





# Análise de marcadores biométricos e sinais vitais através de visão artificial

# Ritmo Cardíaco

Projeto de Licenciatura em Engenharia Biomédica

Ano letivo 2024/2025

# Rodrigo Rosado Gonçalves

Orientador: Professor Doutor Nuno Gonçalves

# Agradecimentos

Gostaria de expressar o meu profundo agradecimento ao Professor Doutor Nuno Gonçalves, pela orientação e pela disponibilidade constante ao longo de todas as fases deste trabalho. O seu rigor científico e apoio foram fundamentais para a concretização deste projeto.

Agradeço também aos investigadores Gustavo Bongiovi e Diogo Nunes. Ao Gustavo, por ter intermediado todos os requerimentos dos datasets que serviram de base ao trabalho. E ao Diogo, pela valiosa ajuda na adaptação dos códigos. A ambos, deixo o meu reconhecimento pela simpatia, partilha de conhecimento e disponibilidade.

Um forte agradecimento também a todos os participantes no dataset privado, em especial aos que colaboraram com a realização da corrida.

Por fim, um agradecimento muito especial à minha colega de projeto Margarida Ramos, pelo apoio contínuo e pelas trocas de ideias ao longo deste percurso.

#### Resumo

A monitorização de sinais vitais, nomeadamente da frequência cardíaca, é essencial em contexto clínico, especialmente no período pós-cirúrgico. A dependência de equipamentos médicos tradicionais e de acompanhamento presencial pode trazer desafios logísticos e riscos acrescidos, tanto para o paciente como para os serviços de saúde. Este projeto teve como objetivo estudar o potencial da fotopletismografia remota (rPPG), uma técnica não invasiva, que não requer contacto físico e que, através de visão artificial, permite estimar biossinais – em particular, componentes do ritmo cardíaco – recorrendo apenas a câmaras RGB comuns.

Com recurso à rPPG-Toolbox, foram testados vários modelos supervisionados e não supervisionados em diferentes conjuntos de dados – PURE, UBFC-rPPG e COHFACE –, tendo-se avaliado o seu desempenho na estimativa da frequência cardíaca com métricas como MAE, RMSE, MAPE e correlação de Pearson. De forma paralela, foi desenvolvido um pequeno dataset privado com vídeos de baixa qualidade, simulando condições mais próximas da realidade da maioria da população.

Os resultados demonstraram que os modelos supervisionados apresentam desempenho elevado em testes intra-dataset. Em contrapartida, a generalização para outros dados (inter-dataset) continua a ser um desafio, especialmente em datasets mais ruidosos como o COHFACE. Já entre os modelos não supervisionados, apesar de resultados ligeiramente inferiores, o método POS (cuja base matemática assenta num plano ortogonal ao tom de pele) destacou-se pela sua robustez.

Este estudo destaca o potencial da rPPG como ferramenta de monitorização remota de biossinais e explora a sua eventual utilização em contexto médico.

**Palavras Chave:** Fotopletismografia Remota, Frequência Cardíaca, Modelos Supervisionados, Modelos Não Supervisionados, Monitorização Remota

#### Lista de Abreviaturas

- AV (Nódulo) Auriculoventricular
- BVP Pressão do Volume Sanguíneo (Blood Volume Pulse)
- **bpm** Batimentos por Minuto (*beats per minute*)
- CPU Unidade Central de Processamento (Central Processing Unit)
- **CUDA** Arquitetura Unificada de Dispositivos de Computação (*Compute Unified Device Architecture*)
- ECG Eletrocardiograma (Electrocardiograma)
- **EVM** Amplificação de Vídeo Euleriana (*Eulerian Video Magnification*)
- FC Frequência Cardíaca
- FFT Transformada rápida de Fourier (Fast Fourier Transform)
- **fps** Frames por Segundo (*frames per second*)
- **GPU** Unidade de Processamento Gráfico (*Graphics Processing Unit*)
- **IA** Inteligência Artificial
- IC Intervalo de Confiança
- ML Aprendizagem Automática/de Máquina (Machine Learning)
- NIR Próximo do Infravermelho (Near Infrared)
- **PPG** Fotopletismografia (*Photoplethysmography*)
- **RGB** Vermelho, Verde e Azul (*Red*, *Green*, *Blue*)
- RR Ritmo Respiratório
- **ROI** Região de Interesse (*Region of Interest*)
- rPPG Fotopletismografia Remota (remote Photoplethysmography)
- **SA** (Nódulo) Sinoauricular
- **VS** Versus (contra, em oposição)

# Lista de Figuras

1	Representação esquemática simples de rPPG [10]. No caso do PPG, ao	
	invés de uma câmara temos um fotodetetor em contacto com a pele	7
2	Representação esquemática simplificada do funcionamento global de um	
	algoritmo de rPPG	8
3	<i>Input</i> que entra nos modelos a partir dos frames dos vídeos	14
4	Labels do sinal PPG e estimativa da FC média por chunk	14
5	Gráfico dos sinais PPG real e estimado num <i>chunk</i>	15
6	Scatter Plot do modelo DeepPhys no dataset PURE	18
7	Difference Plot do modelo Deephys no dataset PURE	18
8	Scatter Plot do modelo Deephys treinado PURE e testado no UBFC	20
9	Difference Plot do modelo Deephys treinado PURE e testado no UBFC	20
10	Scatter Plot do modelo EfficientPhys treinado UBFC e testado no COH-	
	FACE	21
11	Difference Plot do modelo EfficientPhys treinado UBFC e testado no	
	COHFACE	21
12	Scatter Plot do modelo CHROM no dataset UBFC	24
13	Difference Plot do modelo CHROM no dataset UBFC	24
14	Representação da utilização da banda torácica <i>Polar H10</i>	25
15	Fotografia da utilização do oxímetro Entigo OLED TY05, onde entre os	
	valores de Sp $0_2$ e FC, está RR/min	25

# Lista de Tabelas

1	Resultados das métricas de avaliação nos modelos supervisionados	16
2	Resultados das métricas de avaliação nos modelos não supervisionados	22

# Índice

1	Intr	rodução	1				
2	Ritr	no Cardíaco	3				
	2.1	Ritmo Cardíaco VS Frequência Cardíaca	3				
	2.2	Métodos para medir Frequência Cardíaca	4				
	2.3	Importância da Monitorização da Frequência Cardíaca	5				
	2.4	Fatores que Alteram Frequência Cardíaca	6				
3	Foto	opletismografia	7				
	3.1	Técnica de PPG	7				
	3.2	Técnica de rPPG	8				
4	Mét	todos da rPPG-Toolbox	9				
	4.1	Modelos Supervisionados	9				
	4.2	Modelos Não Supervisionados	10				
	4.3	Datasets	12				
	4.4	Ferramentas para Análise dos Resultados	13				
		4.4.1 Métricas de avaliação teste	13				
		4.4.2 Visualização Pré-Processamento	14				
		4.4.3 Visualização Pós-Processamento	15				
5	Res	ultados e Discussão	16				
	5.1	Modelos Supervisionados	16				
		5.1.1 Testes Intra-Dataset	17				
		5.1.2 Testes Inter-Dataset	19				
	5.2	Modelos Não Supervisionados	22				
6	Cria	ação de um Dataset Privado	<b>2</b> 5				
7	Limitações 27						
8	Con	ıclusão	28				
Re	eferê	ncias Bibliográficas	29				

# 1 Introdução

O acompanhamento realizado em contexto pós-cirúrgico é fundamental para avaliar de forma contínua e detalhada o estado de saúde do paciente durante o seu processo de recuperação. Atualmente, este acompanhamento é efetuado maioritariamente de forma presencial nas unidades de saúde, o que pode originar constrangimentos tanto para os pacientes, como para os profissionais e os próprios serviços de saúde. No que concerne aos pacientes, existem riscos associados à deslocação, desde o percurso até ao hospital (que pode constituir um desafio atendendo a fatores como a idade, o tipo de cirurgia, o estado de debilidade e à distância de residência), até à possível exposição a infeções hospitalares. Relativamente aos serviços de saúde, em períodos de maior stress hospitalar e sobrelotação de unidades, a presença de pacientes considerados não urgentes pode contribuir para o desgaste dos profissionais, o consumo acrescido de recursos e o adiamento no atendimento a doentes mais prioritários.

