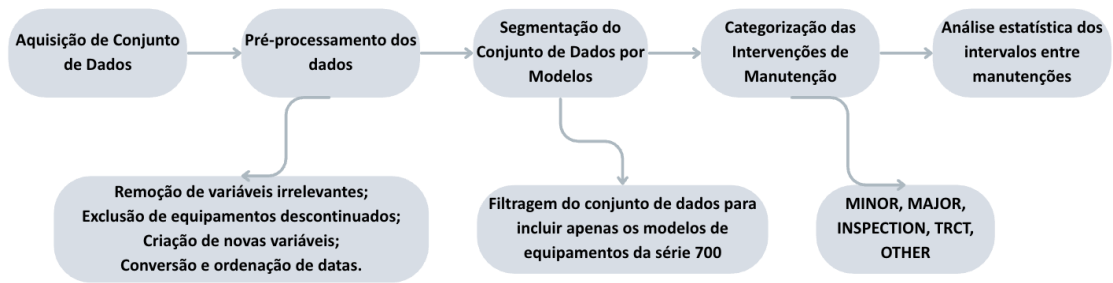


Figura 35 Pipeline pré-processamento de dados em Python

O conjunto de dados também inclui equipamentos das séries 500 e 600, além dos modelos da série 700, que foram submetidos a processos de reparação ou inspeção no centro de serviços da Fujifilm. Contudo, para os fins da análise estatística, a atenção será focada apenas nos equipamentos da série 700, conforme descrito na Tabela 1. Todos os estudos estatísticos serão realizados utilizando a linguagem de programação R, que é amplamente utilizada para a análise de dados devido à sua flexibilidade e capacidade de lidar com grandes volumes de informação, código apresentado no Anexo II.

Tabela 1 Equipamento Série 700

Tipo de Equipamento	Modelo
Colonoscópio	EC-720R/L
	EC-720R/I
	EC-740T/L
	EC-760R-V/L
	EC-760Z-V/L
	EC-760P-V/L
	EC-760R-V/I
	EC-760ZP-V/L
Gastroscópio	EG-720R
	EG-760R
	EG-760Z
	EG-760CT

A amostra abrange 95 clientes, sendo que cada cliente pode possuir um ou mais equipamentos, uma vez que a quantidade de números de série distintos é de 521, indica que, em média, cada cliente possui mais de um equipamento. Além disso, o conjunto de dados inclui doze modelos diferentes de equipamentos, dos quais oito são colonoscópios e os restantes são gastroscópios, Tabela 2.

Tabela 2 Quantidade de equipamentos por modelo no conjunto de dados

Modelo	Equipamentos Únicos
EC-720R/L	142
EG-760R	113
EC-760R-V/L	104
EG-720R	86
EG-760Z	25
EC-760Z-V/L	23
EC-720R/I	12
EC-760P-V/L	9
EG-760CT	3
EC-760ZP-V/L	2
EC-740T/L	1
EC-760R-V/I	1

Dessa forma, o conjunto de dados analisado para este estudo é composto por um total de 1876 amostras, das quais 117 destas correspondem a inspeções, 979 a manutenções *Minor repair* e 773 amostras a manutenções *Major repair*, como demonstra o Gráfico 1. A distribuição entre os tipos de intervenção revela uma predominância de reparações *minor*, o que pode sugerir que a maioria das avarias identificadas nos equipamentos tem uma natureza menos complexa, exigindo intervenções mais simples, como ajustes ou correções de menor escala.

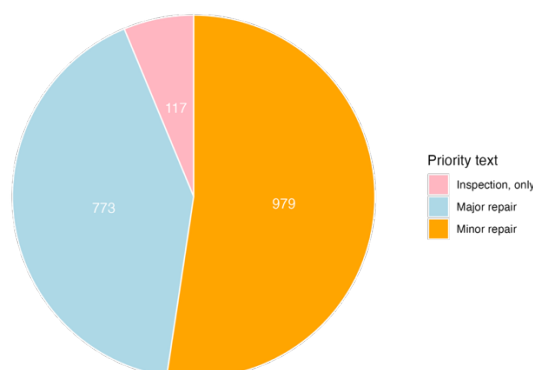


Gráfico 1 Distribuição de Reparções de 2020 a 2024

Considerando que o objetivo deste estudo se centra na gestão da manutenção de equipamentos médicos, o conjunto de dados inclui informações detalhadas acerca do tipo de

manutenção realizada em cada equipamento, na variável *Priority text*, classificada como “*Major repair*”, “*Minor repair*” ou “*Inspection, only*”. Adicionalmente, a variável *Short text* descreve de forma mais detalhada o tipo de reparação que o equipamento foi submetido. Por exemplo, *CHA* indica que a reparação foi mudada a câmara de vídeo do equipamento, o canal de trabalho (FCT), o tubo ar/água (AWT), o tubo de jato, se tiver, entre outras peças necessárias para a reparação, uma vez que a câmara de vídeo é a peça essencial da reparação, o conteúdo da variável *Short text* será *CHA*. Por fim, é registada a data de cada intervenção, permitindo uma análise temporal dos processos de manutenção realizados, este conteúdo está contido na variável *Column1* do conjunto de dados.

Com base na análise do conjunto de dados, foi identificado que o modelo EC-760R-V/L apresenta a maior frequência de manutenções, totalizando 489 reparações, como demonstrado no Gráfico 2. Conclui-se, ainda, que os modelos EC-760R-V/L, EG-760R e EC-720R/L são os modelos que apresentam uma maior quantidade de reparações nas instalações da Fujifilm.

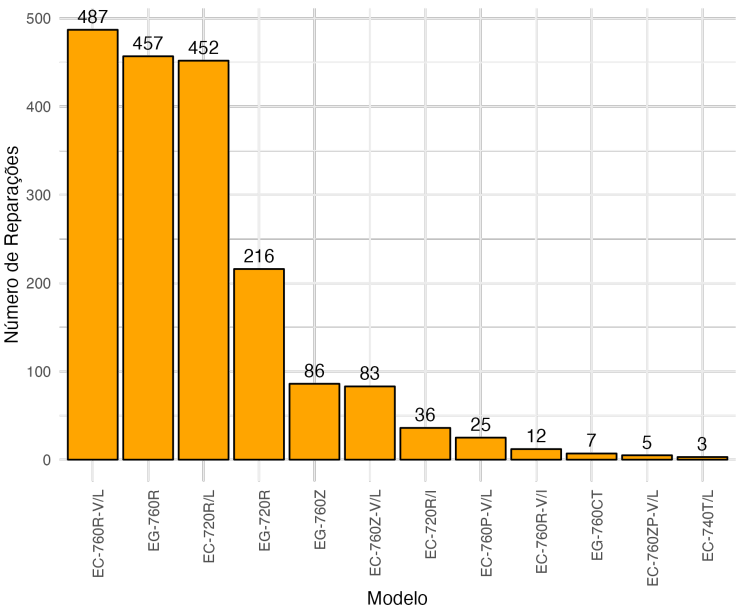


Gráfico 2 Quantidade de Reparções por Modelo

Ao realizar um estudo mais pormenorizado aos tipos de reparações do modelo EC-760R-V/L, conclui-se que 222 reparações realizadas no período compreendido entre 2020 e 2024 são classificadas como *Minor repair*, 234 reparações *Major repair* e as restantes inspeções.

Das 487 reparações realizadas, 120 foram classificadas como *small repairs*, estas estão relacionadas, por exemplo, com o ajuste de angulações e/ou folgas nos comandos, limpeza da lente da câmara de vídeo, desobstrução do canal ar/água. Além disso, 61 das reparações foram identificadas como do tipo *CHA*, envolvendo a substituição da câmara de vídeo do equipamento, Tabela 3.

Tabela 3 Classificação de reparação e respetiva quantidade

Classificação de Reparação	Quantidade
Small repair	120
CHA – Camera Head Assembly	61
CAP + RBS + NOZ	46
BSA – Bending Section Assembly	38
Inspection	33
RBS – Rubber Bending Section	32
FCT – Forceps Channel Tube	27
LGB – Light Guide Bundle	26
FSA – Flexible Section A	25
Complete Handle	21
FSB – Flexible Section B	20
AWT – Air Water Tube	9
DWA – Drum Wire Assembly	5
FSA + FSB	5
CHA + BSA	4
FSB + RBS	4
BSA + FSB	3
FSA + BSA	3
CAP	2
CHA + FSA	1
ISA (Insert Section Assembly) + FSB	1
Replacing all channels	1

Além da análise quantitativa total de equipamentos por modelo, também foi realizado o estudo do número de reparações anuais por modelo, conforme ilustrado no Gráfico 3. A partir da observação do gráfico, é possível identificar distintos padrões, uma cor mais escura representa um maior número de reparações. Os modelos EC-760R, EG-720R e EC-760R-V/L demonstraram uma tendência crescente no número de reparações anuais. De forma similar, o modelo EC-720R/L exibiu um aumento progressivo nas reparações, atingindo o seu pico máximo em 2024.

Por outro lado, os modelos EC-760ZP-V/L e EC-740T/L não registaram nenhuma intervenção de manutenção nos anos de 2021, 2022 e 2024, e nos anos de 2022 e 2024, respetivamente.

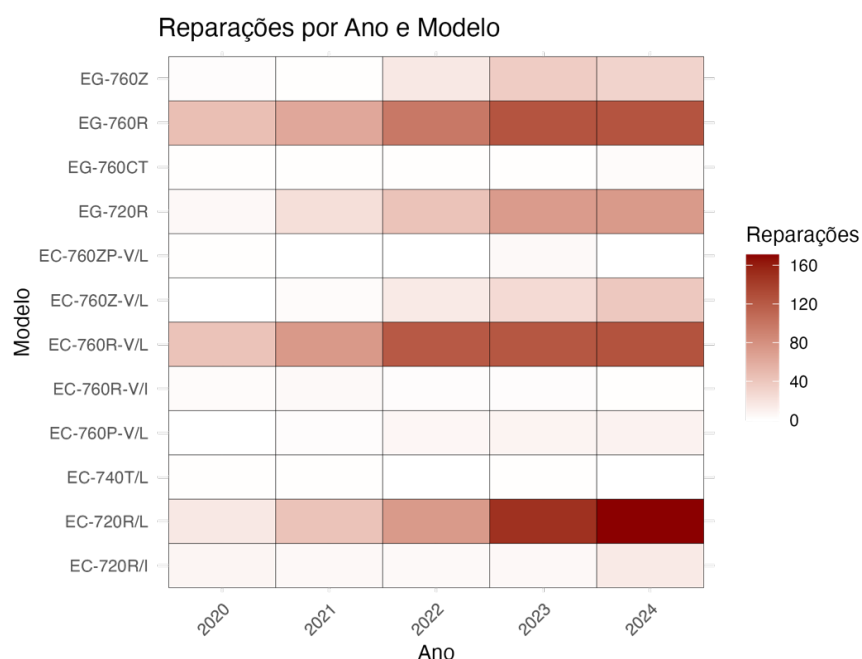


Gráfico 3 Quantidade de reparações por ano e por modelo

De acordo com as conclusões previamente apresentadas, os modelos EC-760R-V/L, EG-760R e EC-720R/L destacam-se como aqueles que registam o maior número de reparações. Esta elevada frequência de intervenções pode indicar maior suscetibilidade a falhas, caso apresentem um tempo médio entre reparações reduzido.

Uma vez que o conjunto de dados constitui informação em relação às datas em que os equipamentos são submetidos a reparações, foi possível estudar o tempo médio entre reparações. Assim, este indicador permite identificar padrões e tendências relevantes para a avaliação da confiabilidade e da necessidade de manutenção dos diferentes modelos.

Conforme se observa no Gráfico 4, o modelo EC-720R-L regista o menor tempo médio entre reparações (130 dias). À exceção dos modelos EC-740T/L e EC-760ZP-V/L, os demais equipamentos apresentam tempos médios entre reparações muito semelhantes, convergindo para um intervalo inferior a seis meses.

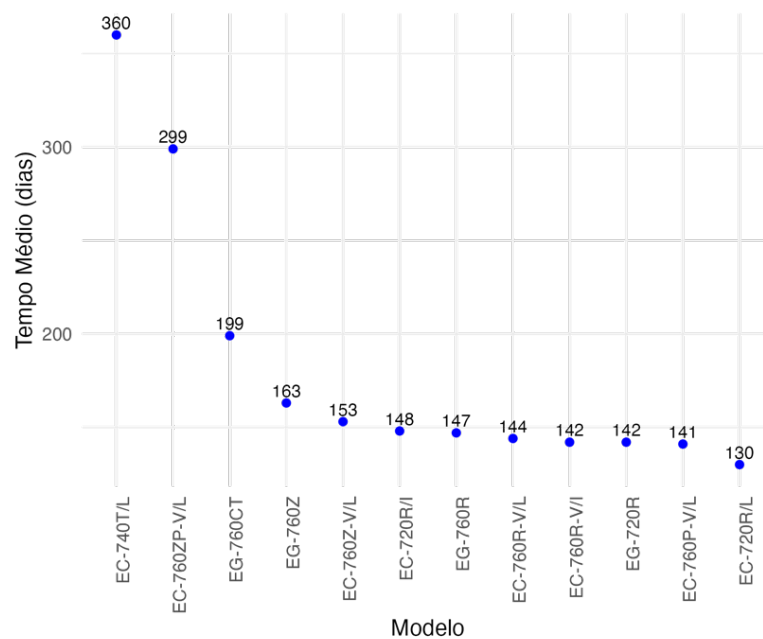


Gráfico 4 Tempo Médio entre Reparações por Modelo

O gráfico boxplot, Figura 36, constitui uma representação gráfica da distribuição de um conjunto de dados. Exibe estatísticas, tais como a mediana, os quartis e os possíveis *outliers*. O valor mínimo (limite inferior) corresponde à quantia mais baixa registada no conjunto de dados, excluindo os *outliers*. O primeiro quartil (Q1) é o valor abaixo do qual estão 25% dos dados, enquanto o terceiro quartil (Q3) é o valor acima do qual se encontram os 25% maiores valores. A mediana, também designada por Q2, corresponde ao ponto médio do conjunto de dados. Metade dos valores encontra-se abaixo deste e a outra metade encontra-se acima.

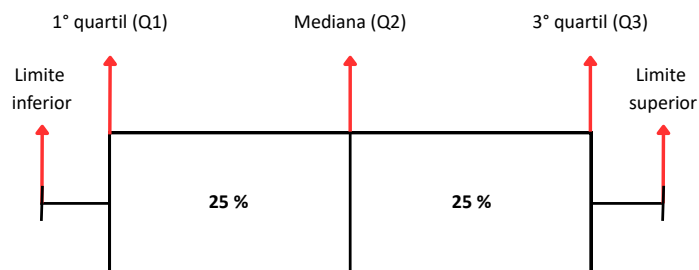


Figura 36 Exemplo de boxplot

Ao analisar o Gráfico 5, observa-se que todos os modelos endoscópicos apresentam um tempo entre reparações muito baixo ou próximo de zero, com exceção do modelo EC-740T/L. Neste caso, apresenta uma mediana visivelmente mais alta e uma grande variabilidade e cerca de 25% dos dados apresentam valores inferiores a aproximadamente 250 dias, indicando maior variabilidade nesse intervalo. Da mesma forma, o quartil superior, representa os 75 % do

conjunto de dados. Na maioria dos casos, 75 % dos dados são inferiores a aproximadamente 230 dias.

Também é possível observar a presença de *outliers*, o que aponta para casos isolados de desempenho atípico. Isso pode indicar, por exemplo, na variabilidade de uso dos equipamentos. Em contexto da análise estatística, seria de interesse a eliminação dos *outliers*. Contudo, considerando o tamanho reduzido do conjunto de dados, essa intervenção não foi realizada neste estudo. Pode-se inferir que, nos modelos em que foram identificados *outliers*, as reparações não ocorrem de forma periódica, caracterizando-se como reparações corretivas. Alternativamente, esses *outliers* podem refletir falhas inesperadas nos equipamentos.

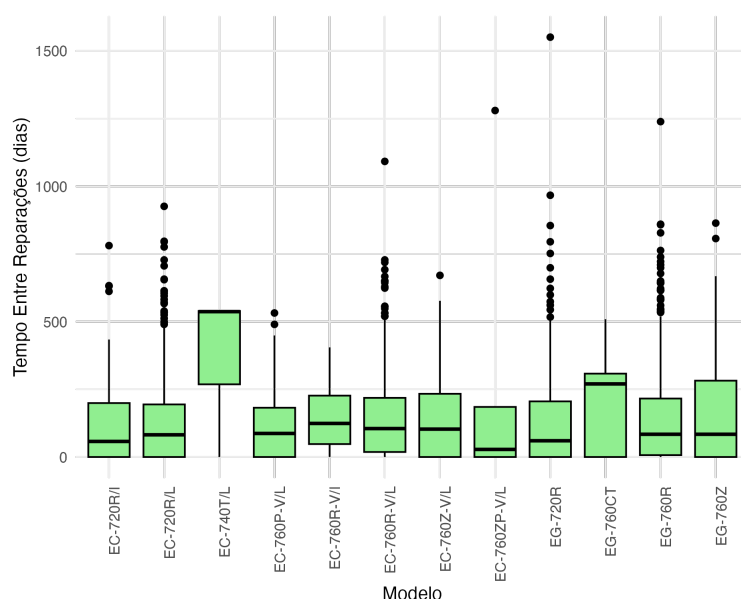


Gráfico 5 Box-plot tempo entre reparações

É importante salientar que um modelo com um elevado número absoluto de reparações pode, ainda assim, apresentar um bom desempenho em termos de fiabilidade, caso o número total de equipamentos em operação seja proporcionalmente elevado. De igual modo, um modelo com poucas reparações pode aparentar maior fiabilidade do que efetivamente possui, especialmente se a amostra for reduzida. Por esta razão, e com o intuito de garantir uma avaliação mais robusta, objetiva e comparável entre diferentes modelos, optou-se por realizar a análise com base em proporções relativas, nomeadamente a razão entre o número de reparações e o número total de unidades em operação por modelo.

O gráfico seguinte, Gráfico 6, apresenta a proporção de reparações por modelo, discriminadas pelos tipos de manutenção — inspeções, reparações *minor* e reparações *major*. O gráfico de barras foi representado devido à diferença significativa de equipamentos representados na amostra, assim o comprimento de cada barra é proporcional ao valor que representa.

Observa-se uma heterogeneidade significativa na frequência de intervenções entre os modelos, evidenciando que alguns modelos demandam um volume significativamente maior de

intervenções do que outros. Essa variação pode estar associada a múltiplos fatores, tais como a complexidade técnica dos equipamentos, a frequência operacional, o tempo de vida útil dos equipamentos e a eficácia dos planos de manutenção preventiva adotados para cada grupo.

Modelos que apresentam uma maior proporção de reparações *major* podem estar sujeitos a falhas estruturais ou problemas mais graves, sugerindo a necessidade de intervenções corretivas mais complexas.

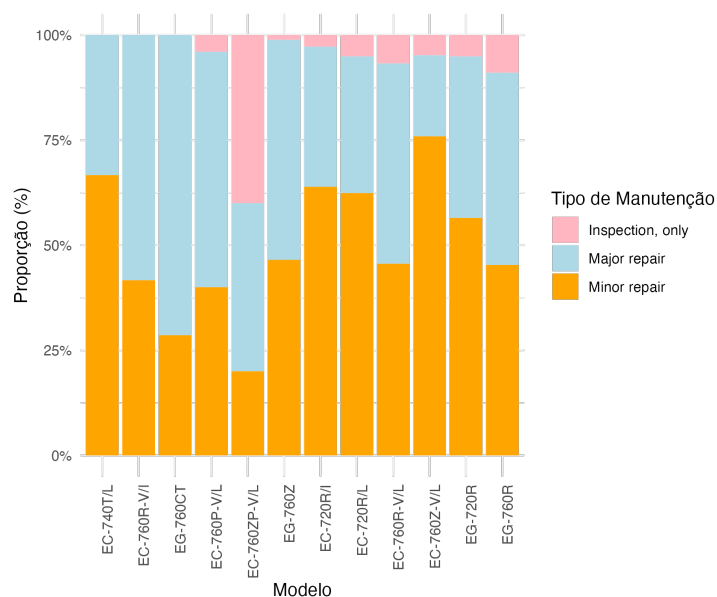


Gráfico 6 Proporção de Reparções *minor* e *major* e inspeções por Modelo

A análise da evolução das atividades de manutenção no período compreendido entre 2020 e 2024 evidencia um crescimento consistente dos três tipos de intervenção a partir de 2021, Tabela 4. Este padrão pode refletir tanto o envelhecimento progressivo dos equipamentos em operação, um aumento da utilização operacional ou da exigência funcional imposta aos mesmos.

As reparações do tipo *minor* apresentam um comportamento de crescimento contínuo ao longo do período, o que pode indicar uma maior atenção às falhas de menor gravidade. As reparações *major*, por sua vez, evidenciam um crescimento mais gradual. Em contraste, o volume de inspeções manteve-se praticamente constante ao longo do período analisado.

Tabela 4 Quantidade de Reparações por ano

Ano	Tipo de Reparação	Quantidade de Reparações
2020	Inspeção	18
	Major repair	53
	Minor repair	54
2021	Inspeção	19
	Major repair	99
	Minor repair	102
2022	Inspeção	22
	Major repair	162
	Minor repair	195
2023	Inspeção	32
	Major repair	226
	Minor repair	294
2024	Inspeção	26
	Major repair	233
	Minor repair	334

6.3 Pré-processamento do conjunto de dados

O pré-processamento de dados é uma etapa fundamental na preparação de conjuntos de dados para a aplicação de algoritmos de *machine learning*. Esta fase inclui a limpeza dos dados, a transformação dos dados e a seleção de características. Cada uma dessas etapas tem um papel crucial na melhoria da qualidade dos dados e no aumento da precisão dos modelos.

O conjunto de dados foi carregado a partir de um ficheiro no formato CSV utilizando o ponto e vírgula (;) como delimitador. A fase inicial de limpeza de dados envolveu a identificação e remoção de valores ausentes, inconsistentes ou duplicados dentro do conjunto de dados. Para lidar com valores ausentes, podem ser adotadas abordagens como a imputação (preenchimento com a média, mediana ou valores previstos) ou a exclusão, dependendo da quantidade e da relevância dos dados ausentes.

Foram identificadas e eliminadas todas as linhas contendo dados ausentes, recorrendo à função `na.omit()`, de forma a assegurar que os dados utilizados nas análises são completos e consistentes. No contexto do conjunto de dados utilizado, foram identificados trinta amostras com informações em falta das variáveis *Priority text* e, consequentemente, da *Short text*.

A partir da variável *Material Description* foi criada a variável *Modelo*, com o intuito de uniformizar a identificação dos modelos dos equipamentos. Após esta transformação, a variável original foi eliminada por se tornar redundante.

No sentido de garantir a consistência semântica entre variáveis, foi realizada uma normalização na variável *Short text*. Observou-se que, em determinadas instâncias, a *Priority text* indicava uma inspeção ("Inspection, only"), mas a *Short text* associada não refletia essa condição. Assim,

a variável *Manutencao* foi criada a partir da variável categórica *Priority text*, com codificação numérica conforme a intervenção: 2 – “*Major repair*”, 1 – “*Minor repair*” e 0 – “*Inspection*”.

Com base na combinação entre as variáveis *Short text*, *Classification* e *Modelo*, foi criada a variável *Pecas*, com o objetivo de descrever de forma mais detalhada os componentes envolvidos em cada intervenção de manutenção. A criação desta variável foi fundamentada na experiência prática obtida no local de estágio, onde foi possível compreender a associação entre determinadas descrições técnicas (*Short text*), a gravidade da falha (*Classification*) e o modelo do equipamento.

A variável *Column1*, que representa a data da intervenção, foi convertida para o formato de data (*Date*). Posteriormente, foi extraído o ano da intervenção (variável *Ano*) para possibilitar análises temporais. Para cada equipamento (*Serial Number*), calculou-se o tempo entre intervenções sucessivas (em dias), com o auxílio da função *diff()*, criando-se a variável *Tempo_Entre_Reparacoes*. A variável *Numero_Mantencao* foi criada para indicar o número sequencial da manutenção de cada equipamento.

6.4 Seleção de características

A seleção de características prepara os dados e auxilia na seleção de um conjunto de características para algoritmos de *machine learning*. O processo de seleção de características simplifica um modelo, identificando as características mais importantes e não redundantes no conjunto de dados. Reduzir o número de características aumenta a eficiência do modelo e melhora o desempenho. Os benefícios da seleção de características incluem a redução do sobreajuste e tempo de treino mais curtos.

Os métodos de filtragem consistem num conjunto de técnicas de seleção de características que se focam exclusivamente nos dados em si, sem consideração direta da otimização do desempenho do modelo. As variáveis de entrada são avaliadas de forma independente em relação à variável-alvo, com o objetivo de determinar qual apresenta a maior correlação.

Como é possível verificar nas bibliotecas de *machine learning*, nomeadamente na *Sklearn*, alguns exemplos de métodos de filtragem comuns são Mutual Information, Chi-square test, Fisher’s score e ANOVA. O método Mutual Information avalia a dependência entre variáveis medindo as informações obtidas sobre uma através da outra; o Chi-square test avalia a relação entre duas variáveis categóricas comparando os valores observados com os valores esperados; o Fisher’s score utiliza derivadas para calcular a importância relativa de cada característica para classificar os dados, uma pontuação mais alta indica maior influência; o método ANOVA determina se diferentes valores de características afetam o valor da variável alvo, (Belcic & Stryker, 2025).

Para identificar as variáveis mais relevantes para as análises, foi calculada a informação mútua entre as variáveis do conjunto de dados e as variáveis dependentes. A função *mutinformation()* foi utilizada para calcular a quantidade de informação compartilhada entre as variáveis. Os

resultados da informação mútua foram visualizados através de histogramas, destacando as variáveis que possuíam maior relevância para prever as variáveis alvo: *Manutencao*, *Pecas* e *Tempo_Entre_Reparacoes*.

6.5 Normalização dos dados: Z Score e Min-Max

Para além do estudo da importância de cada variável, realizou-se uma outra etapa de pré-processamento. Procedeu-se à normalização dos dados, necessária para a manipulação dos dados, que contribui para a otimização da precisão dos modelos de previsão e predição. A normalização de dados promove a consistência e a comparabilidade de diferentes modelos preditivos ao padronizar o intervalo de variáveis ou características independentes num conjunto de dados, originando resultados mais estáveis e fiáveis.

A normalização assegura que todas as variáveis do conjunto de dados contribuam de forma igual durante o processo de aprendizagem. Sem este procedimento, atributos com escalas numéricas superiores podem dominar a função de custo, reduzindo o impacto de características igualmente relevantes, mas com magnitudes inferiores.

A normalização foi realizada para ajustar as variáveis numéricas à mesma escala, facilitando o treino de modelos de *machine learning*. Foram aplicados dois métodos de normalização: Z-score e Min-Max.

O método Min-Max consiste na transformação linear dos dados originais. Os valores mínimos e máximos dos dados são obtidos e cada valor é alterado, utilizando a equação seguinte, Equação 3.

A fórmula em questão consiste na subtração do valor mínimo do valor original, de modo a determinar a distância entre o valor mínimo e o valor máximo. Posteriormente, a diferença é dividida pelo intervalo da variável (a diferença entre os valores máximo e mínimo). Esta divisão permite a quantificação da variável em termos de proporção em relação ao intervalo total. Em consequência, o valor normalizado situa-se entre 0 e 1.

Quando a característica x está no seu mínimo, o valor normalizado é 0. Esta situação deve-se ao facto de o numerador ser igual a zero. Em contraste, quando x está no seu máximo, o valor normalizado é 1, refletindo uma normalização a toda a escala. Para valores entre o mínimo e o máximo, o valor normalizado varia entre 0 e 1.

$$x_{normalizado} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

A normalização por Z-score, transforma os dados de modo que a média seja 0 e o desvio padrão 1. Este processo ajusta os valores dos dados com base na distância a que se desviam da média, medida em unidades de desvio padrão. Os valores são obtidos de acordo com a Equação 4, na qual X representa o valor original, μ a média do conjunto de dados e σ o desvio padrão.

A normalização por Z-score assegura que todas as características são tratadas de forma equivalente, facultando ao algoritmo a identificação de padrões de modo mais eficiente.

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

O pipeline de pré-processamento e tratamento dos dados, que contempla estas etapas, está esquematizado na Figura 37.

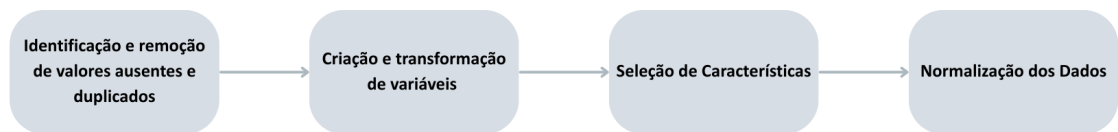


Figura 37 Pipeline pré-processamento de dados em R

7 Resultados e Discussão

O presente estudo envolveu a realização de uma análise estatística e preditiva aplicada a um conjunto de dados reais de manutenção de equipamentos médicos, com o objetivo de identificar padrões e relações relevantes entre variáveis associadas aos processos de manutenção. O estudo teve como foco a compreensão dos fatores que influenciam o tempo entre reparações, a frequência de intervenções, a previsão do tipo de manutenção futura e a necessidade de peças, contribuindo assim para a otimização da gestão de manutenção corretiva e preventiva.

7.1 Métodos estatísticos

No presente estudo, foram aplicados dois métodos estatísticos ANOVA (Análise de Variância) e o teste de Kruskal-Wallis com o objetivo de avaliar a influência de diferentes variáveis explicativas sobre três variáveis dependentes relacionadas à manutenção: Tempo Entre Reparções, Tipo de Manutenção e Peças utilizadas.

As análises estatísticas foram conduzidas com o objetivo de identificar relações significativas entre as variáveis do conjunto de dados e verificar se essas relações poderiam ser generalizadas para uma população mais ampla de equipamentos.

O teste ANOVA é um teste estatístico paramétrico, ou seja, assume que os dados seguem uma distribuição específica, são aplicáveis a apenas variáveis numéricas quantitativas e baseiam-se em parâmetros como a média e o desvio padrão. Permite verificar se existem diferenças estatisticamente significativas entre mais de duas amostras, (Sawyer, 2013).

O teste de Kruskal-Wallis é um teste estatístico não paramétrico utilizado para comparar três ou mais grupos independentes, a fim de determinar se existem diferenças estatisticamente significativas entre eles. Os testes não paramétricos são usados quando os dados não são

distribuídos na forma da curva de Gauss, mas são mais apropriados para amostras de pequena dimensão, (Chicco, Sichenze, & Jurman , 2025).