Com a evolução tecnológica dos últimos anos, tem-se assistido a uma modernização e atualização das metodologias tradicionais em diversos setores, em particular na área da medicina. No âmbito médico, uma área que tem vindo a crescer é a ideia de um acompanhamento remoto via tele-consultas de doentes não urgentes. Este tipo de acompanhamento, que registou um pico durante a pandemia de COVID-19, demonstrou ser uma alternativa fiável e eficaz [1]. O progresso na conceção de componentes eletrónicos e computacionais cada vez mais compactos e potentes, como CPUs e GPUs, tem permitido que computadores acessíveis e utilizados pela grande maioria da população, disponham de maior capacidade de processamento de dados. Paralelamente, tem-se verificado a integração de sistemas de aquisição de imagens com cada vez mais qualidade e resolução, como as câmaras RGB incorporadas nos computadores. Este desenvolvimento rápido de hardware aliado à acelerada "explosão"da IA, permite a criação de softwares cada vez mais sofisticados.

Nesse sentido, uma tecnologia que tem vindo a afirmar-se como promissora e eficaz é a medição de biossinais através de visão computacional, uma alternativa *contactless*, não intrusiva e que, do ponto de vista do paciente, é de fácil utilização. Uma das técnicas utilizadas para tal é a rPPG, que permite, através de processos óticos, recolher com precisão o sinal PPG (análogo ao obtido por oxímetros) e estimar indicadores como FC à distância.

Assim, o objetivo deste projeto consiste em testar a implementação de metodologias baseadas em técnicas de rPPG que, a partir de dados recolhidos por câmaras RGB simples (como as integradas nos computadores pessoais), possam extrair informações acerca de vários sinais vitais do paciente. No caso deste trabalho em específico, dados do ritmo cardíaco do doente. Através da avaliação destas metodologias pretende-se também compreender em que medida será possível colmatar as lacunas provenientes do acompanhamento presencial, combinando-o com um acompanhamento médico virtual, online e à distância.

#### 2 Ritmo Cardíaco

Antes de abordar as metodologias, é fundamental compreender a definição de ritmo cardíaco e perceber como é que, do ponto de vista fisiológico, ocorre o "bater"do coração, quais as reações envolvidas e os diversos fatores que o alteram. Além disso, deve ser reconhecida também a importância da sua monitorização em contexto médico, bem como as diversas estratégias para a sua medição ao longo do tempo.

#### 2.1 Ritmo Cardíaco VS Frequência Cardíaca

Apesar de, frequentemente, o ritmo cardíaco e a frequência cardíaca (FC) serem utilizados como sinónimos, são conceitos distintos. O ritmo cardíaco não é só o número de vezes com que o coração "bate" durante os sucessivos ciclos cardíacos, tem também em conta aspetos relacionados à regularidade desses batimentos, à sua origem elétrica e ao padrão de condução do impulso ao longo de todo o miocárdio.

Podemos começar por definir a FC como o número de ciclos completos de contração (sístole) e relaxamento (diástole) do músculo cardíaco por unidade de tempo (geralmente por minuto). No entanto, o coração humano é um órgão complexo, constituído por quatro cavidades (duas aurículas e dois ventrículos) e quatro válvulas que permitem ou impedem a passagem de sangue de forma sequencial e não simultânea. Cada batimento cardíaco é fruto de um conjunto sequencial de reações fisiológicas de sístoles e diástoles das várias cavidades cardíacas.

As oscilações cíclicas de potencial elétrico que originam o ritmo cardíaco são iniciadas pelas células pacemaker do nódulo sino auricular (SA), localizado na parte superior da aurícula direita do coração. Estas células geram o potencial elétrico responsável pela contração das fibras do miocárdio. Depois, este impulso elétrico é propagado a nível auricular pelas vias internodais até ao nódulo auriculoventricular (AV) e a nível ventricular pelas células de Purkinje. No nódulo AV o impulso sofre uma retardação, o que permite que a sístole auricular ocorra antes da ventricular.

O ritmo cardíaco pode ser caracterizado consoante alguns parâmetros, sendo isso fundamental para avaliar o estado de saúde cardiovascular do paciente.

- Frequência cardíaca (FC) A faixa de valores considerada como normal em repouso segue, habitualmente, a referência clínica convencional de 60 a 100 bpm, embora alguns estudos proponham outras faixas, como 50 a 90 bpm [2], entre outras. As variações refletem sempre as especificidades e o contexto particular abordado por cada estudo. De um modo geral, pacientes cuja FC de repouso é inferior à faixa convencional são classificados como braquicárdicos, enquanto que pacientes com valores superiores são classificados como taquicárdicos. A FC é o parâmetro do ritmo cardíaco mais facilmente mensurável dada a sua natureza numérica e objetiva.
- Variabilidade O ritmo cardíaco é considerado regular quando os intervalos de tempo entre batimentos se mantêm praticamente constantes, o que faz com que os valores da FC fiquem relativamente estáveis ao longo do tempo. Por outro lado, num ritmo cardíaco irregular/arrítmico, os batimentos possuem um padrão imprevisível, que resulta na variação repentina dos valores da FC. Uma reduzida variabilidade pode estar associada a um maior risco de doenças cardiovasculares, enquanto que uma elevada variabilidade poderá indicar a presença de arritmias.
- Origem e direção de propagação do sinal elétrico O ritmo cardíaco é considerado sinusal quando o vetor de propagação elétrico tem origem no nódulo SA, com sentido da zona superior direita para a inferior esquerda do coração. No caso de um ritmo não sinusal (normalmente associado ao sinal elétrico não surgir no SA, mas noutra região cardíaca) a direção do vetor propagação encontra-se comprometida, podendo resultar em disfunções cardíacas. Esta particularidade é detetável através de eletrocardiograma (ECG), visto que um batimento sinusal típico possui uma onda P (que representa a despolarização auricular), seguida de um complexo QRS (despolarização ventricular) e de uma onda T (repolarização ventricular). No caso do ritmo não sinusal, observam-se anomalias na morfologia da onda P no ECG, dada a alteração no vetor de propagação elétrico [3].

# 2.2 Métodos para medir Frequência Cardíaca

A FC pode ser medida através de métodos mais "tradicionais", como a contagem por palpação ou por auscultação. A palpação é geralmente realizada nas regiões anatómicas do punho (artéria radial) e do pescoço (artéria carótida), onde a dilatação do respetivo vaso sanguíneo (pulso) é mais facilmente detetável. Tal fenómeno deve-se à conjugação entre a elasticidade do vaso e o aumento do BVP causado pela ejeção de sangue da sístole ventricular. Nos métodos sonoros de auscultação, como a estetoscopia, recorrem-se a dispositivos médicos para auscultar o som periódico das válvulas cardíacas a fechar.

Atualmente existem também métodos elétricos e óticos, nomeadamente o ECG e a PPG, que, embora possam ser menos práticos do que os métodos tradicionais, oferecem elevada precisão. No caso do ECG, o eletrocardiógrafo regista as variações de potencial elétrico associadas às despolarizações das células pacemaker cardíacas, que são posteriormente representadas sob a forma de ondas. Por sua vez, a PPG baseia-se na deteção das variações da luz refletida nos vasos sanguíneos superficiais, provocadas pelas alterações na quantidade de hemoglobina no sangue. Essas alterações estão associadas ao BVP, uma vez que o aumento da pressão resulta de um maior volume sanguíneo, o qual, por sua vez, transporta uma maior concentração de hemoglobina.

#### 2.3 Importância da Monitorização da Frequência Cardíaca

A monitorização dos sinais vitais do paciente em contexto pós-cirúrgico, com especial destaque para o ritmo cardíaco, é de extrema importância. Diversos estudos revelam que é um parâmetro relevante, uma vez que se verificam alterações no ciclo circadiano, especialmente na resposta cardíaca, as quais podem constituir um indicador prognóstico de possíveis complicações futuras [4]. Adicionalmente, têm sido desenvolvidos modelos que, com base na variabilidade da FC, procuram prever o risco de complicações em contexto pós-cirúrgico [9].

Por outro lado, o estado de maior debilidade em que o doente se poderá encontrar no período após a cirurgia, constitui uma condição de risco acrescido. Não só a nível físico, uma vez que o paciente pode necessitar de um esforço adicional para realizar tarefas básicas do quotidiano, mas também a nível imunitário, dada a maior suscetibilidade à contração de alguma patologia. Para além disso, o controlo da FC pode constituir um indicador útil para avaliar a resposta do paciente não só à recuperação cirúrgica, mas também à medicação ou aos fármacos que lhe são administrados.

Ao nível cardiovascular, a ausência de uma monitorização eficaz pode expor o paciente a diversos riscos, tais como arritmias, taquicardia ou braquicardia, perfusão sanguínea inadequada, eventos isquémicos, insuficiência cardíaca, ou até mesmo paragem cardíaca, frequentemente associados ao estado de debilidade pós-cirurgia. A desregulação do ciclo circadiano, quando não devidamente controlada, pode comprometer a recuperação, tornando-a mais lenta e potencialmente mais perigosa.