**7.1.1 Teste ANOVA e Kruskal-Wallis para a variável dependente
Tempo_Entre_Reparacoes**

A análise de variância (ANOVA) foi realizada para avaliar se existem diferenças estatisticamente significativas no tempo entre reparações com base em várias variáveis do conjunto de dados. A variável dependente considerada foi o tempo entre reparações. O teste ANOVA permite verificar se o tempo entre reparações varia significativamente com base nas diferentes categorias de cada variável explicativa.

A hipótese nula (H_0) do teste é: “não existem diferenças significativas no tempo entre reparações entre as diferentes categorias das variáveis explicativas”. A hipótese alternativa (H_1) é: “existe uma diferença significativa no tempo entre reparações entre as diferentes categorias das variáveis explicativas”.

Para a variável depende *Tempo_Entre_Reparações*, os resultados indicam que as variáveis *Priority text*, *Short text*, *Classification* e *Ano* apresentaram significância estatística consistente em ambos os testes ($p < 0,05$), o que evidencia uma associação estatisticamente importante com o tempo entre reparações. Essas variáveis podem, portanto, ser consideradas fatores explicativos relevantes desse indicador.

As variáveis *Cliente* e *Serial Number* não foram consideradas neste estudo devido a terem sido atribuídos valores aleatórios. A variável *Pecas* não foi incluída na aplicação do teste de Kruskal-Wallis por se tratar de uma variável categórica. Já a variável *Numero_Manutencao* não foi considerada devido à baixa variabilidade, abrangendo um intervalo de valores muito restrito.

A Tabela 5 apresenta os valores de *p-value* obtidos por meio dos testes ANOVA e Kruskal-Wallis para a variável dependente *Tempo_Entre_Reparacoes*.

Tabela 5 Resultados da Análises Estatísticas para a variável *Tempo Entre Reparções* por Variáveis Explicativas

Variável	p-value ANOVA	p-value Kruskal-Wallis
Priority text	6,68E-11	1,09E-8
Short text	<2,0E-16	<2,2E-16
Column1	<2,0E-16	1,18E-1
Modelo	< 2,0E-16	1,99E-1
Classification	3,63E-2	2,07E-7
Pecas	1,18E-1	-
Ano	1,88E-3	<2,2E-16
Numero_Manutencao	3,62E-1	-

7.1.2 Teste ANOVA e Kruskal-Wallis para a variável dependente Manutencao

No que se refere à variável dependente *Manutencao*, os resultados evidenciaram que a variável *Priority text* apresenta uma elevada significância estatística em ambos os testes aplicados (ANOVA e Kruskal-Wallis), o que reforça como fator explicativo neste contexto. Adicionalmente, as variáveis *Short text*, *Column1*, *Modelo*, e *Classification* demonstraram significância apenas no teste de Kruskal-Wallis.

A Tabela 6 apresenta os valores de *p-value* obtidos por meio dos testes ANOVA e Kruskal-Wallis para a variável dependente *Manutencao*.

Tabela 6 Resultados da Análises Estatísticas para a variável Manutencao por Variáveis Explicativas

Variável	p-value ANOVA	p-value Kruskal-Wallis
Priority text	< 2,0E-16	<2,2E-16
Short text	7,96E-1	<2,2E-16
Column1	9,14E-1	4,25E-04
Modelo	3,07E-1	9,28E-6
Classification	9,78E-1	<2,2E-16
Pecas	8,19E-1	-
Ano	2,32E-1	8,18E-1
Tempo_Entre_Reparacoes	9,66E-1	-
Numero_Manutencao	1,10E-1	-

7.1.3 Teste ANOVA e Kruskal-Wallis para a variável dependente Pecas

No caso da variável dependente *Pecas*, a análise foi conduzida exclusivamente por meio do teste de Kruskal-Wallis. Os resultados obtidos indicam que as variáveis *Priority text*, *Short text* e *Classification* apresentaram níveis de significância estatística extremamente elevados ($p < 0,05$), evidenciando uma associação substancial entre essas variáveis explicativas e o tipo de peças utilizadas nos processos de manutenção.

A Tabela 7 apresenta os valores de *p-value* obtidos por meio dos testes Kruskal-Wallis para a variável dependente *Pecas*.

Tabela 7 Resultados da Análises Estatísticas para a variável Peças por Variáveis Explicativas

Variável	p-value Kruskal-Wallis
Priority text	<2,2E-16
Short text	<2,2E-16
Column1	1,32E-1
Modelo	3,84E-1
Classification	<2,2E-16
Ano	1,56E-1

7.1.4 Método Mutual Information para as variáveis dependentes: Manutencao, Pecas e Tempo_Entre_Reparacoes

A análise de informação mútua entre as variáveis do conjunto de dados e a variável-alvo *Manutenção* permitiu identificar aspectos relevantes sobre a relevância das características para a previsão da manutenção. Variáveis como *Priority text* e *Short text*, com valores de *Mutual Information* elevados (0,8773 e 0,6755, respectivamente), demonstram uma forte dependência com a variável-alvo, sugerindo que estas características contêm informações cruciais para a previsão do tipo de manutenção, Tabela 8. Em contraste, variáveis como *Ano* (0,0060) e *Numero_Manutencao* (0,0080) apresentaram valores muito baixos, evidenciando uma dependência quase nula com a manutenção.

Os valores obtidos para a variável alvo *Pecas* obteve um valor de Mutual Information superior para as variáveis *Column1* e *Short text* (1,6624 e 2,3408, respectivamente), conclui-se que serão as variáveis que poderão prever com maior precisão a variável alvo. Por último, para a variável alvo *Tempo_Entre_Reparacoes*, as variáveis que obtiveram o valor mais elevado do estudo são *Column1*, *Numero_Manutencao* e *Pecas*.

A análise realizada tem como objetivo entender as variáveis dependentes que poderão ter maior valor estatístico nos modelos de previsão.

Tabela 8 Mutual Information para as variáveis alvo Manutencao, Pecas e Tempo_Entre_Reparacoes

Variável Independente	Variável Alvo: Manutencao	Variável Alvo: Pecas	Variável Alvo: Tempo_Entre_Reparacoes
	Mutual Information	Mutual Information	Mutual Information
Priority text	0,8773	0,5278	0,2709
Short text	0,6755	2,3408	1,1184
Column1	0,4795	1,6624	3,9237
Modelo	0,2275	0,2282	0,6674
Classification	0,5518	0,7711	0,3145
Manutencao	-	0,5278	0,2709
Pecas	0,5278	-	1,1200
Ano	0,0060	0,0603	0,4858
Tempo_Entre_Reparacoes	0,2709	1,1200	-
Numero_Manutencao	0,0080	0,5252	1,3591

7.2 Análise Preditiva da quantidade de reparações anuais (2025 a 2030)

7.2.1 Regressão Linear

Realizou-se uma análise preditiva da quantidade de reparações anuais no período de 2025 a 2030, com base nos dados históricos de 2020 a 2024. Para isso, procedeu-se à implementação de dois métodos: regressão linear e regressão polinomial de segundo grau.

A análise de regressão linear é utilizada para prever o valor de uma variável com base no valor de outra variável, Equação 5. A variável que se pretende prever (y) é designada por variável dependente, que neste caso é a quantidade de reparações. A variável que está a ser utilizada para prever o valor da outra variável é designada por variável independente, neste caso é o Ano (x).

$$y = ax + b \quad (5)$$

Com base nos resultados da regressão linear aplicada aos dados de quantidade de reparações por ano, verificou-se a existência de uma tendência crescente estatisticamente significativa ao longo do tempo. O modelo apresentou um coeficiente de aproximadamente 126,80, o que indica um aumento médio anual de cerca de 127 reparações ($x \approx 127$), Equação 6. Esta relação temporal revelou-se significativa ($p = 0,001988$), reforçando a confiabilidade da relação observada entre o tempo e o número de reparações.

$$y = 126,80x - 256015,80 \quad (6)$$

Além disso, o modelo obteve um coeficiente de determinação ($R^2 = 0,9722$), sugerindo que aproximadamente 97% da variação na quantidade de reparações pode ser explicada pela variável Ano.

7.2.2 Regressão Polinomial de segundo grau

Para fins comparativos, foi realizado o mesmo estudo recorrendo ao modelo de regressão polinomial de segundo grau, Equação 7. Onde y é a variável dependente, x a variável independente, β_0, β_1 e β_2 são os coeficientes para determinar a melhor curva e ε o termo de erro, que representa a diferença entre os valores observados e os valores previstos pela equação. A Equação 8 representa a equação representativa do modelo de regressão polinomial de segundo grau obtida.

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \varepsilon \quad (7)$$

$$y = 373,80 + 400,98x - 25,12x^2 + 44,57 \quad (8)$$

O modelo de regressão polinomial de segundo grau também evidenciou uma tendência de crescimento ao longo do tempo. O coeficiente linear do modelo polinomial revelou-se estatisticamente significativo ($p = 0,02402$), indicando que a variável “Ano” tem um efeito relevante na previsão do número de reparações.

O modelo polinomial teve um R^2 ajustado de 0,976, o que significa que 97,6% da variação nas reparações pode ser explicada pelo modelo polinomial, um valor ligeiramente superior ao do modelo linear.

Importa salientar que ambas as análises foram realizadas com um número reduzido de observações ($n = 5$), correspondentes aos anos de 2020 a 2024, o que pode limitar a robustez estatística das previsões.

Em resumo, os dois modelos indicam uma tendência crescente nas reparações ao longo dos anos, conforme demonstrado na Tabela 9. O modelo linear estabelece uma relação de proporcionalidade entre o número de reparações e o tempo, prevendo um aumento constante e regular. Em contrapartida, o modelo polinomial exibe um padrão distinto, com uma curva de crescimento mais pronunciada no início, que se estabiliza com o decorrer dos anos. As previsões do modelo polinomial de segundo grau indicam um número menor de reparações em comparação ao modelo linear, refletindo um crescimento mais moderado.

Tabela 9 Resultados obtidos pelos modelos de regressão linear e de regressão polinomial de segundo grau

Ano	Quantidade de Reparações pelo Modelo de Regressão Linear	Quantidade de Reparações pelo Modelo de Regressão Polinomial de 2º Grau
2025	754	707
2026	881	787
2027	1008	853
2028	1135	906
2029	1261	946
2030	1388	972

No Gráfico 7, está representado graficamente os resultados do modelo linear e do modelo polinomial para a previsão do número de reparações de 2025 a 2030.

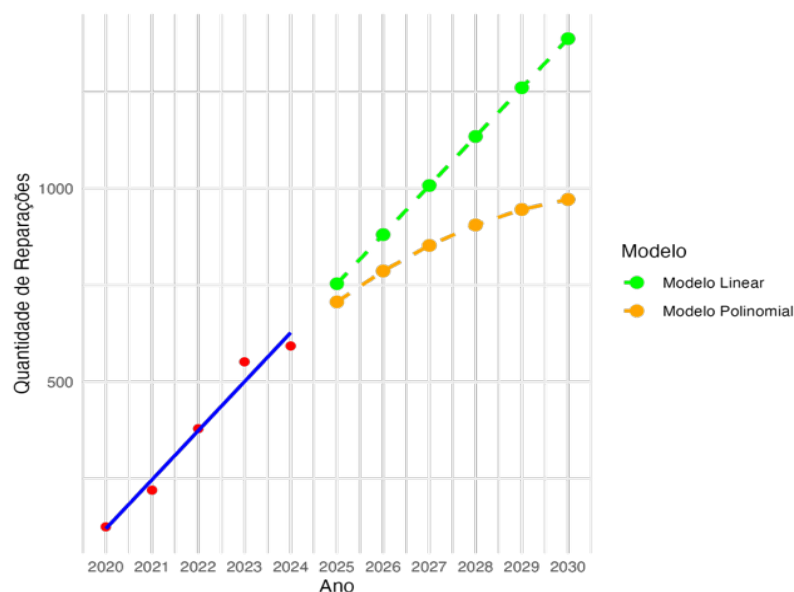


Gráfico 7 Modelo linear e modelo de regressão polinomial de segundo grau para prever a quantidade de reparações

7.3 Análise Preditiva do tempo entre reparações

Realizou-se um estudo adicional com o intuito de prever o tempo entre reparações para cada equipamento. Neste estudo recorreu-se a três modelos, modelo de regressão linear, o modelo de regressão linear com validação cruzada e o modelo Random Forest. No primeiro modelo, as variáveis *Column1*, *Numero_Mantencao*, *Ano*, *Modelo* e *Classification* foram utilizadas como variáveis independentes e a variável *Tempo_Entre_Manutencoes* a variável dependente.

7.3.1 Modelo de regressão linear

O modelo de regressão linear foi aplicado com o propósito de prever o intervalo de tempo entre manutenções consecutivas. Este modelo parte do princípio de que o tempo entre as reparações segue uma tendência linear.

Os resultados obtidos pelo modelo de regressão linear indicam que, à medida que o número de manutenções aumenta, o tempo entre reparações tende a diminuir, conforme apresentado na Equação 9. Onde y é o tempo entre reparações, x_1 a data de manutenção (*Column1*), x_2 o número de manutenções realizadas (*Numero_Mantencao*) e x_3 o ano (*Ano*), β_j são os coeficientes associados às categorias das variáveis *Modelo*, *Classification* e *Short_text*.

$$y = 91,08 + 0,0007518x_1 - 0,09816x_2 - 0,05145x_3 + \sum \beta_j \times \text{Variáveis Categóricas} \quad (9)$$

O valor de p indica que algumas variáveis independentes são estatisticamente significativas no modelo, como a data da manutenção (*Column1*, $p = 0,0299$), o número de manutenções

realizadas (*Numero_Mantencao*, $p < 0,05$), algumas categorias do Modelo (por exemplo, ModeloEC-720R/L, $p = 0,0253$).

O coeficiente de determinação ajustado é de 0,1242, indicando que o modelo explica aproximadamente 12,4 % da variação no tempo entre reparações.

7.3.2 Modelo de regressão linear com validação cruzada

O modelo de regressão linear com validação cruzada de 15 *folds* foi ajustado utilizando as variáveis independentes *Column1*, *Numero_Mantencao* e *Classification*. O conjunto de dados de treino continha 556 amostras e três variáveis preditoras. Os resultados mostram um erro quadrático médio (RMSE) de aproximadamente 0,95 e um coeficiente de determinação (R^2) de cerca de 0,11, indicando que o modelo explica aproximadamente 11% da variabilidade no tempo entre reparações. O erro absoluto médio (MAE) foi de aproximadamente 0,72.

É importante reconhecer que prever o tempo exato entre reparações de um equipamento é um procedimento complexo. Diversos fatores influenciam a necessidade de manutenção, como o modo de utilização, desgaste de utilização e quantidade de utilização dos equipamentos, não estão refletidos nos dados disponíveis. A ausência dessas variáveis pode justificar o baixo valor de R^2 observado.

7.3.3 Modelo Random Forest

Por último, recorreu-se ao modelo Random Forest. Inicialmente, foi realizada a extração de variáveis temporais a partir da variável *Column1*, nomeadamente o ano, mês e dia do mês. A coluna referente às peças associadas a cada reparação foi adaptada para um formato adequado ao modelo. Dado que cada reparação pode conter múltiplas peças, separadas pelo símbolo "+", foi utilizada a função *separate_rows* para dividir os valores contidos em uma ou mais colunas em linhas distintas, com base num delimitador previamente especificado, o carácter "+". Ao aplicar esta função, é possível reestruturar os dados de forma a facilitar a análise estatística, permitindo uma melhor manipulação e interpretação dos dados.