#### 2.4 Fatores que Alteram Frequência Cardíaca

Nesta secção, serão abordados os principais mecanismos que podem provocar um aumento ou diminuição significativa da FC de um indivíduo, que é o parâmetro mais absoluto na caracterização do ritmo cardíaco. As outras duas fontes de caracterização abordadas em 2.2 estão normalmente associadas a patologias cardiovasculares.

O valor da FC de um indivíduo pode aumentar muito rapidamente caso este seja submetido à pratica de atividade física de intensidade moderada/intensa, uma vez que o organismo requer um maior transporte de nutrientes e oxigénio para os tecidos. Em contraste, técnicas de meditação ou respiração profunda tendem à sua diminuição, devido à diminuição dos níveis de cortisol e à redução do stress.

O estado psicológico e emocional do indivíduo também influencia a FC [6]. Algumas emoções intensas e situações de stress levam ao seu aumento, tal como a estimulação cognitiva via interação com jogos ou resolução de problemas matemáticos (como é realizado em alguns datasets [28]). De forma semelhante, pesquisas sobre a exposição a conteúdos violentos e/ou chocantes demonstraram também aumentos na FC, refletindo a ativação do sistema nervoso simpático associada a estados de medo, excitação ou alerta [8].

Para além destes, a FC é também influenciada por uma variedade de outros fatores. Segundo os autores do artigo [7], estes fatores podem ser agrupados em seis grandes categorias: fisiológicos (como a idade, o sexo e o ritmo circadiano); patológicos (como inflamações, infeções e doenças cardiovasculares); ambientais (como ruído, stress social e temperatura); relacionados com o estilo de vida (o consumo de álcool, tabaco, cafeína ou medicamentos, o nível de atividade física e de sedentarismo); psicológicos (como stress, ansiedade e depressão); e genéticos.

# 3 Fotopletismografia

Dentro da técnica de fotopletismografia distinguem-se dois grandes domínios: a PPG e a rPPG. Em ambos os casos, é utilizado o mesmo princípio ótico para a extração do sinal fisiológico. A principal diferença reside no facto de, na rPPG o foto-detetor, que neste caso é uma câmara, não estar em contacto direto com o indivíduo, o que permite a estimativa remota do sinal.

#### 3.1 Técnica de PPG

A PPG utiliza uma fonte de luz e um fotodetetor para medir as variações de volume sanguíneo nos vasos subcutâneos periféricos. Quando a luz incide na pele, o fotodetetor capta microvariações na luz refletida, correspondentes ao sinal PPG, provenientes das variações no volume do sangue. Os constituintes do sangue, em particular a hemoglobina, absorvem parte da luz incidente. De acordo com a lei de Lambert-Beer, quanto maior a concentração ou a espessura da coluna de sangue, maior será a absorvância (fração de luz incidente que é absorvida num material) da hemoglobina [5].

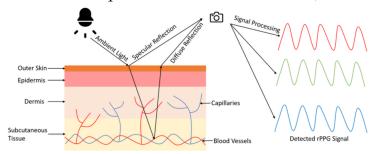


Figura 1: Representação esquemática simples de rPPG [10]. No caso do PPG, ao invés de uma câmara temos um fotodetetor em contacto com a pele.

Durante os batimentos cardíacos, a quantidade de hemoglobina varia. Quando há ejeção ventricular de sangue temos um pico de BVP, a quantidade de hemoglobina que transita pelos vasos aumenta, o que faz aumentar a absorvância e diminuir a intensidade da luz refletida que chega ao fotodetetor.

Os sistemas que traduzem o sinal PPG estão geralmente configurados para inverter o sinal captado. Assim, embora a intensidade da luz que atinge o fotodetetor diminua, essa redução é representada no sinal PPG como um pico positivo. Desta forma, o traçado do sinal PPG apresenta uma elevada semelhança com os sinais obtidos por ECG ou até mesmo com o gráfico de BVP ao longo de cada ciclo cardíaco.

Esta técnica é amplamente utilizada em contextos médicos e desportivos, por ser extremamente prática e não invasiva, seja em oxímetros de dedo ou de pulso, ou até mesmo em *smartwatches* [10].

#### 3.2 Técnica de rPPG

Podem ainda ser consideradas metodologias de rPPG, que se baseiam na extração remota do sinal de PPG, ou seja, à distância. Neste contexto, o sinal PPG é obtido a partir da análise de vídeos faciais ou, em tempo real, através de uma câmara. A análise recorre à avaliação dos sucessivos frames, bem como às micro variações na matriz RGB dos pixeis das ROIs faciais, onde a extração do sinal é mais eficaz. Estas regiões correspondem, geralmente, às bochechas e à testa, por apresentarem maior densidade de vasos sanguíneos superficiais [11]. Diversos métodos com distintas abordagens de rPPG serão discutidos na secção 4.1, sendo que todos partilham a estrutura geral ilustrada na Figura 2.

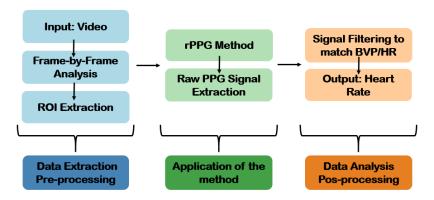


Figura 2: Representação esquemática simplificada do funcionamento global de um algoritmo de rPPG.

Esta técnica depende de muitos fatores externos que podem comprometer a sua eficácia: a distância do sujeito à câmara, se este está estático ou se se está a mover, a qualidade e resolução da câmara, a faixa de comprimentos de onda captados pela câmara (algumas câmaras RGB captam NIR), o tipo de luz ambiente, a coloração da pele, a presença de materiais entre a superfície da pele e a câmara (produtos de maquilhagem ou pelos faciais) entre outros. Estes fatores introduzem ruído no sinal, tornando essencial a sua filtragem com recurso à FFT, entre as frequências de referência para a FC (filtro passa-banda entre 0.6Hz e de 3.3Hz, correspondente à faixa 36-198bpm [26]).

A utilização de EVM pode facilitar a captura do sinal rPPG, ao amplificar as micro variações das cores dos pixeis e dos micro movimentos faciais [13][14]. Estes amplificadores permitem aumentar a quantidade e a variedade de dados utilizados, diminuir a qualidade de captura e evidenciar características mais ocultas que possam ser determinantes. Isto permite aproximar o vídeo a situações extremas para forçar o algoritmo a captar as características que melhor se enquadram para estimar o PPG em situações não ideias. Assim garantimos que o modelo generaliza não só para as situações ideais (indivíduo estático), mas também para situações mais próximas ao real.

#### 4 Métodos da rPPG-Toolbox

Para estimar o ritmo cardíaco através de rPPG, recorreu-se à rPPG-Toolbox [26]. Uma ferramenta desenvolvida para medir biossinais, com especial ênfase na FC, através do sinal PPG extraído de vídeos faciais. Esse sinal, como já referido, possui uma relação com o BVP, permitindo a posterior estimativa da FC com um grau elevado de confiança. A toolbox é principalmente desenvolvida em Python e inclui suporte para CUDA, dada a integração de bibliotecas compatíveis com processamento em GPU.

A toolbox disponibiliza vários métodos classificados como "supervisionados" ou "não supervisionados". Em ambos os casos, o teste de um método num determinado conjunto de dados requer a prévia configuração de um ficheiro "config.yaml". Após a testagem, o desempenho do modelo é avaliado com base nas métricas de avaliação da toolbox, sendo também gerados gráficos Bland-Altman e um ficheiro ".pickle" com dados relativos às previsões dos valores de FC obtidas pelo modelo e com os respetivos valores reais. Este último ficheiro pode ser posteriormente utilizado em códigos de visualização, que permitem uma análise comparativa mais simples e intuitiva entre o sinal previsto e o sinal real, nos diferentes segmentos dos vídeos.

No caso de se pretender treinar um método supervisionado, é necessário começar por configurar um ficheiro "train.yaml". Durante o treino vão sendo apresentadas algumas métricas que permitem acompanhar a evolução do modelo. No final, são também gerados gráficos das perdas ao longo das épocas (Loss vs. Epoch) e da taxa de aprendizagem ao longo das atualizações (Learning Rate vs. Scheduler Step).

# 4.1 Modelos Supervisionados

Os modelos supervisionados recorrem a algoritmos de aprendizagem com redes neuronais profundas e dados rotulados, ou seja, cada frame de vídeo possui, dependendo do dataset, algum tipo de informação que permite a futura estimativa da FC. Durante o treino, estes modelos aprendem a identificar padrões relevantes, a maioria ocultos ao olho humano, para a extração do sinal de PPG.

• **DeepPhys** [15]: É uma rede neuronal convolucional com duas ramificações 2D. Uma delas analisa o frame estático, enquanto que a outra analisa as diferenças entre frames, e, consequentemente, as variações ao longo do tempo. O objetivo do modelo é prever a variação instantânea (primeira derivada) do sinal de pulso obtido na PPG.