Posteriormente, foi aplicada uma codificação *multi-hot*, resultando na geração de colunas binárias que indicam a presença ou ausência de cada peça, permitindo ao modelo identificar padrões específicos associados a combinações de componentes. Esta técnica é utilizada para converter variáveis categóricas numa matriz binária, onde cada valor único de uma variável categórica recebe uma coluna distinta. Assim, cada peça do conjunto de dados é representada por uma coluna própria. Em cada linha da matriz, se o valor correspondente a uma categoria estiver presente, a coluna associada é marcada com o valor 1, enquanto todas as outras colunas dessa linha são atribuídas o valor 0. Esta abordagem permite representar dados categóricos de forma numérica, o que facilita a sua utilização em modelos de *machine learning*, garantindo que os algoritmos possam processar esses dados de forma eficiente.

O conjunto de dados foi segmentado em dois grupos: um para o treino (80%) e outro para o teste (20%). Foi estabelecido um conjunto de hiperparâmetros para a construção do modelo, variando o número de variáveis a serem consideradas em cada divisão (mtry) e o tamanho mínimo de árvores (nodesize). A variável nodesize foi estabelecida nos valores 5, 12 e 15, a variável mtry foi definida nos valores 2, 4 e 6 e o número de árvores (ntree) foi fixado em 4000.

O desempenho do modelo Random Forest foi avaliado através do erro quadrático médio (RMSE, Root Mean Squared Error), que quantifica a discrepância média entre os valores previstos e os valores reais do tempo entre reparações, Equação 10. Onde y_i é o vetor dos valores reais e $\hat{y}_i$ o vetor dos valores previstos.

Ao elevar os resíduos ao quadrado antes de calcular a média, o RMSE penaliza os erros maiores mais fortemente do que os menores. O RMSE é sempre não negativo, e valores mais baixos indicam um modelo mais adequado. O RMSE determina a precisão da previsão do modelo. Indica, em média, a distância entre as previsões e os valores reais, na mesma escala que a variável de resultado.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

O erro absoluto médio (MAE) também foi usado como recurso de comparação do desempenho do modelo. É uma métrica usada para medir a magnitude média dos erros entre os valores previstos e reais na previsão de séries temporais, Equação 11. Calcula a diferença absoluta entre cada valor previsto e o valor observado correspondente e, em seguida, calcula a média dessas diferenças. Ao contrário das métricas que elevam os erros ao quadrado (por exemplo, RMSE), o MAE trata todos os erros de forma igual. Onde x_i é o valor real e y_i o valor previsto. Em contraste com a métrica RMSE, o MAE atribui o mesmo peso a todos os erros, independentemente de sua magnitude, não penalizando de forma desproporcional os desvios mais elevados. Assim como o RMSE, valores mais baixos de MAE indicam melhor desempenho preditivo.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (11)$$

O modelo Random Forest desenvolvido para prever o tempo entre reparações apresentou um erro médio quadrático (RMSE) em z-score próximo de 0,90 e um erro médio absoluto (MAE) de 0,69. A partir dos resultados obtidos, observa-se que o valor de MAE é inferior ao do erro RMSE por ser menos sensível a *outliers*.

Ao converter os resultados para a escala original (dias), o RMSE variou em torno de 158 a 159 dias, enquanto o MAE ficou próximo de 122 a 123 dias. Isso significa que, em média, o modelo erra a previsão do tempo entre reparações em cerca de 4 a 5 meses.

Os resultados obtidos dos dois métodos estão representados na Tabela 10.

Tabela 10 Resultados obtidos RMSE e MAE para o estudo do tempo entre reparações

mtry	nodesize	RMSE_z	MAE_z	RMSE em dias	MAE em dias
1	22	0,9024948	0.6982063	159	123
3		0,9001597	0.6967562	158	123
4		0,9016232	0.6945071	159	122
1	25	0,9024037	0.6980124	159	123
3		0,8988214	0.6951224	158	122
4		0,8977872	0.6924679	158	122
1	32	0,9036455	0.6992206	159	123
3		0,8959625	0.6932302	158	122
4		0,8953568	0.6907251	158	122

7.4 Peças mais substituídas por modelo endoscópico

Com base na representação binária dos dados, tornou-se possível identificar as peças mais frequentemente substituídas em cada modelo de endoscópio, através da soma dos valores binários por modelo, resultando numa análise quantitativa da frequência de uso de cada componente, Tabela 11.

De acordo com a Tabela 11, conclui-se que a peça RBS é, de forma consistente, a mais frequentemente substituída. Esta elevada frequência sugere que a RBS pode estar sujeita a maior desgaste.

Tabela 11 Peças mais substituídas por modelo

Modelo	Peça	Quantidade
EC-720R/I	Sem peças	4
EC-720R/L	RBS	103
EC-740T/L	RBS	1
	Sem peças	
EC-760P-V/L	RBS	6
EC-760R-V/I	FCT	2
	AWJ	
	WJT	
	RBS	
	FSA	
	Sem peças	
EC-760R-V/L	RBS	122
EC-760Z-V/L	RBS	22
EC-760ZP-V/L	RBS	2
EG-720R	RBS	57
EG-760CT	RBS	3
EG-760R	RBS	114
EG-760Z	RBS	25

7.5 Análise Preditiva do tipo de manutenção

Adicionalmente, foi realizado um estudo com o objetivo de prever o tipo de manutenção a ser realizada, utilizando modelos de Random Forest e de redes neurais. Estas abordagens visam identificar, com base em dados históricos de intervenções, a categoria mais provável de manutenção futura.

7.5.1 Modelo Random Forest

Inicialmente, procedeu-se à preparação dos dados, assegurando que a variável alvo, *Manutencao*, estivesse devidamente codificada como variável categórica (fator). Em seguida, foram removidas variáveis potencialmente redundantes ou suscetíveis a causar vazamento de informação, tais como *Classification* e *Priority_text*.

O conjunto de dados foi dividido em subconjuntos de treino (80 %) e de teste (20 %), utilizando a função *createDataPartition()* da biblioteca *caret*, preservando a distribuição das classes da variável alvo entre os subconjuntos, o que é fundamental para garantir a representatividade e validade da avaliação do modelo.

Para a construção dos modelos, foi definido um conjunto de hiperparâmetros com diferentes valores para *mtry* (número de variáveis consideradas em cada divisão) e *nodesize* (tamanho mínimo dos nós). Definiu-se os valores 15, 20 e 22 para a variável *mtry* e 1, 2 e 5 para *nodesize*.

Os resultados obtidos indicaram que a exatidão variara entre 77,37 % e 78,10 %. O aumento substancial no desempenho dos modelos Random Forest demonstra a capacidade destes em identificar padrões significativos presentes nos dados históricos para a previsão da reparação. Adicionalmente, as métricas de precisão, *recall* e *F1-score* apresentaram valores relativamente próximos, indicando que o modelo mantém um equilíbrio adequado entre a identificação correta dos diferentes tipos de manutenção (*recall*) e a minimização de falsos positivos (precisão).

Observa-se ainda que o aumento do parâmetro *mtry* tende a melhorar ligeiramente o desempenho, sugerindo que a consideração de um maior número de variáveis em cada divisão da árvore contribui para uma melhor generalização. Por outro lado, o parâmetro *nodesize* não apresenta uma influência significativa no desempenho.

A Tabela 12 representa os resultados obtidos da exatidão (ou *accuracy*), precisão, *recall* e *F1-score* do modelo Random Forest para a previsão do tipo de manutenção da próxima intervenção de cada equipamento.

Tabela 12 Resultados obtidos de accuracy para o estudo de previsão do tipo de reparação

Mtry	nodesize	Accuracy	Precision	Recall	F1-score
15	1	0,7737	0,6566	0,7854	0,6561
20		0,7810	0,6625	0,7909	0,6623
22		0,7810	0,6625	0,7909	0,6623
15	2	0,7664	0,6623	0,7800	0,6482
20		0,7737	0,6582	0,7854	0,6544
22		0,7810	0,6625	0,7909	0,6623
15	5	0,7591	0,6520	0,7745	0,6393
20		0,7664	0,6542	0,7800	0,6468
22		0,7737	0,6582	0,7854	0,6544

7.5.2 Modelo de redes neurais

Para além do modelo Random Forest, procedeu-se à utilização do modelo de rede neuronal. Inicialmente, procedeu-se à preparação dos dados para o modelo de redes neurais, visto que este requer exclusivamente dados numéricos. As variáveis categóricas presentes no conjunto de dados, nomeadamente *Modelo*, *Priority_text*, *Short_text* e *Classification*, foram convertidas em variáveis *dummy* (codificação *one-hot*) através da função *dummyVars()* da biblioteca *caret*. Esta técnica consiste em representar cada variável por colunas binárias distintas, com valores 0 ou 1, permitindo assim a sua utilização no modelo.

Para garantir a avaliação adequada do desempenho do modelo, o conjunto de dados foi dividido aleatoriamente em dois subconjuntos: treino (80 %) e teste (20 %), usando a função *createDataPartition* da biblioteca *caret*. Foi definida uma fórmula de modelação na qual a variável alvo *Manutencao_target* é explicada por todas as variáveis preditoras, excluindo-se as variáveis originais categóricas e a variável alvo.

De seguida, foram treinados cinco modelos de redes neurais artificiais usando a biblioteca *neuralnet*, variando o número de neurónios na segunda camada oculta, enquanto a primeira camada oculta manteve-se fixa com 5 neurónios. Esta configuração de duas camadas ocultas com estruturas (5, i), onde i varia entre 1 e 5.

Após o treino, os modelos foram avaliados nos conjuntos de treino e teste. Calculou-se o erro quadrático médio da raiz (RMSE) das previsões, permitindo a comparação dos modelos segundo o seu desempenho preditivo.

Com base nos resultados obtidos, observa-se que os cinco modelos de rede neuronal, variando o número de neurónios na segunda camada oculta entre 1 e 5, apresentaram valores de RMSE praticamente idênticos. No conjunto de validação (teste), o erro médio quadrático (RMSE) foi aproximadamente de 0,58 nos testes de validação e de 0,59 nos testes de treino. Dado que os valores são próximos entre si, pode-se afirmar que os modelos não estão a sofrer de sobreajuste (*overfitting*).

Tabela 13 Resultados obtidos RMSE para o estudo de previsão do tipo de reparação

Modelo	Camadas Ocultas	RMSE de validação	RMSE de treino
1	(5,1)	0,5801052	0,5850053
2	(5,2)	0,5801052	
3	(5,3)	0,5801052	
4	(5,4)	0,5801053	
5	(5,5)	0,5801052	

7.6 Análise Preditiva das peças necessárias na futuras intervenções

Para realizar a previsão das peças que serão necessárias nas futuras intervenções de manutenção, dois modelos de redes neurais foram aplicados: um utilizando a biblioteca *neuralnet* e outro utilizando o *framework Keras*. Uma vez que os modelos de redes neurais requerem variáveis numéricas, começou-se por transformar as variáveis relevantes para o estudo (*Modelo*, *Priority_text*, *Short_text* e *Classification*) em numéricas, através da criação de variáveis *dummies*. A transformação foi feita utilizando a função *dummyVars()* da biblioteca *caret*, que aplica a codificação *one-hot*. Esta técnica cria uma nova coluna com valores binários (0 ou 1), permitindo que cada categoria seja representada de forma independente. Após a criação das novas variáveis, as colunas originais são removidas do conjunto de dados.

A variável alvo deste estudo são as peças, que foram anteriormente divididas em várias colunas através da codificação *multi-hot*. As restantes variáveis do conjunto de dados são definidas como as variáveis preditoras, à exceção das variáveis: *Numero_Manutencao*, *Mes* e *Dia_mes*.

Da mesma forma, o conjunto de dados foi dividido aleatoriamente em dois subconjuntos: treino (80 %) e teste (20 %).

7.6.1 Rede neuronal com a biblioteca neuralnet

O primeiro modelo de rede neuronal foi treinado utilizando a biblioteca *neuralnet*, aplicando uma arquitetura de rede com duas camadas ocultas. A primeira camada oculta foi configurada com 5 neurónios, enquanto a segunda camada oculta variou de 1 a 5 neurónios, gerando um total de 5 modelos distintos. Todos os modelos foram treinados com a mesma função de perda, o erro quadrático médio (RMSE) para avaliar o desempenho preditivo.

Os resultados obtidos para o RMSE de validação não variam significativamente entre os cinco modelos, aproximadamente um valor de 0,27. Assim, conclui-se que o aumento da complexidade do modelo, traduzido pela adição progressiva de neurónios na segunda camada oculta, não apresenta melhorias relevantes no desempenho do modelo. Adicionalmente, uma vez que os valores de RMSE de validação são muito próximos dos valores obtidos no conjunto

de treino, pode-se deduzir que o modelo não está a sofrer de sobreajuste (*overfitting*). Esta proximidade entre os erros de treino e validação sugere uma boa capacidade de generalização do modelo.

Tabela 14 Resultados obtidos RMSE para o estudo de previsão de peças

Modelo	Camadas Ocultas	RMSE de validação	RMSE de treino
1	(5,1)	0,2722033	0,2687639
2	(5,2)	0,2722046	
3	(5,3)	0,2722041	
4	(5,4)	0,2722037	
5	(5,5)	0,2722049	

7.6.2 Rede neuronal com a framework Keras

Para além dos modelos baseados na biblioteca *neuralnet*, foi também explorada a utilização de uma arquitetura de rede neuronal mais avançada, implementada com a biblioteca Keras.

A estrutura do modelo *Keras* foi definida utilizando uma rede neuronal sequencial, composta por duas camadas densas ocultas e uma camada densa de saída. A primeira camada densa contém 128 neurónios e utiliza a função de ativação ReLU. A camada de entrada possui um número de unidade igual ao número de variáveis preditoras do conjunto de dados. De seguida, foi adicionada uma camada de *dropout*, com uma taxa de 30 %. Esta camada foi usada com o objetivo de minimizar um sobreajuste do modelo.

A segunda camada densa contém 32 neurónios, com a mesma função de ativação ReLU. A camada de saída contém um número de unidades igual ao número total de peças, com a função de ativação sigmoide. A função sigmoide é adequada para problemas de classificação binária, uma vez que retorna valores entre 0 e 1, interpretados como a probabilidade da presença de cada classe.

Após a definição da arquitetura da rede, o modelo foi compilado com a função de perda *binary_crossentropy*. Essa função de perda mede a discrepância entre as previsões do modelo e os valores reais, penalizando as previsões incorretas.

O treino do modelo foi realizado utilizando os dados de treino, com um valor definido de 200 *epochs* e *batch size* de 64, o que determina o número de amostras processadas antes que o modelo atualize seus pesos. Ou seja, o modelo passou por todo o conjunto de dados de treino 200 vezes, atualizando seus parâmetros após cada lote de 64 amostras. Além disso, foi configurado um *early stopping* para interromper o treino caso o erro de validação não melhorasse após 15 *epochs* consecutivos. Esta técnica impede que o modelo se sobreajuste, pois interrompe o processo de treino quando o desempenho no conjunto de validação começa a diminuir.

Após o treino, o modelo foi avaliado no conjunto de dados de teste. As saídas foram transformadas em valores binários (0 ou 1) com um limiar de 0,35. Isso significa que, para cada classe, o modelo atribui 1 (presença) se a probabilidade predita for maior que 0,35, e 0 (ausência) caso contrário.

O valor da função de perda (*loss*) obtido foi de aproximadamente 0,3766, enquanto a exatidão (*accuracy*) foi aproximadamente 26,81 %. O erro quadrático médio (RMSE) foi cerca de 0,3274. Embora a exatidão obtida não seja muito elevada, o modelo apresenta um erro moderado.

8 Considerações Finais

Neste último capítulo são apresentadas as conclusões obtidas, bem como sugestões de trabalhos futuros e as principais limitações que surgiram.