- PhysFormer [16]: É um modelo baseado em transformadores que combinam informações espaciais e temporais, tanto de curto como de longo prazo, para melhorar a estimativa do sinal PPG, permitindo dar mais importância a padrões globais no vídeo. Para o treino, utiliza uma função de erro mais complexa, que inclui vários termos: um que penaliza a falta de correlação com o sinal real, outro que compara os espectros de frequência e ainda um que avalia a distribuição esperada dos valores.
- TS-CAN [17]: Tal como o DeepPhys, é uma rede com duas ramificações e possui um mecanismo que considera o deslocamento temporal entre frames. Isto permite que o modelo aprenda padrões espaciais e temporais de forma eficiente. Foi desenvolvido para funcionar em tempo real, diretamente em dispositivos com poucos recursos. O sinal que estima é também a primeira derivada da forma de onda do PPG.
- EfficientPhys [18]: É uma rede neuronal leve e com apenas uma ramificação, desenhada para ser executada em tempo real em dispositivos com capacidade limitada. Inclui um módulo de normalização que calcula as diferenças entre frames que o próprio modelo ajusta. Também possui mecanismos que identificam os pixeis da pele mais relevantes para captar o sinal de pulso.

A toolbox dispõe de modelos supervisionados pré-treinados. Isto possibilita que utilizadores sem acesso a hardware computacional sofisticado, como CPUs de elevada capacidade de processamento ou GPUs, possam testar os métodos sem efetuar treino. A principal desvantagem na utilização destes modelos pré-treinados é o utilizador ficar limitado aos parâmetros de treino definidos por terceiros e não possuir informações acerca da evolução do processo de treino.

## 4.2 Modelos Não Supervisionados

Os métodos não supervisionados recorrem a operações algébricas e a técnicas de processamento de sinal para a extração do sinal PPG através da informação presente nos pixeis de cada frame dos vídeos.

• Green [19]: Ao analisar vídeos RGB, utiliza apenas o canal verde para estimar o sinal PPG, visto que, este possui dados mais relacionados com a pulsação. Para tal, realiza-se uma média espacial dos valores RGB, o que suaviza a imagem e facilita a análise do sinal extraído.

- ICA (Análise em Componentes Independentes) [20]: É aplicada aos canais de cor, depois de aplicadas normalizações e médias espaciais. Esta técnica separa os sinais misturados e sobrepostos, para recuperar e analisar as suas componentes independentes, o que inclui o sinal PPG.
- CHROM: [21] Baseia-se na análise da combinação linear dos sinais de crominância (variações de cor) extraídos do vídeo RGB. A ideia é estimar o sinal PPG a partir das diferenças subtis nas cores da pele, que variam com as diferenças do fluxo sanguíneo, associado aos batimentos cardíacos.
- POS (Plano Ortogonal ao Tom de Pele) [22]: Este método projeta os valores dos pixeis, já normalizados e depois de aplicadas médias espaciais, num plano que é ortogonal (perpendicular) ao tom de pele. A projeção é feita com base nos princípios fisiológicos e ópticos associados ao rPPG. Isto permite a captação do sinal por rPPG com menos influência de variações de iluminação e movimento.
- PBV (Volume de Sangue no Pulso) [23]: Utiliza o padrão característico do BVP, que depende do espetro de luz ambiente e das variações no volume sanguíneo para extrair o sinal PPG ao mesmo tempo que tenta compensar o ruído causado por movimento e outras fontes nos vídeos RGB.
- LGI (Local de Gradientes Invariante) [24]: Trata-se de uma forma de representar o sinal que é resistente ao movimento. Usa transformações locais diferenciáveis que permitem manter a informação fisiológica relevante, mesmo quando há deslocamentos ou alterações na posição do rosto.
- OMIT (Transformação Ortogonal da Matriz Imagem) [25]: Começa por recolher o sinal RGB e depois aplica uma primeira projeção matricial num subespaço para remover as variações principais dos dados (ruído por movimento ou luz). Depois é aplicada outra projeção ortogonal que extrai o sinal BVP.

#### 4.3 Datasets

A toolbox inclui ainda suporte para alguns datasets já pré-definido. Assim, basta ter acesso a algum destes datasets e conseguimos logo facilmente utilizá-los quer para treino quer para teste dos modelos. Os seguintes datasets foram utilizados:

- UBFC-rPPG [28]: A componente pública deste dataset inclui 42 sujeitos, cada um gravado durante 1 minuto a cerca de 1 metro de uma câmara *RGB Logitech C920 HD Pro webcam* com resolução 640x480 e 30 fps. Os vídeos encontram-se em formato .avi e os participantes realizavam um jogo matemático cujo objetivo visava aumentar a FC e simular uma interação humano-computador mais natural. Enquanto isso, foi recolhida informação através de um oxímetro *CMS50E*, a uma taxa de aquisição de 30 fps. Os dados foram armazenados num ficheiro .txt e dispostos em três linhas sinal PPG normalizado, FC, instante de tempo.
- PURE [29]: Este dataset inclui 10 participantes, cada um com seis vídeos de 1 minuto cada, gravados em diferentes condições com o indivíduo quieto, a falar, translação lenta, translação rápida, rotação pequena e rotação média (estes últimos quatro relativos ao movimento da cabeça). Ao contrário do dataset anterior, os vídeos estão decompostos em imagens .png, correspondentes a cada frame. Os vídeos foram gravados com uma câmara eco274CVGE SVS-Vistek GmbH a 30 fps com uma resolução 640x480 pixeis. Durante a recolha os indivíduos mantiveram-se a cerca de 1.1 metros da câmara. Para a recolha de dados biométricos foi utilizado um oxímetro CMS50E, que forneceu dados do sinal PPG, consequentemente da FC e ainda de SpO2, a uma taxa de aquisição de 60 Hz. Estes dados foram guardados em ficheiros .json.

Um dos datasets utilizados no estudo foi o COHFACE, tendo sido necessário implementar código adicional, uma vez que não estava previamente integrado na toolbox. (1)

• **COHFACE** [30]: Composto por 40 participantes, cada um gravou um conjunto de quatro vídeos de 1 minuto cada, dois deles em "boas condições"e os restantes "em condições mais naturais (degradadas)". Os vídeos foram gravados com uma câmara *Logitech HD C525*, com resolução 640x480 pixeis a 20 fps e armazenados em formato *.avi*. Quanto aos dados biométricos, para além do sinal BVP (análogo ao sinal PPG) que permite calcular FC, também foi extraído sinal que permite calcular RR. Para tal, foi utilizado um sensor de pulso *SA9308M* a 256Hz e um cinto respiratório *SA9311M* a 32Hz. Estes dados bem como informações de género e data de nascimento dos sujeitos foram guardados em ficheiros *.hdf*5.

<sup>&</sup>lt;sup>(1)</sup>A toolbox possui algumas instruções caso o utilizador queira introduzir um novo dataset ou caso o utilizador queira desenvolver o seu próprio modelo. Para isto convém que o utilizador possua algumas competências de programação.

#### 4.4 Ferramentas para Análise dos Resultados

#### 4.4.1 Métricas de avaliação teste

Foram utilizadas métricas como o MAE, RMSE, MAPE e a correlação de Pearson para avaliar o desempenho dos métodos num determinado conjunto de dados de teste.

• Erro Absoluto Médio (MAE): Mede a média dos valores absolutos das diferenças entre os valores reais  $R_g$  (*ground truth*) e os valores preditos  $R_p$ . Quanto menor o seu valor, melhor o desempenho.

$$MAE = \frac{1}{N} \sum_{n=1}^{N} \left| R_g - R_p \right| \tag{1}$$

• Raiz do Erro Quadrático Médio (RMSE): Calcula a raiz quadrada da média dos quadrados dos erros, o intuito é semelhante ao do MAE, mas é mais sensível a desvios elevados.

$$RMSE = \sqrt{\left(\frac{1}{N}\sum_{n=1}^{N}(R_g - R_p)^2\right)}$$
 (2)

• Erro Percentual Absoluto Médio (MAPE): Representa o erro médio relativo entre os valores reais e preditos, em percentagem. Útil para comparação entre diferentes escalas, visto que permite avaliar o erro de forma relativa.

$$MAPE = \frac{1}{N} \sum_{n=1}^{N} \left| \frac{R_g - R_p}{R_g} \right| \tag{3}$$

• Correlação de Pearson (ρ): Quantifica a correlação linear entre os valores reais e os estimados. Varia entre +1 e -1, sendo que +1 significa correlação linear perfeita positiva; -1 correlação linear perfeita negativa; e 0 ausência de correlação linear.

$$\rho = \frac{\sum_{n=1}^{N} \left( R_{g,n} - \overline{R_g} \right) \left( R_{p,n} - \overline{R_p} \right)}{\sqrt{\left( \sum_{n=1}^{N} \left( R_{g,n} - \overline{R_g} \right)^2 \right) \left( \sum_{n=1}^{N} \left( R_{p,n} - \overline{R_p} \right)^2 \right)}}$$
(4)

#### 4.4.2 Visualização Pré-Processamento

A toolbox disponibiliza códigos em formato *Jupyter notebook* que facilitam a visualização dos dados de entrada utilizados pelos algoritmos no pré-processamento dos datasets. Na Figura 3, observa-se um frame pixelizado com a resolução ajustada, bem como o respetivo frame após as normalizações algébricas aplicadas. Já na Figura 4, observam-se as *labels* de entrada, correspondentes aos valores de PPG e FC média (frequência dominante), associados a cada *chunk* (180 frames). A interface do código permite que o utilizador pesquise por sujeito, por *chunk* e ainda, no caso da Figura 3, por frame. <sup>(2)</sup>



Figura 3: *Input* que entra nos modelos a partir dos frames dos vídeos.

Figura 4: *Labels* do sinal PPG e estimativa da FC média por chunk.

<sup>&</sup>lt;sup>(2)</sup>As Figuras 3, 4 e 5 apresentam dados relativos ao sujeito 44 do dataset UBFC-rPPG. No caso da Figura 5, esta resulta da aplicação do método DeepPhys, treinado sobre o mesmo conjunto de dados. De acordo com os autores do dataset, o participante 44 consente com a utilização dos seus dados.

#### 4.4.3 Visualização Pós-Processamento

Após os ciclos de teste dos modelos, são gerados dois gráficos Bland-Altman por conjunto de resultados (ver pares de Figuras 6 e 7; 8 e 9; 10 e 11; 12 e 13). Um dos gráficos compara os valores da FC prevista relativamente ao real, para cada observação (valor médio da FC, num vídeo total). A reta y=x corresponde ao ideal, se os pontos incidirem sobre essa reta, significa que o valor estimado é igual ao real. O segundo gráfico representa as observações consoante a diferença média entre o valor estimado e o real, em função da média entre esses dois valores. Este último gráfico possui ainda as retas relativas ao erro médio de todas as observações e os IC a 95%.

De forma semelhante às Figuras da subsecção 4.4.2, a toolbox também disponibiliza outro *Jupyter Notebook*, com uma interface que permite selecionar o sujeito e o *chunk* e onde é possível observar o sinal PPG estimado sobreposto com o sinal real.

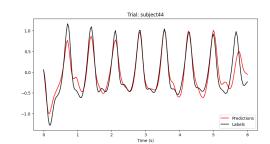


Figura 5: Gráfico dos sinais PPG real e estimado num *chunk*.

## 5 Resultados e Discussão

Este capítulo está organizado em duas secções principais. Na primeira, são abordados os resultados obtidos pelos modelos supervisionados, que recorrem a ML. Na segunda secção, são testados modelos não supervisionados, que não recorrem a métodos de ML, mas sim a algoritmos algébricos de processamento de sinal.

### 5.1 Modelos Supervisionados

Os resultados obtidos para os modelos supervisionados encontram-se organizados na Tabela 1:

			Dataset Teste					
Modelo Supervisionado	Dataset Treino	Métricas	UBF	с	PU	RE	COHFACE	
			Resultado	Incerteza	Resultado	Incerteza	Resultado	Incerteza
	UBFC	MAE	0,98	0,92	4,13	1,85	3,40	0,66
		RMSE	2,93	2,84	14,78	11,39	9,02	5,07
		MAPE (%)	1,11	1,05	3,77	1,44	4,38	0,80
EFFICIENTPHYS		Pearson	0,97	0,08	0,80	0,08	0,72	0,06
EFFICIENTENTS		MAE	2,68	1,45	0,59	0,30	9,77	1,42
	PURE	RMSE	9,76	9,29	1,19	0,99	20,40	8,66
	PURE	MAPE (%)	2,65	1,34	0,97	0,54	13,60	2,06
		Pearson	0,87	0,08	1,00	0,02	0,39	0,07
		MAE	2,25	1,42	5,78	2,38	6,03	1,00
	UBFC	RMSE	4,81	3,87	19,20	13,84	14,05	6,69
	OBFC	MAPE (%)	2,37	1,52	5,66	1,99	7,55	1,19
DEEPPHYS		Pearson	0,94	0,13	0,65	0,10	0,46	0,07
DEEPPHIS	PURE	MAE	1,20	0,42	0,29	0,16	23,94	1,40
		RMSE	2,95	1,97	0,62	0,50	29,78	8,41
		MAPE (%)	1,42	0,51	0,42	0,23	31,67	1,64
		Pearson	0,99	0,02	1,00	0,01	-0,21	0,08
	UBFC	MAE	1,37	1,19	10,48	2,48	6,75	0,79
		RMSE	3,82	3,70	21,75	11,01	12,06	5,23
		MAPE (%)	1,53	1,35	20,21	4,95	9,05	1,02
PHYSFORMER		Pearson	0,96	0,11	0,61	0,10	0,46	0,07
PHISTORIVIER	PURE	MAE	1,38	0,54	0,51	0,28	14,87	1,03
		RMSE	3,74	2,82	1,11	0,99	19,74	7,26
		MAPE (%)	1,60	0,62	0,86	0,52	22,62	1,68
		Pearson	0,98	0,03	1,00	0,02	0,25	0,08
		MAE	1,95	1,22	1,75	1,18	5,31	0,89
	UBFC	RMSE	4,14	3,27	9,24	9,11	12,40	5,98
		MAPE (%)	2,04	1,28	1,79	0,95	6,65	1,05
TSCAN		Pearson	0,95	0,12	0,92	0,05	0,55	0,07
ISCAN	PURE	MAE	1,42	0,40	0,59	0,30	21,98	1,29
		RMSE	2,98	1,75	1,19	0,99	27,39	8,05
		MAPE (%)	1,62	0,47	0,97	0,54	29,68	1,54
		Pearson	0,99	0,02	1,00	0,02	0,09	0,08

Tabela 1: Resultados das métricas de avaliação nos modelos supervisionados.

Todos os testes realizados no estudo utilizaram modelos pré-treinados, isto é, modelos disponibilizados pelos autores da toolbox. Daí o dataset COHFACE não aparecer na coluna dos datasets de treino. Todos os modelos possuem um treino de 30 *epochs*.

#### 5.1.1 Testes Intra-Dataset

Nesta subsecção apenas serão abordados modelos que foram testados dentro do mesmo dataset onde foram treinados  $\begin{cases} A_{80\%} \to A_{20\%} \\ B_{80\%} \to B_{20\%} \end{cases}$  (3), com 80% para treino e 20% para teste, sem interseção dos conjuntos. Os resultados abordados nesta subsecção estão inseridos nas duas primeiras colunas da Tabela 1, que também contém resultados de cruzamento de datasets (abordados na subsecção 5.1.2).

Todos os modelos obtiveram desempenhos muito elevados nas respetivas métricas de avaliação em ambos os datasets. Destacaram-se quatro conjuntos de resultados com correlações de Pearson muito próximas do ideal<sup>(4)</sup>, todos para o dataset PURE. Isto significa que, para cada modelo e dataset, durante o treino, foram captadas as características que permitem uma boa estimativa da FC dentro do conjunto de dados.

Os modelos treinados e testados no PURE obtiveram melhores resultados nas métricas que os do UBFC, o que parece contraditório, uma vez que no PURE há indução de movimento (ruído) nos dados. Por outro lado, isto faz com que o modelo capte as características mais fundamentais do sinal PPG, já que parâmetros associados ao movimento são diferentes de vídeo, para vídeo. O facto de possuir menos sujeitos também pode ter contribuído para alguma eventual sobrevalorização nos resultados, uma vez que há menor variabilidade nos dados de testes.

O conjunto que obteve os melhores resultados foi o modelo DEEPPHYS no dataset PURE. Podemos analisar mais detalhadamente os resultados obtidos neste caso:

<sup>(3)</sup> A, B e C (aparecerá mais à frente), serão usados para esquematização e simplificação da ideia. A e B, correspondem aos datasets UBFC e PURE, enquanto que C ao COHFACE.

<sup>&</sup>lt;sup>(4)</sup>Nota: os valores de correlação de Pearson apresentados não foram ideais como apresentados na Tabela 1. Isso ocorreu fruto de arredondamento, os valores aproximados são, respetivamente 0,9981; 0,9986; 0,9985 e 0,9981.