8.1 Conclusões

A gestão da manutenção de equipamentos médicos, em particular dos equipamentos de endoscopia, desempenha um papel fundamental para assegurar o funcionamento adequado dos equipamentos médicos. No entanto, a gestão da manutenção de equipamentos médicos pode revelar-se complexa e prolongada. Essa complexidade surge devido a vários fatores que devem ser considerados na gestão da manutenção de equipamentos, como a frequência da manutenção, o tipo de manutenção necessária e a disponibilidade de peças sobresselentes.

O principal objetivo deste projeto é o desenvolvimento de modelos preditivos baseados em técnicas de *machine learning* para prever o tipo de manutenção seguinte para cada equipamento, o tempo até à próxima manutenção e auxiliar a gestão de peças sobresselentes. Os modelos permitem prever quando é que pode haver um maior número de reparações, o tipo de reparações e garantir que as peças necessárias estejam disponíveis quando necessário e reduzir os custos associados ao excesso ou à falta de peças.

De acordo com a Organização Mundial de Saúde o número de procedimentos endoscópicos, na Europa, deverá manter-se praticamente constante até 2030. Sendo que a quantidade de procedimentos é muito elevada, a necessidade de intervenções de reparação a equipamentos de endoscopia deverá aumentar ou manter-se constante. Por essa razão, os centros de reparação devem estar disponíveis para a receção de equipamentos endoscópicos e ter em stock as peças necessárias para reparação.

Para a previsão da quantidade anual de reparações entre 2025 e 2030, os modelos de regressão linear e polinomial demonstraram uma tendência crescente significativa ao longo do tempo. O modelo linear apresentou um coeficiente de determinação elevado, aproximadamente de 0,97, indicando que a variável ano explica a maior parte da variação na quantidade de reparações. O modelo polinomial teve um desempenho ligeiramente superior, sugerindo uma relação não estritamente linear.

Para a previsão do tempo entre reparações recorreu-se a três modelos estatísticos: regressão linear, regressão linear com validação cruzada e Random Forest. Nos dois modelos de regressão linear obteve-se valores para o coeficiente de determinação ajustado baixos, 0,12 e 0,11, respetivamente. O reduzido poder explicativo pode ser atribuído a fatores que não constam no conjunto de dados, como o modo de utilização dos equipamentos, o desgaste natural e a quantidade de uso. Por outro lado, o modelo Random Forest foi avaliado por meio de duas métricas: erro médio quadrático, RMSE, e erro médio absoluto, MAE, cujos valores foram obtidos na escala da variável alvo. O modelo apresentou um RMSE de 0,90 e um MAE de 0,69, que, ao serem convertidos para a unidade original (dias), corresponderam a erros médios de aproximadamente 158 a 159 dias (RMSE) e 122 a 123 dias (MAE).

No que diz respeito à classificação do tipo de manutenção, o modelo Random Forest mostrou-se eficiente, com uma taxa de exatidão próxima de 78 %, evidenciando a capacidade do modelo em identificar corretamente as categorias de manutenção. As redes neuronais aplicadas também apresentaram desempenho consistente e sem sinais de sobreajuste, confirmando a fiabilidade dos modelos.

Relativamente à previsão das peças sobresselentes necessárias para futuras intervenções, foram implementados dois modelos baseados em redes neuronais: biblioteca neuralnet e o Framework Keras. O primeiro modelo usou duas camadas ocultas com cinco neurónios na primeira camada e um número variável na segunda camada. No entanto, de acordo com os resultados obtidos, conclui-se que o número de neurónios na segunda camada não influencia o desempenho do modelo de forma considerável. O erro RMSE foi de aproximadamente de 0,27. O segundo modelo possui duas camadas densas ocultas (com 128 e 32 neurónios, respetivamente), uma camada de *dropout* e uma camada densa de saída com função de ativação sigmoide. Para este modelo obteve-se uma exatidão de 26,81 %, mas em contrapartida obteve um valor de RMSE de aproximadamente de 0,33.

O presente estudo demonstra a viabilidade da aplicação de técnicas de *machine learning* na previsão de intervenções de manutenção e na gestão de inventário de componentes sobresselentes no contexto dos equipamentos médicos.

8.2 Trabalhos Futuros

O enriquecimento do conjunto de dados, através da inclusão de variáveis relevantes, como o histórico detalhado de avarias, a frequência de utilização dos equipamentos, as peças orçamentadas e substituídas em cada reparação, o registo da queixa referente à reparação, o

tipo de utilização (uso intensivo ou uso regular) e condições ambientais de funcionamento poderá traduzir-se num aumento significativo da capacidade preditiva dos modelos e, consequentemente, numa gestão mais eficiente.

Para além dos modelos utilizados no estudo, a utilização de outros modelos de *machine learning*, a otimização dos hiperparâmetros dos modelos utilizados no estudo ou a combinação de vários modelos poderá melhorar a robustez e a precisão das previsões.

A implementação de um sistema de apoio à decisão baseado nos modelos preditivos poderá ser implementado para uma gestão mais eficiente, permitindo a recomendação de manutenção preventiva. Para facilitar a monitorização e interpretação dos resultados, poderá ser desenvolvido um *dashboard* interativo que permita visualizar em tempo real o estado dos equipamentos, as previsões de falhas, as necessidades de manutenção e a disponibilidade de peças.

Por fim, a validação dos modelos preditivos com um conjunto de dados da INNOVO de clientes de Espanha, França e Itália para verificar a sua aplicabilidade prática e impacto operacional, avaliar a sua eficácia na redução de falhas não planeadas.

Bibliografia

- (s.d.). Obtido de ISO - International Organization for Standardization: <https://www.iso.org/home.html>
- 90th Anniversary - History*. (2024). Obtido de Fujifilm Holdings Corporation: <https://holdings.fujifilm.com/special/90th/en/history/>
- Abd-Almonem, O. (s.d.). Endoscopy – the scientific principle behind a medical application. *Gharbiya STEM Physics Journal*.
- All cancers – Fact sheet*. (s.d.). Obtido de Global Cancer Observatory: <https://gco.iarc.who.int/media/globocan/factsheets/cancers/39-all-cancers-fact-sheet.pdf>
- Ang, T. L., Li, J. W., Wong, Y. J., Tan, Y.-L. J., Fock, K. M., Tan, M. T., . . . Wang, L. M. (2019). A prospective randomized study of colonoscopy using blue laser imaging and white light imaging in detection and differentiation of colonic polyps. *Endoscopy*.
- Belcic, I., & Stryker, C. (Março de 2025). *What is feature selection?* Obtido de <https://www.ibm.com/think/topics/feature-selection>
- Boeriu, A., Boeriu, C., Drasovean, S., Pascarenco, O., Mocan, S., Stoian, M., & Dobru, D. (2015). Narrow-band imaging with magnifying endoscopy for the evaluation of gastrointestinal lesions. *World J Gastrointest Endosc.*, 110-120.
- Boese, A., Wex, C., Croner, R., Liehr, U. B., Wendler, J. J., Weigt, J., . . . Illanes, A. (2022). Endoscopic Imaging Technology Today. *Diagnostics*, 12(5).
- Bray, F., Laversanne, M., Weiderpass, E., & Soerjomataram, I. (2021). The ever-increasing importance of cancer as a leading cause of premature death worldwide. *An International Interdisciplinary Journal of the American Cancer Society*, 127(16), 3029-3030.
- Browning, C. M., Cloutier, R., Rich, T. C., & Leavesley, S. J. (2022). Endoscopy Lifetime Systems Architecture: Scoping Out the Past to Diagnose the Future Technology. *Systems*, 10(5).
- Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. d., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137.
- Chen, S., Cao, Z., & Prettnner, K. (2023). Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. *JAMA Oncol.*, 9(4), 465–472.

Corporation, F. (s.d.). *Healthcare*. Obtido de Fujifilm Global: <https://global.fujifilm.com/en/about/corporate/field/healthcare>

Decreto-Lei n.º 145/2009, de 17 de junho. (2009). Portugal.

Endoscopy Procedures Estimates Market (2025 - 2030). (s.d.). Obtido de <https://www.grandviewresearch.com/industry-analysis/endoscopy-procedures-estimates-market-report#>

Fujifilm. (s.d.). *Discovery Light Enhanced Endoscopy*. Obtido de <https://sw.dk/wp-content/uploads/2024/08/eluxeo-multi-light-technology-2.pdf>

Fujifilm distinguida como empresa solidária pela Liga Portuguesa Contra o Cancro. (7 de fevereiro de 2025). Obtido de Health News: <https://healthnews.pt/2025/02/07/fujifilm-distinguida-como-empresa-solidaria-pela-liga-portuguesa-contra-o-cancro/>

Fujifilm. (s.d.). *Endoscopy processors - 7000 system*. Obtido de Fujifilm Healthcare: <https://www.fujifilm.com/tw/en/healthcare/endoscopy/endoscopy-processors/7000system>

Gerges, C., Sr, H. N., Ishaq, S., Sivanathan, V., Galle, P. R., Neuhaus, H., & Neumann, H. (2021). Evaluation of a novel colonoscope offering flexibility adjuster – a retrospective observational study. *Therapeutic Advances in Gastroenterology*.

Gounella, R., Granado, T. C., Junior, O. H., Luporini, D. L., Gazziro, M., & Carmo, J. P. (2023). Endoscope Capsules: The Present Situation and Future Outlooks. *Bioengineering*, 10(12).

Ikematsu, H., Sakamoto, T., Togashi, K., Yoshida, N., Hisabe, T., Kiriya, S., . . . Kato. (2017). Detectability of colorectal neoplastic lesions using a novel endoscopic system with blue laser imaging: a multicenter randomized controlled trial. *Gastrointestinal Endoscopy*, 86(2), 386–394.

International Organization for Standardization. (2016). *ISO 13485:2016 - Medical devices — Quality management systems — Requirements for regulatory purposes*. Switzerland.

ISO - International Organization for Standardization. (s.d.).

Izadi, A., Bakhshali, M. A., & Sarrafzadeh, O. (2023). An Intelligent System for Management of Medical Equipment Maintenance. *Journal of Patient Safety and Quality Improvement*, 11(3), 161-167.

Jamshidi, A., Rahimi, S. A., & Ait-kadi, D. (2014). Medical Devices Inspection and Maintenance; a literature Review. *Proceedings of the 2014 Industrial and Systems Engineering Research Conference* (pp. 3895–3904). Institute of Industrial and Systems Engineers (IISE).

- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- Khider, M. O., & Hamza, A. O. (2023). Medical Equipment Maintenance Management System Design and Implementation. *Journal of Clinical Engineering*, 43(3), 107-117.
- Kleisiaris, C. F., Sfakianakis, C., & Papathanasiou, I. V. (Março de 2014). Health care practices in ancient Greece: The Hippocratic ideal. *Journal of Medical Ethics and History of Medicine*, 7.
- Liao, L., & Köttig, F. (2016). A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Applied Soft Computing*, 44, 191-199.
- Life science plant ISO 13485 certified. (s.d.). Obtido de <https://www.fujifilm.com/ef/en/news/iso-certification-lsm>
- Lin, E. W., Karakasheva, T. A., Hicks, P. D., Bass, A. J., & Rustgi, A. K. (2016). The tumor microenvironment in esophageal cancer. *Oncogene*, 5337–5349.
- Ma, G., Cao, B., & Lu, H. (2024). Design and Implementation of Intelligent Medical Equipment Maintenance System - Double Consideration of Improving Efficiency and Reducing Cost. *International Journal of Big Data Intelligent Technology*, 5(1), 201-209.
- Marcelissen, T., & Vijgen, G. H. (2023). The endoscope. *British Journal of Surgery*, 110(12), 1641-1643.
- Neumann, H., Kreft, A., Sivanathan, V., Rahman, F., & Galle, P. R. (2021). Evaluation of novel LCI CAD EYE system for real time detection of colon polyps. *PLOS ONE*, 16(8).
- Osawa, H., Miura, Y., Takezawa, T., Ino, Y., Khurelbaatar, T., Sagara, Y., . . . Yamamoto, H. (2018). Linked Color Imaging and Blue Laser Imaging for Upper Gastrointestinal Screening. *Clin Endosc.*, 51(6), 513-526.
- Painter, F. R., Baretich, M. F., & Velazquez-Berumen, A. (s.d.). *Medical equipment maintenance programme overview WHO Medical device technical series*.
- Qualidade, I. P. (s.d.). *Metrologia*. Obtido de Instituto Português da Qualidade (IPQ): <https://www.ipq.pt/metrologia/>
- Quality management. (s.d.). Obtido de Wako Chemicals: <https://www.wako-chemicals.de/en/quality-management.php>
- Raghavendra, P., & Pullaiah, T. (2018). Chapter 4 - Biomedical Imaging Role in Cellular and Molecular Diagnostics. Em P. Raghavendra, & T. Pullaiah, *Advances in Cell and Molecular Diagnostics* (pp. 85-111). Academic Press.

- Rahman, N. H., Zaki, M. H., Hasikin, K., Razak, N. A., Ibrahim, A. K., & Lai, K. W. (2023). Predicting medical device failure: a promise to reduce healthcare facilities cost through smart healthcare management. *PeerJ Computer Science*.
- Rutter, M. D., Senore, C., Bisschops, R., Domagk, D., Valori, R., Kaminski, M. F., . . . Dinis-Ribeiro, M. (2015). The European Society of Gastrointestinal Endoscopy Quality Improvement Initiative: developing performance measures. *United European Gastroenterol Journal*, 4(1), 30-41.
- Săftoiu, A., Hassan, C., Areia, M., Bhutani, M. S., Bisschops, R., Bories, E., . . . Me. (2020). Role of gastrointestinal endoscopy in the screening of digestive tract cancers in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*, 52(4), 293-304.
- Santos, C., Malaman, D., Yoshida, N., Pereira-Lima, J. C., Onófrío, F. Q., Furlan, R. G., . . . Malafaia, O. (2018). Blue laser imaging: a new image-enhanced endoscopy for the diagnosis of colorectal lesions. *Eur J Gastroenterol Hepatol.*, 30(12), 1514-1520.
- Serradilla, O., Zugasti, E., Rodriguez, J., & Zurutuza, U. (2022). Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. *Applied Intelligence*, 52, 10934-10964.
- Shamayleh, A., Awad, M., & Farhat, J. (2020). IoT Based Predictive Maintenance Management of Medical Equipment. *Journal of Medical Systems*, 44.
- Ucin, N. M., Nimpagaritse, M., Maru, S. M., Makhado, M., & Tebeka, S. M. (2024). Practices and factors affecting on-site medical equipment maintenance at Wau Teaching Hospital, South Sudan. *BMC Public Health*, 24.
- Velliangiri, S., Alagumuthukrishnan, S., & Joseph, S. I. (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Computer Science*, 104-111.
- Yazici, İ., Shayea, I., & Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. *Engineering Science and Technology, an International Journal*, 44.

Anexos

Anexo I

```
import pandas as pd

# Carregar os dados
data = pd.read_csv('/Users/anafernandes/Desktop/TESE/DATA_ENDO.csv',
delimeter=';', on_bad_lines='skip')

print("Column names:", data.columns)

# Eliminar as colunas não relevantes - Contrato, Plant section e SrvOrdCost
data = data.drop(columns=["Contract", "Plant section", "SrvOrdCost"],
errors='ignore')

# Remover espaços extras nos nomes das colunas
data.columns = data.columns.str.strip()

# Eliminar os equipamentos descontinuados, não são relevantes para o estudo
data = data[~data['Material
Description'].str.contains(r'discontinued|disc\.', case=False, na=False)]

# Criar a nova coluna Modelo
import re

def extract_model(description):
    if isinstance(description, str):
        match = re.search(r'([A-Za-z0-9\-\./]+)', description)
        return match.group(1) if match else None
    return None

data['Modelo'] = data['Material Description'].apply(extract_model)
print(data[['Material Description', 'Modelo']].head())

data.to_csv('/Users/anafernandes/Desktop/TESE/nova_data.csv', index=False,
sep=';')

# Função para contar todos os modelos
def count_all_models(df):
    model_names = df['Material Description'].str.extract(r'([A-Za-z0-9\-\./]+)').dropna()[0].unique()
    model_counts = {model: df['Material Description'].str.contains(model,
case=False, na=False).sum() for model in model_names}
    return model_counts

model_counts = count_all_models(data)

for model, count in model_counts.items():
    print(f"Quantidade de '{model}' é: {count}")

# Filtrar o conjunto de dados para apenas modelos 720, 740 e 760 (Modelos
700's)
```

```

data_filtered = data[data['Material
Description'].str.contains(r'720|740|760', case=False, na=False)].copy()

model_counts_filtered = count_all_models(data_filtered)

for model, count in model_counts_filtered.items():
    print(f"Quantidade de '{model}' no conjunto de dados filtrado é:
{count}")

data_filtered.to_csv('/Users/anafernandes/Desktop/TESE/filtered_data_720_74
0_760.csv', index=False, sep=';')

# Número de equipamentos por cliente
equipment_per_client = data_filtered.groupby('Cliente')['Serial
Number'].nunique()
print(equipment_per_client)

# Exibir os termos únicos na coluna Short text
print(data_filtered['Short text'].unique())

# Função para classificar os termos na nova coluna Classification
def classify_short_text(row):
    if not isinstance(row, str):
        return 'OTHER'

    major_terms = ['ISA', 'CHA', 'FCT', 'BSA', 'FCT + FSB', 'FSA', 'AWJ',
                    'BSA + DWA', 'FSA+BSA', 'LGB', 'CHA + BSA', 'ISA+FSB',
                    'FSA+FSB', 'CHA + FSA',
                    'Replacing all channels', 'BSA+FSB', 'AWJ + FSB', 'CHA +
FSM',
                    'LGB + FSB']
    minor_terms = ['CAP+RBS + NOZ', 'Small repair', 'CAP', 'RBS',
                    'FSB+RBS', 'DWA', 'FSB']
    trct_terms = ['Push-Pull FCT', 'Push-Pull AWJ']
    inspection_terms = ['Inspection']

    if any(term in row for term in major_terms):
        return 'MAJOR'
    elif any(term in row for term in minor_terms):
        return 'MINOR'
    elif any(term in row for term in inspection_terms):
        return 'INSPECTION'
    elif any(term in row for term in trct_terms):
        return 'TRCT'
    else:
        return 'OTHER' # Caso não corresponda a nenhum termo, classificar
como 'OTHER'

data_filtered.loc[:, 'Classification'] = data_filtered['Short
text'].apply(classify_short_text)

print(data_filtered.head())
data_filtered.to_csv('/Users/anafernandes/Desktop/TESE/novo_dataset_com_cla
ssification.csv', index=False, sep=';')

# Contagem das classificações
classification_counts = data_filtered['Classification'].value_counts()
print(classification_counts)

```

```

# Adicionar a coluna Model (apenas modelos 720, 740 e 760)
data_filtered = data[data['Material
Description'].str.contains(r'720|740|760', case=False, na=False)].copy()
data_filtered.loc[:, 'Model'] = data_filtered['Material
Description'].str.extract(r'(720|740|760)')
data_filtered.loc[:, 'Classification'] = data_filtered['Short
text'].apply(classify_short_text)

# Contagem por classificação e modelo
classification_counts_by_model = data_filtered.groupby(['Model',
'Classification']).size().unstack(fill_value=0)
print(classification_counts_by_model)

data_filtered = data_filtered.rename(columns={'Column1': 'Repair Date'})

data_filtered['Repair Date'] = pd.to_datetime(data_filtered['Repair Date'],
errors='coerce')
data_filtered = data_filtered.sort_values(by=['Serial Number', 'Repair
Date'])

# Calcular a diferença entre as datas de reparação para cada equipamento
data_filtered['Repair Date Difference'] = data_filtered.groupby('Serial
Number')['Repair Date'].diff()
print(data_filtered[['Serial Number', 'Repair Date', 'Repair Date
Difference']].head())
data_filtered.to_csv('/Users/anafernandes/Desktop/TESE/DATA.csv',
index=False, sep=';')

# Calcular a diferença entre as datas de reparação para cada equipamento
data_filtered['Repair Date Difference'] = data_filtered.groupby('Serial
Number')['Repair Date'].diff()
data_filtered['Repair Date Difference (Days)'] = data_filtered['Repair Date
Difference'].dt.days
avg_repair_time_per_model = data_filtered.groupby('Model')['Repair Date
Difference (Days)'].mean().astype(int)

print(avg_repair_time_per_model)
avg_repair_time_per_model.to_csv('/Users/anafernandes/Desktop/TESE/tempo_me
dio_reparacao_por_modelo_dias_inteiros.csv', sep=';')

```

Anexo II

```
library(readr)
library(psych)
library(ggplot2)
library(dplyr)
library(infotheo)
library(randomForest)
library(caret)
library(lubridate)
library(tidyr)
library(neuralnet)

# ----- bibliotecas -----

libs <- c("tidyverse", "survival", "survminer", "MASS", "lme4",
"randomForest",
"arules", "cluster", "factoextra", "tsibble", "fable",
"forecast", "qcc")

installed <- rownames(installed.packages())

for (p in libs) {
  if (!(p %in% installed)) {
    try(install.packages(p), silent = TRUE)
  }
}

for (p in libs) {
  if (p %in% rownames(installed.packages())) {
    library(p, character.only = TRUE)
  } else {
    message(paste("Pacote", p, "não está instalado."))
  }
}

# Definir a semente para reprodutibilidade
set.seed(123)

# Carregar os dados CSV, com delimitador ;
Data <-
read.csv('/Users/anafernandes/Desktop/TESE/novo_dataset_com_classification.
csv')
Data <-
read_delim("/Users/anafernandes/Desktop/TESE/novo_dataset_com_classification.csv", delim = ";")

# Verificar valores ausentes
colSums(is.na(Data))

# Eliminar as linhas com valores ausentes
Data <- na.omit(Data)

# Características do dataset
summary(Data)
describe(Data)
length(unique(Data$Cliente)) # número de clientes únicos
length(unique(Data$`Serial Number`)) # número de equipamentos únicos
```

```

length(unique(Data$Modelo)) #número de modelos únicos

# número de equipamentos únicos por modelo
equipamentos_por_modelo <- Data %>%
  group_by(Modelo) %>%
  summarise(Equipamentos_Unicos = n_distinct(`Serial Number`)) %>%
  arrange(desc(Equipamentos_Unicos))

print(equipamentos_por_modelo)

# Remover coluna não necessária
Data <- subset(Data, select = -`Material Description`)

# Há algumas incongruência, onde priority text apresenta inspeção, mas
short text não corresponde a inspeção,
# assim em todas essas situações vai ser alterado para inspeção
# Estas situações podem ser justificadas por engano ou reparações rápidas
realizadas durante a inspeção
Data <- Data %>%
  mutate(`Short text` = ifelse(`Priority text` == "Inspection, only" &
`Short text` != "Inspection", "Inspection", `Short text`))

# Criar uma nova variável para tipo de manutenção (0 - inspeção, 1 - minor
repair, 2- major repair)
# Para fins de previsão
Data$Manutencao <- ifelse(Data$`Priority text` == "Major repair", 2,
                          ifelse(Data$`Priority text` == "Minor repair", 1,
0))

# Criar uma nova variável - Peças
# Para fins de previsão
# Esta variável é criada de forma manual, depende das variáveis: short
text, classification e modelo do equipamento

Data <- Data %>%
  mutate(Pecas = case_when(
    `Short text` == "ASPI" ~ "ASPI",
    `Short text` == "AWJ" ~ "AWT",
    `Short text` == "BSA" & `Classification` == "MINOR" ~ "BSA + RBS",
    `Short text` == "BSA" & `Classification` == "MAJOR" & `Modelo` != "EG-
720R" ~ "BSA + FCT + AWT + WJT + RBS",
    `Short text` == "BSA" & `Classification` == "MAJOR" & `Modelo` == "EG-
720R" ~ "BSA + FCT + AWT + RBS",
    `Short text` == "CAP" ~ "CAP",
    `Short text` == "BSA+FSB" & `Modelo` != "EG-720R" ~ "BSA + FCT + AWT +
WJT + RBS + FSB + SUCTUBE + AWT_FSB",
    `Short text` == "BSA + FSB" & `Modelo` == "EG-720R" ~ "BSA + FCT + AWT
+ RBS + FSB + SUCTUBE + AWT_FSB",
    `Short text` == "CAP+RBS + NOZ" ~ "CAP + RBS + NOZ",
    `Short text` == "CHA" & `Modelo` != "EG-720R" ~ "CHA + FCT + AWT + WJT
+ RBS",
    `Short text` == "CHA" & `Modelo` == "EG-720R" ~ "CHA + FCT + AWT +
RBS",
    `Short text` == "CHA + BSA" & `Modelo` != "EG-720R" ~ "CHA + BSA + FCT
+ AWT + WJT + RBS",
    `Short text` == "CHA + BSA" & `Modelo` == "EG-720R" ~ "CHA + BSA + FCT
+ AWT + RBS",
    `Short text` == "COMPLETE HANLDE" ~ "COMPLETE HANLDE",

```

```

    `Short text` == "DWA" ~ "DWA",
    `Short text` == "FCT" & `Modelo` != "EG-720R" ~ "FCT + AWT + WJT +
RBS",
    `Short text` == "FCT" & `Modelo` == "EG-720R" ~ "FCT + AWT + RBS",
    `Short text` == "FCT + FSB" & `Modelo` != "EG-720R" ~ "FCT + AWT + WJT
+ RBS + FSB + SUCTUBE + AWT_FSB",
    `Short text` == "FCT + FSB" & `Modelo` == "EG-720R" ~ "FCT + AWT + RBS
+ FSB + SUCTUBE + AWT_FSB",
    `Short text` == "FSA" & `Modelo` != "EG-720R" ~ "FSA + FCT + AWT + WJT
+ RBS",
    `Short text` == "FSA" & `Modelo` == "EG-720R" ~ "FSA + FCT + AWT +
RBS",
    `Short text` == "FSA+BSA" & `Modelo` != "EG-720R" ~ "FSA + BSA + FCT +
AWT + WJT + RBS",
    `Short text` == "FSA+BSA" & `Modelo` == "EG-720R" ~ "FSA + BSA + FCT +
AWT + RBS",
    `Short text` == "FSA+FSB" & `Modelo` != "EG-720R" ~ "FSA + FSB + FCT +
AWT + WJT + RBS + SUCTUBE + AWT_FSB",
    `Short text` == "FSA+FSB" & `Modelo` == "EG-720R" ~ "FSA + FSB + FCT +
AWT + RBS + SUCTUBE + AWT_FSB",
    `Short text` == "FSB" ~ "FSB + SUCTUBE + AWT_FSB",
    `Short text` == "FSB + RBS" ~ "FSB + RBS + SUCTUBE + AWT_FSB",
    `Short text` == "ISA + FSB" ~ "ISA + FSB + SUCTUBE + AWT_FSB",
    `Short text` == "LGB" & `Modelo` != "EG-720R" ~ "LGB + CAP + NOZ + RBS
+ FCT + AWT + WJT",
    `Short text` == "LGB" & `Modelo` == "EG-720R" ~ "LGB + CAP + NOZ + RBS
+ FCT + AWT",
    `Short text` == "Push-Pull AWJ" ~ "AWT",
    `Short text` == "Push-Pull FCT" ~ "FCT",
    `Short text` == "RBS" ~ "RBS",
    `Short text` == "Suction cage" ~ "Suction cage",
    `Short text` == "FSB+RBS" ~ "FSB + RBS + SUCTUBE + AWT_FSB",
    `Short text` == "CHA + FSA" & `Modelo` != "EG-720R" ~ "CHA + FSA + FCT
+ AWT + WJT + RBS",
    `Short text` == "Replacing all channels" & `Modelo` == "EG-720R" ~ "FCT
+ AWT + RBS",
    `Short text` == "Replacing all channels" & `Modelo` != "EG-720R" ~ "FCT
+ AWT + WJT + RBS",
    `Short text` == "Inspection" ~ "Sem peças",
    `Short text` == "Small repair" ~ "Sem peças",
    `Short text` == "ISA+FSB" ~ "ISA + FSB + SUCTUBE + AWT_FSB",
  ))

```

```

# Verificar valores ausentes
colSums(is.na(Data))

```

```

# Eliminar as linhas com valores ausentes
Data <- na.omit(Data)

```

```

summary(Data)
describe(Data)
length(unique(Data$Cliente))
length(unique(Data$`Serial Number`))
length(unique(Data$Modelo))

```

```

equipamentos_por_modelo <- Data %>%
  group_by(Modelo) %>%
  summarise(Equipamentos_Unicos = n_distinct(`Serial Number`)) %>%

```

```

    arrange(desc(Equipamentos_Unicos))

print(equipamentos_por_modelo)

# Contagem de reparações por modelo
contagem_reparacoes <- table(Data$Modelo)
contagem_reparacoes <- sort(contagem_reparacoes, decreasing = TRUE)
contagem_reparacoes

# Contagem de tipos de manutenção no total
contagem_tipo_manutencao <- table(Data$Manutencao)
contagem_tipo_manutencao

# Identificar o modelo com mais reparações
modelo_mais_reparacoes <-
names(contagem_reparacoes)[which.max(contagem_reparacoes)]
modelo_mais_reparacoes

# Filtrar os dados para o modelo com mais reparações
dados_modelo <- Data[Data$Modelo == modelo_mais_reparacoes, ]

# Verificar contagens de tipos de manutenção e short text para o modelo com
mais reparações
contagem_tipo_manutencao_modelo1 <- table(dados_modelo$Manutencao)
contagem_short_text <- table(dados_modelo$`Short text`)
contagem_short_text <- sort(contagem_short_text, decreasing = TRUE)

# Converter o formato da variável Column1 para Date
Data$Column1 <- as.Date(Data$Column1, format="%d/%m/%y")

# Criar a coluna Ano a partir da data de reparação
Data$Ano <- format(Data$Column1, "%Y")

# Calcular o tempo entre reparações (em dias) para cada Serial Number
Data <- Data %>%
  arrange(`Serial Number`, Column1) %>% # ordena os dados de forma
  crescente, primeiro por nº de série e depois por data
  group_by(`Serial Number`) %>%
  mutate(Tempo_Entre_Reparacoes = c(0, diff(Column1))) #fiff(Column1)
  calcula a diferença entre datas consecutivas
                                                    # adiciona NA na
  primeira reparação

# Criar a coluna Numero Manutencao
Data <- Data %>%
  arrange(`Serial Number`, Column1) %>% # Ordena os dados por Serial
  Number e pela data de reparação
  group_by(`Serial Number`) %>%
  mutate(Numero_Manutencao = row_number()) %>% # Gera um número sequencial
  de manutenção para cada Serial Number
  ungroup()

# Calcular o tempo médio entre reparações por modelo
tempo_medio_modelo <- Data %>%
  group_by(Modelo) %>%
  summarise(Medio = round(mean(Tempo_Entre_Reparacoes, na.rm = TRUE))) %>%
  arrange(desc(Medio))