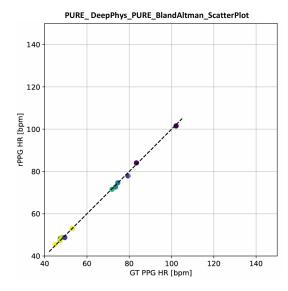


Figura 6: Scatter Plot do modelo DeepPhys no dataset PURE.

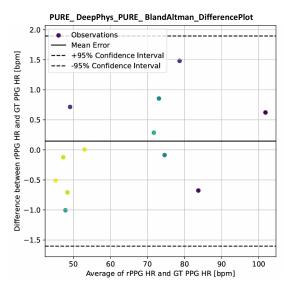


Figura 7: Difference Plot do modelo Deephys no dataset PURE.

O gráfico da esquerda coloca o valor da FC previsto em função do real para cada vídeo (valor médio). Como podemos observar os pontos acompanham a reta linear y=x, que representa a faixa ideal, onde as estimativas são iguais aos valores do ground truth. No gráfico da direita temos o valor da diferença entre o estimado e o real em função da média destes valores. Como expectável pelos resultados anteriores, todas as observações estão dentro do intervalo de confiança de 95% e o erro médio (reta horizontal) está muito próximo de zero.

Por outro lado, como os testes foram efetuados para indivíduos que, embora diferentes dos utilizados no treino (grupos de treino e de teste independentes), pertencem ao mesmo dataset, já seria de esperar este tipo de resultados um pouco "inflacionados". Nos modelos de ML ocorre sempre um ajuste do modelo aos dados. Dentro de uma base de dados as condições são semelhantes em todas as recolhas (espaço, luminosidade, etc.), logo, o modelo ao ajustar alguns dos seus pesos em função destas variáveis, como estes parâmetros são semelhantes, ele consegue generalizar facilmente. Se o submetermos a condições radicalmente diferentes, possivelmente não terá a mesma eficácia.

#### 5.1.2 Testes Inter-Dataset

Dados os resultados obtidos na subsecção 5.1.1, vamos avaliar a generalização dos modelos, utilizando um modelo treinado num dataset e testado noutro. Os datasets de treino ficaram divididos em 80% treino e 20% validação, enquanto que os de teste foram usados na sua totalidade.

#### Datasets Internos

Aqui, testámos os modelos para datasets "internos" (UBFC-rPPG e PURE), isto é, datasets já suportados pela toolbox.  $\left\{ \begin{array}{l} A \to B \\ B \to A \end{array} \right.$ . Os resultados estão também eles nas duas primeiras colunas da Tabela 1.

Tal como previsto, o facto de testarmos o modelo com um dataset diferente fez com que a maioria deles não apresentasse resultados tão robustos como os da subsecção anterior. Ainda assim tivemos alguns modelos que apresentaram resultados bastante bons, com correlações de Pearson acima de 98%, tendo sido ambas as combinações do modelo TS-CAN e, para o DEEPPHYS e o PHYSFORMER, os casos treinados no PURE e testados no UBFC.

Neste caso particular, os modelos treinados no PURE e testados no UBFC apresentaram melhores resultados para todos os modelos. Isto pode dever-se a dois fatores relativos às características de cada dataset: o PURE, apesar de possuir um menor número de indivíduos (menor variabilidade de sujeitos), possui um número superior de vídeos e, consequentemente, de dados; paralelamente, o PURE possui também seis gravações por participante, onde são induzidos movimentos específicos diferentes na cabeça de cada um (ruído). O que permite que, para um maior conjunto de treino, os modelos sejam forçados a generalizar consoante características que não dependam tanto de movimentos ruidosos.

O conjunto que obteve melhores resultados foi o modelo DEEPPHYS treinado no dataset PURE e testado no UBFC. Com os respetivos gráficos associados:

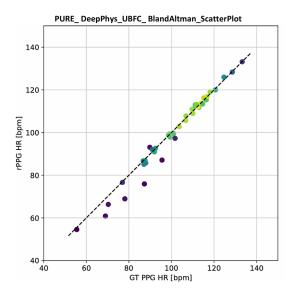


Figura 8: Scatter Plot do modelo Deephys treinado PURE e testado no UBFC.

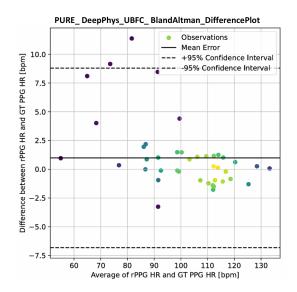


Figura 9: Difference Plot do modelo Deephys treinado PURE e testado no UBFC.

A maioria dos pontos encontra-se sobre a reta ideal, o que está em conformidade com os resultados obtidos pelas métricas de avaliação. Embora, no Scatter Plot, se verifiquem alguns pontos ligeiramente mais afastados da reta ideal, também importa ressalvar que, como testámos para um dataset completo, o número de observações é substancialmente mais elevado, o que aumenta proporcionalmente a dispersão. No que diz respeito ao Difference Plot, e conforme expectável, a maioria dos pontos encontra-se relativamente próxima de zero, apesar de existirem alguns outliers, dado o elevado número de vídeos.

#### Dataset Externo

Aqui, testámos num dataset para o qual a toolbox não possuía recursos pré-definidos, COHFACE, tendo sido necessárias algumas implementações manuais  $\begin{cases} A \to C \\ B \to C \end{cases} .$  Os resultados ocupam a última coluna da Tabela 1.

O desempenho dos modelos foi muito inferior ao antecipado. Como os resultados da generalização obtidos anteriormente tinham sido promissores, tendo em conta a mudança de dataset, era de esperar que aqui também fosse semelhante (apesar das diferenças intrínsecas entre os datasets, nomeadamente as diferentes taxas de aquisição dos vídeos). Na realidade o que observámos foi que nenhum dos modelos conseguiu generalizar face aos parâmetros do dataset COHFACE.

Quando testámos para os datasets internos, o modelo com os melhores resultados tinha sido o TSCAN, treinado no PURE. Aqui, nesta bateria de testes, este foi dos que apresentou piores resultados nas métricas de avaliação estipuladas. Outro facto curioso foi que, os modelos treinados no UBFC apresentaram melhores resultados relativos aos treinados no PURE, contrariamente ao que se observou anteriormente.

O conjunto que obteve melhor resultados foi o modelo do EFFICIENTPHYS treinado no UBFC, cujos respetivos gráficos foram:

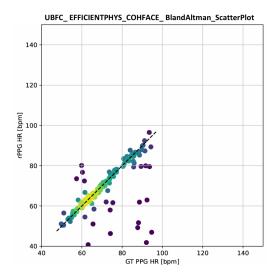


Figura 10: Scatter Plot do modelo EfficientPhys treinado UBFC e testado no COHFACE.

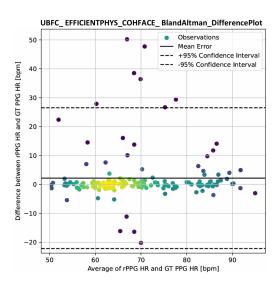


Figura 11: Difference Plot do modelo EfficientPhys treinado UBFC e testado no COHFACE.

Como podemos ver pelos gráficos, há um número muito mais elevado de outliers e de pontos longe da reta ideal no Scatter Plot, apesar da maioria permanecer próximo dela. Por outro lado, no Difference Plot, o intervalo de confiança a 95% esperado para as observações também aumentou. Embora o erro médio, do ponto de vista gráfico, se encontre próximo de zero, importa ter em consideração que o fator de escala do eixo vertical (eixo y) é inferior ao dos gráficos anteriormente apresentados. Esta diferença resulta num erro médio ligeiramente superior aos obtidos anteriormente. Estes resultados estão em conformidade com os obtidos pelas métricas de avaliação, foram piores que os anteriores.

Podemos atribuir os maus resultados ao facto dos vídeos no COHFACE possuírem fraca qualidade e, principalmente, por possuírem uma menor taxa de aquisição (20 fps ao invés de 30 fps), diferente da utilizada pelos datasets de treino. Com a agravante de que metade dos vídeos foram propositadamente gravados em "condições naturais", ou seja, mais degradados. Isto dificulta a captura e o processamento dos parâmetros que permitem que os modelos estimem com rigor o sinal PPG para cada vídeo.

## 5.2 Modelos Não Supervisionados

Nesta secção, testámos diferentes modelos não supervisionados, para os três datasets já referidos nas abordagens supervisionadas. Os algoritmos não utilizam ML para estimar a FC, mas sim técnicas de processamento fundamentadas por princípios algébricos de manipulação e análise de imagens. Os modelos utilizados foram sucintamente explicados na secção 4.2.

		Dataset Teste					
Modelo Não Supervisionado	Métricas	UBFC Resultados Incerteza		PURE		COHFACE Resultados Incerteza	
	DAAF.						
	MAE	16,19	3,53	4,78	2,09	10,82	0,86
ICA	RMSE	28,02	14,50	16,73	12,40	15,31	5,28
	MAPE (%)	15,35	3,20	4,48	1,66	15,08	1,17
	Pearson	0,45	0,14	0,72	0,09	0,12	0,08
	MAE	3,62	1,25	2,62	1,28	12,52	0,71
POS	RMSE	8,87	7,11	10,14	7,61	15,44	4,75
	MAPE (%)	3,58	1,22	5,26	2,66	19,01	1,21
	Pearson	0,90	0,07	0,91	0,05	0,22	0,08
	MAE	3,31	1,00	4,51	1,76	11,33	0,69
CHROM	RMSE	7,28	4,93	14,25	9,51	14,30	4,44
	MAPE (%)	3,19	0,93	8,96	3,65	17,04	1,17
	Pearson	0,92	0,06	0,83	0,07	0,18	0,08
	MAE	23,86	4,42	11,52	3,17	12,78	1,38
GREEN	RMSE	37,30	18,47	26,93	16,59	21,67	7,90
	MAPE (%)	22,65	3,88	12,10	2,61	15,98	1,66
	Pearson	0,28	0,15	0,21	0,13	-0,07	0,08
	MAE	18,98	4,21	6,49	2,59	22,30	1,18
LGI	RMSE	33,21	17,54	20,93	15,05	26,85	7,78
	MAPE (%)	17,55	3,64	6,63	2,20	30,29	1,43
	Pearson	0,26	0,15	0,57	0,11	-0,07	0,08
	MAE	18,44	4,03	3,51	1,53	19,11	1,22
PBV	RMSE	31,96	16,71	12,24	10,96	24,55	7,75
	MAPE (%)	17,70	3,61	4,44	1,41	25,74	1,45
	Pearson	0,42	0,14	0,86	0,07	0,01	0,08
	MAE	18,98	4,21	4,98	2,13	22,30	1,18
ОМІТ	RMSE	33,21	17,54	17,14	13,40	26,85	7,78
OWIT	MAPE (%)	17,55	3,64	5,48	1,91	30,29	1,43
	Pearson	0,26	0,15	0,72	0,09	-0,07	0,08

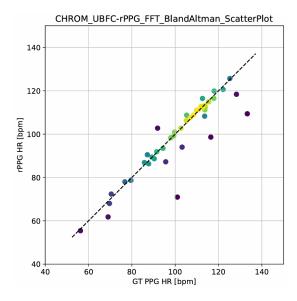
Tabela 2: Resultados das métricas de avaliação nos modelos não supervisionados

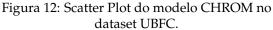
De um modo geral, os resultados obtidos não foram tão bons quanto os obtidos pelos modelos supervisionados, testados e treinados no mesmo dataset (subsecção 5.1.1), ou generalizados para datasets internos em 5.1.2. O que nos permite supor que, possivelmente modelos que utilizem abordagens de ML são mais precisos do que os modelos não supervisionados.

Relativamente à comparação entre datasets, observamos novamente que, os resultados obtidos no COHFACE foram piores do que os do PURE e UBFC. Assim, continuamos a considerar como explicação mais plausível para esta discrepância entre datasets, o facto dos vídeos do COHFACE apresentarem uma qualidade inferior.

Olhando agora apenas para os modelos não supervisionados, o método que obteve melhores resultados, considerando os 3 datasets foi o modelo POS. A essência deste algoritmo permite eliminar a componente constante da tonalidade da pele, que não serve para o cálculo da FC e realçar as variações dinâmicas nos pixeis. O facto de possuir uma projeção fixa e normalizar o espaço confere-lhe maior robustez e menor suscetibilidade a ruído, para os diversos tipos de dados.

Contudo, os melhores resultados surgiram para o método CHROM no dataset UBFC. O CHROM utiliza uma combinação linear de sinais de crominância extraídos dos canais de RGB dos vídeos, isto confere robustez para ambientes mais degradados ao nível da iluminação ambiente. No entanto, o facto desse método não ter conseguido resultados semelhantes para nenhum dos restantes datasets poderá indicar um comprometimento ao nível do ruído provocado pelo movimento da face, visto que os outros dois datastes induzem isso nos participantes.





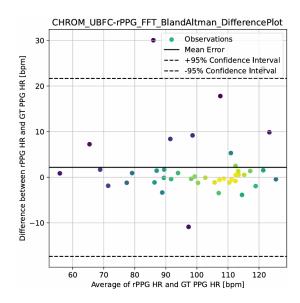


Figura 13: Difference Plot do modelo CHROM no dataset UBFC.

Estes foram os gráficos obtidos ao correr o algoritmo na toolbox. À esquerda, o Scatter Plot onde vemos que a maioria dos pontos está sobre a reta ideal, salvo algumas exceções. E finalmente, à direita, apesar de alguns pontos outliers, observa-se que a maioria dos pontos se distribui sobre uma reta horizontal próxima ao zero. Gráficos compatíveis com os resultados obtidos nas métricas da Tabela 2.

## 6 Criação de um Dataset Privado

Paralelamente ao trabalho desenvolvido, também levámos a cabo a criação de um dataset privado, juntamente com a Ana Ramos [31]. O dataset possui vídeos de baixa qualidade, gravados com as *webcams* integradas dos nossos computadores pessoais, com o intuito de avaliar, em trabalhos futuros, a exequibilidade da aplicação de modelos de rPPG mais robustos e adaptados a contextos realistas. Esta abordagem procura refletir as condições da grande parte da população, que não dispõe de câmaras de elevada qualidade, ou cujo acesso a tais equipamentos é limitado.

Cada vídeo enquadra a face do participante e a região dos ombros (alguns métodos que estimam parâmetros do RR, dependem do seu movimento). As gravações foram realizadas com câmaras que possuíam uma resolução de 1280x720 (720p) e uma taxa de aquisição de 30 fps. Cada vídeo tem uma duração aproximada de 1 minuto.

Para além da FC, recolhida através de uma banda de peito *Polar H10* (Figura 14 <sup>(5)</sup>), também foi medida a frequência do RR, recorrendo a um oxímetro *Enttgo OLED TY05* <sup>(6)</sup> (Figura 15). Ambos os biossinais foram adquiridos a uma taxa de aquisição de 1 Hz. A banda torácica referida é amplamente utilizada para monitorização cardíaca em contexto de investigação [27] com elevada precisão, no entanto, é classificada como um dispositivo de grau recreativo, não sendo certificada como dispositivo médico. Tal como a banda, o oxímetro encontra-se comercialmente disponível e também carece de certificação médica.



Figura 14: Representação da utilização da banda torácica *Polar H10* 



Figura 15: Fotografia da utilização do oxímetro *Enttgo OLED TY05*, onde entre os valores de Sp0<sub>2</sub> e FC, está RR/min

<sup>&</sup>lt;sup>(5)</sup>Para evitar colocar uma imagem real, da região do tórax de um dos participantes, foi gerado um "manequim"com recurso a IA e depois editada a imagem com uma fotografia do dispositivo real.

<sup>&</sup>lt;sup>(6)</sup>Apesar de também estimar FC, apenas foi utilizada para RR, uma vez que a banda é mais precisa.

#### Dividimos o dataset em dois subdatasets:

→ O primeiro consistiu numa gravação simples de um vídeo facial de 1 minuto enquanto medíamos os biossinais referidos. Aqui, procurámos ter algum equilíbrio entre os sexos, dos 42 participantes 23 pertencem ao sexo masculino e 19 ao sexo feminino. E também uma faixa de idades variada, com idades compreendidas entre os 18 e os 82 anos, embora a maior incidência seja a camada jovem (18-22 anos). Foi pedido aos participantes que permanecessem estáticos, para controlo de ruído, apesar de que, em alguns casos, os participantes, de forma imprevisível, tenham movido a cabeça e/ou começado a falar. Foram também gravados em diferentes locais, com o intuito de diversificar as condições ambientais e externas.

→ O segundo subdataset incluiu 10 participantes, 8 do sexo masculino e apenas 2 do sexo feminino, com idades compreendidas entre os 19 e os 21 anos. Para cada participante foram gravados três vídeos, o primeiro em repouso, o segundo logo após uma corrida de 1 minuto de elevada intensidade e o terceiro, 15 minutos após a corrida.

Os vídeos foram armazenados em formato *mp4* e os restantes dados num ficheiro *excel* para cada indivíduo. O dataset com estes dados de todos os participantes permaneceu privado durante todo o desenvolvimento do projeto, uma vez que aguardamos por um parecer da Comissão de Ética da Universidade de Coimbra.