```

```

ggplot(tempo_medio_modelo, aes(x = reorder(Modelo, -Medio), y = Medio)) +
  geom_point(color = "blue") +
  geom_text(aes(label = round(Medio, 1)), vjust = -0.5, size = 3) +
  labs(x = "Modelo",
       y = "Tempo Médio (dias)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
ggsave("tempo_medio_entre_reparacoes_por_modelo.png", plot = last_plot(),
device = "png", dpi = 300)

# Calcular a quantidade de reparações por modelo
quantidade_reparacoes_modelo <- Data %>%
  group_by(Modelo) %>%
  summarise(Quantidade = n()) %>%
  arrange(desc(Quantidade))

ggplot(quantidade_reparacoes_modelo, aes(x = reorder(Modelo, -Quantidade),
y = Quantidade)) +
  geom_bar(stat = "identity", fill = "orange", color = "black") +
  geom_text(aes(label = Quantidade), vjust = -0.5) +
  theme_minimal() +
  labs(x = "Modelo",
       y = "Número de Reparções") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
ggsave("quantidade_reparacoes_modelo.png", plot = last_plot(), device =
"png", dpi = 300)

# Quantidade de inspeções/minor/major repair
tipo_manutencao_proporcao <- Data %>%
  group_by(`Priority text`) %>%
  summarise(Quantidade = n()) %>%
  mutate(Proporcao = Quantidade / sum(Quantidade))

ggplot(tipo_manutencao_proporcao, aes(x = "", y = Proporcao, fill =
`Priority text`)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar(theta = "y") +
  theme_void() +
  scale_fill_manual(values = c("lightpink", "lightblue", "orange")) +
  geom_text(aes(label = Quantidade), position = position_stack(vjust =
0.5), color = "white") +
  theme(plot.title = element_text(size = 14, face = "bold", hjust = 0.5))
ggsave("proporcao_tipos_manutencao.png", plot = last_plot(), device =
"png", dpi = 300)

#Boxplot do Tempo Entre Reparções por Modelo
ggplot(Data, aes(x = Modelo, y = Tempo_Entre_Reparacoes)) +
  geom_boxplot(fill = "lightgreen", color = "black") +
  labs(x = "Modelo",
       y = "Tempo Entre Reparções (dias)") +
  coord_cartesian(ylim = c(-200, 1600))+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
ggsave("boxplot_tempo_entre_reparacoes_por_modelo.png", plot = last_plot(),
device = "png", dpi = 300)

```

```

#Distribuição de Tipos de Manutenção por Modelo
tipo_manutencao_por_modelo <- Data %>%
  group_by(Modelo, `Priority text`) %>%
  summarise(Quantidade = n()) %>%
  group_by(Modelo) %>%
  mutate(Proporcao = Quantidade / sum(Quantidade)) %>%
  ungroup()

ggplot(tipo_manutencao_por_modelo, aes(x = reorder(Modelo, -Proporcao), y =
Proporcao, fill = `Priority text`)) +
  geom_bar(stat = "identity", position = "fill") +
  scale_y_continuous(labels = scales::percent) +      # eixo y em
percentuais
  labs(x = "Modelo",
       y = "Proporção (%)",
       fill = "Tipo de Manutenção") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_manual(values = c("lightpink", "lightblue", "orange"))
ggsave("proporcao_modelo_manutencao.png", plot = last_plot(), device =
"png", dpi = 300)

#Frequência de Reparações por Modelo e Ano - Heatmap
reparacoes_ano_modelo <- Data %>%
  count(Ano, Modelo, name = "Quantidade") %>%
  complete(Ano, Modelo, fill = list(Quantidade = 0))

ggplot(reparacoes_ano_modelo, aes(x = Ano, y = Modelo, fill = Quantidade))
+
  geom_tile(color = "black") +
  scale_fill_gradient(low = "white", high = "darkred", name = "Reparações")
+
  labs(x = "Ano",
       y = "Modelo",
       title = "Reparações por Ano e Modelo") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

ggsave("heatmap_reparacoes_ano_modelo.png", plot = last_plot(), device =
"png", dpi = 300)

# Quantidade de reparações por ano
quantidade_reparacoes_ano <- Data %>%
  group_by(Ano) %>%
  summarise(Quantidade = n()) %>%
  arrange(Ano)

ggplot(quantidade_reparacoes_ano, aes(x = Ano, y = Quantidade)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "red", size = 2) +
  labs(x = "Ano", y = "Quantidade de Reparações") +
  theme_minimal()
ggsave("quantidade_reparacoes_ano.png", plot = last_plot(), device = "png",
dpi = 300)

#Frequência de reparação por ano e tipo de manutenção
reparacoes_ano_tipo <- Data %>%

```

```

group_by(Ano, `Priority text`) %>%
  summarise(Quantidade = n())

ggplot(reparacoes_ano_tipo, aes(x = Ano, y = Quantidade, fill = `Priority
text`)) +
  geom_bar(stat = "identity") +
  labs(x = "Ano", y = "Quantidade de Reparções", fill = "Tipo de
Manutenção") +
  theme_minimal()
ggsave("reparacoes_ano_tipo_manutencao.png", plot = last_plot(), device =
"png", dpi = 300)

# ----- ANÁLISE ESTATÍSTICA -----

# Análise para identificar as relações entre as variáveis e o Tempo entre
reparações
# MÉTODOS ANOVA

anova_fit <- aov(Tempo_Entre_Reparacoes ~ ., data = Data)
summary(anova_fit)

anova_fit <- aov(Manutencao ~ ., data = Data)
summary(anova_fit)

# teste Kruskal-Wallis - teste não paramétrico
# Variáveis independentes para o teste
variaveis <- c("Modelo", "Priority text",
              "Short text", "Column1", "Classification", "Ano")

# variáveis dependentes
variaveis_dependentes <- c("Tempo_Entre_Reparacoes", "Manutencao", "Pecas")

for (dep_var in variaveis_dependentes) {
  cat("\nTeste de Kruskal-Wallis para a variável dependente:", dep_var,
      "\n")

  for (indep_var in variaveis) {
    formula <- as.formula(paste(dep_var, "~ `", indep_var, "`", sep = ""))
    resultado <- kruskal.test(formula, data = Data)

    cat("\nVariável independente:", indep_var, "\n")
    print(resultado)
  }
}

# ----- FEATURE SELECTION -----
# ----- Mutual Information -----

calculate_mutual_info <- function(Data, target_var) {

  # Criar um dataframe vazio para armazenar os resultados

```

```

mutual_info <- data.frame(Variable = character(), Mutual_Info =
numeric())

# Loop para iterar sobre todas as colunas do dataframe
for (var in colnames(Data[1:12])) {

  # Calcular a informação mútua entre a variável-alvo e cada coluna
  mi = mutinformation(Data[[target_var]], Data[[var]])
  mutual_info <- rbind(mutual_info, data.frame(Variable = var,
Mutual_Info = mi))
}
print(mutual_info)
return(mutual_info)
}

#target_var = Manutencao
mutual_info_manutencao <- calculate_mutual_info(Data, "Manutencao")

ggplot(mutual_info_results, aes(x=Mutual_Info)) +
  geom_histogram(binwidth=0.1, fill="pink", color="pink3") +
  labs(x="Mutual Information", y="Count")

#target_var = Pecas
mutual_info_pecas <- calculate_mutual_info(Data, "Pecas")

ggplot(mutual_info_results, aes(x=Mutual_Info)) +
  geom_histogram(binwidth=0.1, fill="pink", color="pink3") +
  labs(x="Mutual Information", y="Count")

#target_var = Tempo_Entre_Reparacoes
mutual_info_tempo <- calculate_mutual_info(Data, "Tempo_Entre_Reparacoes")

ggplot(mutual_info_results, aes(x=Mutual_Info)) +
  geom_histogram(binwidth=0.1, fill="pink", color="pink3") +
  labs(x="Mutual Information", y="Count")

# ----- NORMALIZAÇÃO -----
-
#Min-max normalization: Guarantees all features will have the exact same
scale but does not handle outliers well.
#Z-score normalization: Handles outliers, but does not produce normalized
data with the exact same scale.

# Normalizar apenas as variáveis numéricas
# Alterar o tipo de variável para numérico
Data$Cliente <- as.numeric(Data$Cliente)
Data$`Serial Number` <- as.numeric(Data$`Serial Number`)
Data$Manutencao <- as.numeric(Data$Manutencao)
Data$Ano <- as.numeric(Data$Ano)
Data$Tempo_Entre_Reparacoes <- as.numeric(Data$Tempo_Entre_Reparacoes)

# Lista das colunas numéricas que serão normalizadas
numeric_attrib <- c('Cliente', 'Serial Number', 'Manutencao', 'Ano',
'Tempo_Entre_Reparacoes')

normalize_and_scale <- function(Data, numeric_attrib) {
  # Função Z-Score
  z_score_function <- function(x) {

```

```

    (x - mean(x, na.rm = TRUE)) / (sd(x, na.rm = TRUE)) # Subtrai a média e
divide pelo desvio padrão para cada valor
  }

  # Z-score
  # Aplicar a função z_score_function (Z-Score) às colunas (valor 2)
numéricas (numeric_attrib)
  z_score_data <- Data
  z_score_data[, numeric_attrib] <- apply(Data[, numeric_attrib], 2,
z_score_function)

  #Min-Max
  # Calcular os valores mínimos e máximos de cada coluna numérica
  min_values <- apply(Data[, numeric_attrib], 2, min)
  max_values <- apply(Data[, numeric_attrib], 2, max)
  min_max_scaled_data <- Data
  min_max_scaled_data[, numeric_attrib] <- scale(Data[, numeric_attrib],
center = min_values, scale = max_values - min_values)

  return(list(z_score_data, min_max_scaled_data))
}

# Normalizar os dados
result <- normalize_and_scale(Data, numeric_attrib)
z_score_data <- result[[1]] # Dados normalizados pelo Z-Score
min_max_scaled_data <- result[[2]] # min max

# Layout para os gráficos: 2 linhas e 3 colunas
par(mfrow = c(2, 3))

# Histogramas para Z-score
for (col in numeric_attrib) {
  hist(z_score_data[[col]],
      main = paste("Histograma Z-Score:", col),
      xlab = paste("Valores de", col),
      col = "lightblue",
      breaks = 20,
      border = "darkblue")
}

# Histogramas para Min-Max
for (col in numeric_attrib) {
  hist(min_max_scaled_data[[col]],
      main = paste("Histograma Min-Max:", col),
      xlab = paste("Valores de", col),
      col = "lightgreen",
      breaks = 20,
      border = "darkgreen")
}

# Boxplot comparativo para 'Tempo_Entre_Reparacoes'
boxplot(z_score_data$Tempo_Entre_Reparacoes,
min_max_scaled_data$Tempo_Entre_Reparacoes,
      names = c("Z-Score", "Min-Max"),
      main = "Boxplot: Tempo Entre Reparações",
      col = c("lightblue", "lightgreen"))

# Boxplot comparativo para 'Manutencao'
```

```

boxplot(z_score_data$Manutencao, min_max_scaled_data$Manutencao,
        names = c("Z-Score", "Min-Max"),
        main = "Boxplot: Manutenção",
        col = c("lightblue", "lightgreen"))

Data_new <- Data[, c('Cliente', 'Serial Number', 'Modelo', 'Priority text',
                    'Short text',
                    'Classification', 'Pecas', 'Ano',
                    'Tempo_Entre_Reparacoes', 'Manutencao', 'Numero_Manutencao', 'Column1')]

# ----- ANÁLISE ESTATÍSTICA -----

# Prever a quantidade de reparações por ano, de 2025 a 2030
future_years <- 2025:2030
future_data <- data.frame(Ano = future_years)

# Criar a coluna Numero_Manutencao para os anos futuros,
# iniciando um número sequencial a partir do maior número já existente + 1
future_data$Numero_Manutencao <- seq(from = max(Data_new$Numero_Manutencao)
+ 1,
                                     by = 1, length.out =
length(future_data$Ano))

# Modelo de Regressão Linear
# y=ax+b
quantidade_reparacoes_ano$Ano <-
as.numeric(as.character(quantidade_reparacoes_ano$Ano))
lm_model <- lm(Quantidade ~ Ano, data = quantidade_reparacoes_ano)
summary(lm_model)

# Previsões com o modelo linear
predictions_lm <- predict(lm_model, newdata = future_data)
predictions_lm_rounded <- round(predictions_lm)
future_predictions_lm <- data.frame(Ano = future_years,
Predicted_Quantidade = predictions_lm_rounded)

# Modelo Polinomial de grau 2 - Prever a quantidade de reparações por ano,
de 2025 a 2030
poly_model <- lm(Quantidade ~ poly(Ano, 2), data =
quantidade_reparacoes_ano)
summary(poly_model)

# Previsões com o modelo polinomial
predictions_poly <- predict(poly_model, newdata = future_data)
predictions_poly_rounded <- round(predictions_poly)
future_predictions_poly <- data.frame(Ano = future_years,
Predicted_Quantidade = predictions_poly_rounded)

# Comparar os resultados obtidos pelos dois modelos, graficamente
ggplot(quantidade_reparacoes_ano, aes(x = Ano, y = Quantidade)) +
  geom_point(color = "red", size = 2) +
  geom_line(aes(y = predict(lm_model)), color = "blue", size = 1) +

```

```

    geom_line(data = future_predictions_lm, aes(x = Ano, y =
Predicted_Quantidade, color = "Modelo Linear"), size = 1, linetype =
"dashed") +
    geom_line(data = future_predictions_poly, aes(x = Ano, y =
Predicted_Quantidade, color = "Modelo Polinomial"), size = 1, linetype =
"dashed") +
    geom_point(data = future_predictions_lm, aes(x = Ano, y =
Predicted_Quantidade, color = "Modelo Linear"), size = 3) +
    geom_point(data = future_predictions_poly, aes(x = Ano, y =
Predicted_Quantidade, color = "Modelo Polinomial"), size = 3) +
    labs(x = "Ano", y = "Quantidade de Reparacões") +
    scale_color_manual(values = c("Modelo Linear" = "green", "Modelo
Polinomial" = "orange"), name = "Modelo") +
    scale_x_continuous(breaks = 2020:2030, limits = c(2020, 2030)) +
    theme_minimal()
ggsave("comparacao_modelos.png", plot = last_plot(), device = "png", dpi =
300)

```

```

print(future_predictions_lm)
print(future_predictions_poly)

```

```

# ----- PREVER O TEMPO ENTRE REPARAÇÕES EM DIAS PARA CADA
EQUIPAMENTO -----
# Usar dados normalizados (z_score_data), transformar Column1 em data,
ordenar e numerar manutenções por equipamento
Data_new <- z_score_data %>%
  mutate(Column1 = as.Date(Column1)) %>%
  arrange(`Serial Number`, Column1) %>% # Ordena pelo número de série e
data
  group_by(`Serial Number`) %>% # Agrupa por equipamento Serial Number
  mutate(Numero_Manutencao = row_number()) %>% # Gera número sequencial da
manutenção para cada equipamento
  ungroup()

# Filtragem dos dados para intervalos de tempo entre manutenções > 0
df_model <- Data_new %>% filter(Tempo_Entre_Reparacoes > 0)

df_model <- df_model %>%
  rename(
    Short_text = `Short text`,
    Priority_text = `Priority text`,
  )

# Modelo regressão linear
lm_model_numero <- lm(Tempo_Entre_Reparacoes ~ Column1 + Numero_Manutencao
+ Ano + Modelo + Classification, data = df_model)

summary(lm_model_numero)

```

```

set.seed(123)
trainIndex <- createDataPartition(df_model$Tempo_Entre_Reparacoes, p = 0.8,
list = FALSE)
train_data <- df_model[trainIndex, ]
test_data <- df_model[-trainIndex, ]
cv_model <- train(Tempo_Entre_Reparacoes ~ Column1 + Numero_Manutencao +
Manutencao,

```

```

        data = train_data,
        method = "lm", # Regressão Linear
        trControl = trainControl(method = "cv", number = 15)) #
15-fold Cross-validation

print(cv_model)

# MODELO RANDOM FOREST
# Criar features temporais a partir da coluna de data (Column1)
df_model <- df_model %>%
  mutate(
    Ano = year(Column1), # Extrai o ano da data
    Mes = month(Column1), # Extrai o mês da data
    Dia_mes = day(Column1) # Extrai o dia data
  )

# Separar múltiplas peças em linhas separadas
df_novo <- df_model %>%
  mutate(Pecas = as.character(Pecas)) %>% # Garante que a coluna Pecas seja
do tipo caracter
  separate_rows(Pecas, sep = " \\+ ") # Separa as peças que estão
separadas pelo " + " em linhas distintas

# Criar colunas binárias (multi-hot) para cada peça
df_multi_hot <- df_novo %>%
  mutate(value = 1) %>% # Adiciona uma coluna 'value' com valor 1 para
indicar a presença de uma peça
  pivot_wider(names_from = Pecas, values_from = value, values_fill = 0) #
Cria colunas para cada peça, preenchendo com 0 ou 1

# Converter as colunas de peças para numericas
peca_cols <- setdiff(colnames(df_multi_hot), colnames(df_model))
df_multi_hot[peca_cols] <- lapply(df_multi_hot[peca_cols], as.numeric) #
Converte as colunas de peças para numéricas

# Remover as variáveis do dataset
df_model_final <- dplyr::select(df_multi_hot, -`Serial Number`, -Cliente)

# QUAIS SÃO AS PEÇAS MAIS USADAS POR MODELO
pecas_por_modelo <- df_model_final %>%
  group_by(Modelo) %>%
  summarise(across(all_of(peca_cols), sum, na.rm = TRUE))

pecas_mais_usadas_por_modelo <- pecas_por_modelo %>%
  pivot_longer(cols = -Modelo, names_to = "Peca", values_to = "Frequencia")
%>%
  filter(Frequencia > 0) %>%
  arrange(Modelo, desc(Frequencia))

print(pecas_mais_usadas_por_modelo)

# remover espaços e caracteres inválidos
colnames(df_model_final) <- make.names(colnames(df_model_final))

df_model_final_rf <- df_model_final

```

```

#80% treino e 20% teste
set.seed(123)
trainIndex <- createDataPartition(df_model_final_rf$Tempo_Entre_Reparacoes,
p = 0.8, list = FALSE)
train_data <- df_model_final_rf[trainIndex, ]
test_data <- df_model_final_rf[-trainIndex, ]

# Grid de hiperparâmetros
rf_grid <- expand.grid(
  mtry = c(1, 3, 4),
  nodesize = c(22, 25, 32)
)

# randomForest
train_rf <- function(mtry, nodesize) {
  randomForest(
    Tempo_Entre_Reparacoes ~ ., # Variável alvo Tempo_Entre_Reparacoes
    data = train_data,
    ntree = 4000,
    mtry = mtry,
    nodesize = nodesize
  )
}

media_tempo <- mean(Data$Tempo_Entre_Reparacoes, na.rm = TRUE)
sd_tempo <- sd(Data$Tempo_Entre_Reparacoes, na.rm = TRUE)

results_rf <- data.frame(mtry = integer(), nodesize = integer(), RMSE_z =
numeric(), MAE_z = numeric(), RMSE_dias = numeric(), MAE_dias = numeric())

for(i in 1:nrow(rf_grid)) {
  cat("mtry =", rf_grid$mtry[i], "e nodesize =", rf_grid$nodesize[i], "\n")
  model <- train_rf(rf_grid$mtry[i], rf_grid$nodesize[i])
  pred <- predict(model, newdata = test_data)
  rmse_z <- sqrt(mean((pred - test_data$Tempo_Entre_Reparacoes)^2)) #RMSE
normalizado
  mae_z <- mean(abs(pred - test_data$Tempo_Entre_Reparacoes))
  pred_dias <- pred * sd_tempo + media_tempo
  real_dias <- test_data$Tempo_Entre_Reparacoes * sd_tempo + media_tempo
  rmse_dias <- sqrt(mean((pred_dias - real_dias)^2))
  mae_dias <- mean(abs(pred_dias - real_dias))
  results_rf <- rbind(
    results_rf,
    data.frame(
      mtry = rf_grid$mtry[i],
      nodesize = rf_grid$nodesize[i],
      RMSE_z = rmse_z,
      MAE_z = mae_z,
      RMSE_dias = rmse_dias,
      MAE_dias = mae_dias
    )
  )
}
print(results_rf)

mean_value <- mean(df_model_final_rf$Tempo_Entre_Reparacoes)

```

```

rmse_naive_z <- sqrt(mean((mean_value -
test_data$Tempo_Entre_Reparacoes)^2))
mae_naive_z <- mean(abs(mean_value - test_data$Tempo_Entre_Reparacoes))

mean_value_dias <- mean_value * sd_tempo + media_tempo
real_dias <- test_data$Tempo_Entre_Reparacoes * sd_tempo + media_tempo

rmse_naive_dias <- sqrt(mean((mean_value_dias - real_dias)^2))
mae_naive_dias <- mean(abs(mean_value_dias - real_dias))

cat("Baseline - RMSE z-score:", rmse_naive_z, ", MAE z-score:",
mae_naive_z, "\n")
cat("Baseline - RMSE dias:", rmse_naive_dias, ", MAE dias:",
mae_naive_dias, "\n")

# ----- PREVER O TIPO DE MANUTENÇÃO
# Garantir que a variável seja fator
df_model_final$Manutencao <- as.factor(df_model_final$Manutencao)

# Separar preditores e alvo
df_model_sem_target <- df_model_final %>% dplyr::select(-Manutencao, -
Classification, -Priority_text)
manutencao_target <- df_model_final$Manutencao

# Dividir os dados
set.seed(123)
trainIndex_manu <- createDataPartition(manutencao_target, p = 0.8, list =
FALSE)

trainData_manu <- df_model_sem_target[trainIndex_manu, ]
trainTarget_manu <- manutencao_target[trainIndex_manu]

testData_manu <- df_model_sem_target[-trainIndex_manu, ]
testTarget_manu <- manutencao_target[-trainIndex_manu]

# Grid de hiperparâmetros
rf_grid_manu <- expand.grid(mtry = c(15, 20, 22), nodesize = c(1, 2, 5))

# Função de treino
train_rf_manu <- function(mtry, nodesize) {
  randomForest(
    x = trainData_manu,
    y = trainTarget_manu,
    ntree = 2000,
    mtry = mtry,
    nodesize = nodesize
  )
}

results_rf_manu <- data.frame(
  mtry = integer(),
  nodesize = integer(),
  Accuracy = numeric(),
  Precision = numeric(),
  Recall = numeric(),
  F1 = numeric()
)

```

```

for (i in 1:nrow(rf_grid_manu)) {
  cat("mtry =", rf_grid_manu$mtry[i],
      "e nodesize =", rf_grid_manu$nodesize[i], "\n")

  model_manu <- train_rf_manu(
    rf_grid_manu$mtry[i],
    rf_grid_manu$nodesize[i]
  )

  pred_manu <- predict(model_manu, newdata = testData_manu)

  cm <- confusionMatrix(pred_manu, testTarget_manu)

  acc <- cm$overall["Accuracy"]
  precision <- mean(cm$byClass[, "Precision"], na.rm = TRUE)
  recall <- mean(cm$byClass[, "Recall"], na.rm = TRUE)
  f1 <- mean(cm$byClass[, "F1"], na.rm = TRUE)

  results_rf_manu <- rbind(
    results_rf_manu,
    data.frame(
      mtry = rf_grid_manu$mtry[i],
      nodesize = rf_grid_manu$nodesize[i],
      Accuracy = round(acc, 4),
      Precision = round(precision, 4),
      Recall = round(recall, 4),
      F1 = round(f1, 4)
    )
  )
}

print(results_rf_manu)

# ----- PREVER O TIPO DE MANUTENÇÃO

library(neuralnet)
library(caret)
library(dplyr)
library(lubridate)

# dummies para variáveis categóricas
dummies <- dummyVars(~ Modelo + Short_text, data = df_model_final)
dummy_data <- predict(dummies, newdata = df_model_final)

df_model_final_num <- cbind(
  dplyr::select(df_model_final, -Modelo, -Priority_text, -Short_text, -
Classification),
  dummy_data
)

colnames(df_model_final_num) <- make.names(colnames(df_model_final_num))
df_model_final_num[] <- lapply(df_model_final_num, as.numeric)

# Separar variável-alvo
df_model_final_num$Manutencao_target <- df_model_final_num$Manutencao

# Divisão treino/teste
set.seed(123)

```

```

trainIndex <- createDataPartition(df_model_final_num$Manutencao_target, p =
0.8, list = FALSE)
train_data <- df_model_final_num[trainIndex, ]
test_data <- df_model_final_num[-trainIndex, ]

predictors <- setdiff(names(df_model_final_num), c("Manutencao_target",
"Manutencao"))
formula_nn <- as.formula(paste("Manutencao_target ~", paste(predictors,
collapse = " + ")))

model_list <- lapply(1:5, function(i) {
  cat("Camada hidden = c(5,", i, ")\n")
  neuralnet(
    formula = formula_nn, # Variável alvo Manutencao
    data = train_data,
    hidden = c(5, i),
    linear.output = TRUE,
    stepmax = 1e9,
    lifesign = "full",
    algorithm = "rprop+",
    threshold = 0.01,
    rep = 1
  )
})

valid_results <- lapply(model_list, function(model) {
  pred <- compute(model, test_data[, predictors])$net.result[,1]
  rmse <- sqrt(mean((pred - test_data$Manutencao_target)^2))
  return(rmse)
})

valid_results_df <- data.frame(
  Model = 1:5,
  Hidden_Layers = paste("(", 5, ",", 1:5, ")", sep = ""),
  RMSE = unlist(valid_results)
)
print(valid_results_df)

train_results <- lapply(model_list, function(model) {
  pred <- compute(model, train_data[, predictors])$net.result[,1]
  rmse <- sqrt(mean((pred - train_data$Manutencao_target)^2))
  return(rmse)
})

train_results_df <- data.frame(
  Model = 1:5,
  Hidden_Layers = paste("(", 5, ",", 1:5, ")", sep = ""),
  RMSE = unlist(train_results)
)
print(train_results_df)

best_model_index <- which.min(unlist(valid_results))
best_model <- model_list[[best_model_index]]

cat("Melhor modelo selecionado: camada hidden =",
valid_results_df$Hidden_Layers[best_model_index], "\n")

test_predictions <- compute(best_model, test_data[,
predictors])$net.result[,1]

```

```

final_rmse <- sqrt(mean((test_predictions -
test_data$Manutencao_target)^2))
cat("RMSE final no teste:", final_rmse, "\n")

# PREVER AS PEÇAS QUE SERÃO USADAS NA PRÓXIMA INTERVENÇÃO

dummies <- dummyVars(~ Modelo + Short_text, data = df_multi_hot)
dummy_data <- predict(dummies, newdata = df_multi_hot)

df_final <- cbind(
  dplyr::select(df_multi_hot, -Cliente, -`Serial Number`, -Modelo, -
Priority_text, -Short_text, -Classification), # Exclui as variáveis
originais
  dummy_data
)

colnames(df_final) <- make.names(colnames(df_final))
df_final[] <- lapply(df_final, as.numeric)

colnames(df_final) <- gsub(" ", "_", colnames(df_final))
colnames(df_final) <- make.names(colnames(df_final))
peca_cols <- setdiff(colnames(df_final), colnames(df_model))
predictors <- setdiff(names(df_final), peca_cols)
predictors <- setdiff(predictors, c("Numero_Manutencao", "Mes", "Dia_mes",
"Serial_Number", "Cliente"))
target_cols <- peca_cols

formula_nn <- as.formula(
  paste(paste(target_cols, collapse = " + "), "~", paste(predictors,
collapse = " + "))
)

set.seed(123)
trainIndex <- createDataPartition(df_final[[target_cols[1]]], p = 0.8, list
= FALSE)
train_data <- df_final[trainIndex, ]
test_data <- df_final[-trainIndex, ]

model_list <- lapply(1:5, function(i) {
  neuralnet(
    formula = formula_nn,
    data = train_data,
    hidden = c(5, 3),
    linear.output = FALSE,
    stepmax = 1e6,
    lifesign = "full",
    algorithm = "rprop+",
    threshold = 0.01
  )
})

rmse_multi_label <- function(pred_probs, real_values) {
  sqrt(mean((pred_probs - real_values)^2))
}

valid_results <- lapply(model_list, function(model) {

```

```

    pred <- compute(model, test_data[, predictors])$net.result
    real <- as.matrix(test_data[, target_cols])
    rmse <- rmse_multi_label(pred, real)
    return(rmse)
  })

valid_results_df <- data.frame(
  Model = 1:5,
  Hidden_Layers = paste("(", 5, ",", 1:5, ")", sep = ""),
  RMSE = unlist(valid_results)
)
print(valid_results_df)

train_results <- lapply(model_list, function(model) {
  pred <- compute(model, train_data[, predictors])$net.result
  real <- as.matrix(train_data[, target_cols])
  rmse <- rmse_multi_label(pred, real)
  return(rmse)
})

train_results_df <- data.frame(
  Model = 1:5,
  Hidden_Layers = paste("(", 5, ",", 1:5, ")", sep = ""),
  RMSE = unlist(train_results)
)
print(train_results_df)

best_model_index <- which.min(valid_results_df$RMSE)
best_model <- model_list[[best_model_index]]

test_predictions <- compute(best_model, test_data[, predictors])$net.result
final_rmse <- rmse_multi_label(test_predictions, as.matrix(test_data[,
target_cols]))

cat("RMSE final no teste:", final_rmse, "\n")

# ----- MODELO COM KERAS -----

library(keras)

x_train <- as.matrix(train_data[, predictors])
y_train <- as.matrix(train_data[, target_cols])
x_test <- as.matrix(test_data[, predictors])
y_test <- as.matrix(test_data[, target_cols])

model_keras <- keras_model_sequential() %>%
  layer_dense(units = 128, activation = "relu", input_shape =
ncol(x_train)) %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 32, activation = "relu") %>%
  layer_dense(units = length(target_cols), activation = "sigmoid") # Saída
binária para cada peça

model_keras %>% compile(
  loss = "binary_crossentropy", # Função de perda binária para multi-label
  optimizer = optimizer_adam(learning_rate = 0.001),
  metrics = c("accuracy", "AUC")
)

```

```

early_stopping <- callback_early_stopping(monitor = "val_loss", patience =
15, restore_best_weights = TRUE)

history <- model_keras %>% fit(
  x_train, y_train,
  epochs = 200,
  batch_size = 64,
  validation_split = 0.2,
  callbacks = list(early_stopping)
)

score_keras <- model_keras %>% evaluate(x_test, y_test)
cat("Loss (Keras):", score_keras[1], " Accuracy (Keras):", score_keras[2],
"\n")

predictions_keras <- model_keras %>% predict(x_test)
predictions_bin_keras <- ifelse(predictions_keras > 0.35, 1, 0)

rmse_keras <- sqrt(mean((predictions_bin_keras - y_test)^2))
cat("RMSE no conjunto de teste (Keras):", rmse_keras, "\n")

```