# 7 Limitações

Neste capítulo pretendo, de forma sucinta, expor algumas das limitações encontradas ao longo do desenvolvimento do trabalho. Embora estas não tenham influenciado diretamente os resultados obtidos, condicionaram a abordagem metódica adotada. Também considero relevante a partilha destas dificuldades pois, poderá ser útil para futuros investigadores interessados nesta área compreenderem os principais desafios que, possivelmente, irão encontrar e algumas formas de os contornar.

Começando pelo hardware, nesta área recorremos à utilização de modelos que exigem um pré-processamento de grandes volumes de dados, o que pode constituir um desafio dependendo do computador utilizado (se possui uma CPU potente ou se possui suporte para a utilização de uma GPU). Para além disto, muitos dos modelos mais robustos e disponíveis na comunidade científica funcionam exclusivamente com suporte CUDA (ou equivalente) e por isso, exigem utilização de GPU. Modelos que utilizem CPU, para além de demorarem mais tempo a serem executados, normalmente estão associados a modelos não supervisionados que apresentam menor robustez face aos supervisionados.

Outro fator de elevada importância prende-se com o facto de, para a avaliação destes modelos, serem necessários datasets que possuam, no mínimo: vídeos ou sequencias de frames, e informações relativas ao sinal PPG para extração e posterior cálculo de FC. O que resulta, não só, em datasets muito pesados, como também, por possuírem as imagens faciais de pessoas, devido a questões de privacidade e proteção de dados, serem de difícil acesso.

Neste caso, alguns dos modelos eram demasiado "pesados" para o computador utilizado (que possui apenas CPU) o que fez com que perdesse muitos ciclos de teste, dado que o computador deixava de responder. Também foi esse o motivo que me levou a escolher a rPPG-Toolbox, visto que esta possuía já modelos pré-treinados e não tinha de submeter o meu computador a treino um a um (cada treino levaria dias a correr). A utilização da toolbox, apesar de ter sido benéfica (não necessitei de programar modelos de raiz), também possui algumas desvantagens. Se o utilizador não possuir um conhecimento avançado em programação fica restrito aos resultados fornecidos pela toolbox. Para além disso, há pouca informação acerca do treino dos modelos pré-treinados (não são fornecidos gráficos de perda, ou outros parâmetros relevantes de treino) que já vinham com a toolbox.

#### 8 Conclusão

Após a análise dos resultados obtidos, verificou-se que os modelos supervisionados obtiveram resultados significativamente mais fiáveis e robustos do que os não supervisionados. Apesar de três dos modelos não supervisionados terem obtido, para datasets específicos, correlações de Pearson superiores a 90%, as métricas de erro revelaram-se algo elevadas, tendo em conta a finalidade do trabalho. Quando colocados à prova num dataset de menor qualidade, os resultados foram substancialmente inferiores, o que não lhes confere fiabilidade para utilização em contexto médico.

Relativamente aos modelos supervisionados, alguns atingiram excelentes resultados em datasets específicos, bem como em alguns dos testes de generalização, nos quais as métricas de avaliação foram promissoras, com elevadas correlações de Pearson e métricas de erro bastante reduzidas. Por outro lado, quando testados em datasets de pior qualidade, verificou-se igualmente uma redução na precisão e na fiabilidade do modelo, o que revela uma limitação, dado o objetivo final do estudo.

Para finalizar, considerando que o objetivo deste projeto é avaliar a aplicabilidade deste tipo de métodos em pacientes pós-cirurgia, para fins de monitorização, é necessário atentar a alguns aspetos fundamentais. Em contexto médico, não podem ocorrer falhas significativas e, dada a variação dos resultados obtidos, enquanto não for desenvolvido um modelo suficientemente robusto face a diferentes ruídos, variações fisiológicas entre pacientes, qualidade das câmaras utilizadas, entre outros fatores, será necessário garantir condições que mitiguem possíveis fontes de imprecisão nos algoritmos. Uma das opções poderá passar por garantir que o paciente disponha de uma câmara que cumpra com os requisitos mínimos de qualidade (no caso da utilização de um modelo supervisionado, com características técnicas compatíveis com as utilizadas nos datasets de treino dos modelos), bem como assegurar condições ambientais adequadas (espaço e iluminação) no local onde iria realizar as consultas (muito provavelmente na residência do doente).

Caso contrário, atualmente, ainda não consideraria a utilização de qualquer um destes modelos em contexto hospitalar, dado que nestas situações não podem ocorrer falhas graves. No entanto, trata-se de uma técnica muito promissora, que poderá certamente, no futuro, constituir uma mais-valia importante para a monitorização remota de pacientes.

## Referências

- [1] S. Carrillo de Albornoz, K. L. Sia, and A. Harris, "The effectiveness of teleconsultations in primary care: systematic review," Family Practice, vol. 39, no. 1, pp. 168–182, Jan. 2022, doi: 10.1093/fampra/cmab077.
- [2] D. H. Spodick, "Normal sinus heart rate: appropriate rate thresholds for sinus tachycardia and bradycardia," Southern Medical Journal, vol. 89, no. 7, pp. 666–667, Jul. 1996, doi: 10.1097/00007611-199607000-00003.
- [3] "Sinus rhythm," ScienceDirect, [Online]. Available: https://www.sciencedirect.com/topics/nursing-and-health-professions/sinus-rhythm. [Accessed: Jul. 7, 2025].
- [4] E. H. J. Mestrom et al., "Postoperative circadian patterns in wearable sensor measured heart rate: a prospective observational study," Journal of Clinical Monitoring and Computing, vol. 38, no. 1, pp. 147–156, Feb. 2024, doi: 10.1007/s10877-023-01089-z.
- [5] H. Xiao, T. Liu, Y. Sun, Y. Li, S. Zhao, and A. Avolio, "Remote photoplethysmography for heart rate measurement: A review," Biomedical Signal Processing and Control, vol. 88, no. Part B, Art. no. 105608, 2024, doi: 10.1016/j.bspc.2023.105608.
- [6] Sassenrath C, Barthelmäs M, Saur J, Keller J. Inducing empathy affects cardio-vascular reactivity reflected in changes in high-frequency heart rate variability. Cogn Emot. 2021 Mar;35(2):393-399. doi: 10.1080/02699931.2020.1826910. Epub 2020 Oct 1. PMID: 32998630.
- [7] R. Tiwari, R. Kumar, S. Malik, T. Raj, and P. Kumar, "Analysis of heart rate variability and implication of different factors on heart rate variability," Current Cardiology Reviews, vol. 17, no. 5, Art. no. e160721189770, 2021, doi: 10.2174/1573403X16999201231203854.
- [8] S. Beutler and I. Croy, "Psychophysiological reactions during the traumafilm paradigm and their predictive value for intrusions," European Journal of Psychotraumatology, vol. 14, no. 2, Art. no. 2281753, 2023, doi: 10.1080/20008066.2023.2281753.
- [9] C. H. Koo, B. Xue, V. Yik, I. Seow-En, M. E. H. Ong, and E. K. Tan, "Using heart rate variability to develop a predictive model for post-operative cardiovascular complications following major abdominal surgery: A pilot study," World Journal of Surgery, vol. 48, no. 3, pp. 585–597, Mar. 2024, doi: 10.1002/wjs.12077.

- [10] C.-H. Cheng, K.-L. Wong, J.-W. Chin, T.-T. Chan, and R. H. Y. So, "Deep learning methods for remote heart rate measurement: A review and future research agenda," Sensors, vol. 21, Art. no. 6296, 2021, doi: 10.3390/s21186296.
- [11] S. Kwon, J. Kim, D. Lee, and K. Park, "ROI analysis for remote photoplethysmography on facial video," in Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Milan, Italy, 2015, pp. 4938–4941, doi: 10.1109/EMBC.2015.7319499.
- [12] X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, Y. Wang, S. Sengupta, S. Patel, and D. McDuff, "rPPG-Toolbox: Deep Remote PPG Toolbox," \*arXiv preprint\* arXiv:2210.00716, 2022. [Online]. Available: https://arxiv.org/abs/2210.00716. [Accessed: Jul. 7, 2025].
- [13] M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Z. Rashad, "Video-based real-time monitoring for heart rate and respiration rate," *Expert Systems with Applications*, vol. 225, Art. no. 120135, 2023, doi: 10.1016/j.eswa.2023.120135.
- [14] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," ACM Transactions on Graphics (Proc. SIGGRAPH 2012), vol. 31, no. 4, 2012
- [15] W. Chen and D. McDuff, "DeepPhys: Video-based physiological measurement using convolutional attention networks," arXiv preprint arXiv:1805.07888, Aug. 2018.
- [16] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. S. Torr, and G. Zhao, "Physformer: Facial video-based physiological measurement with temporal difference transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 4186–4196.
- [17] X. Liu, J. Fromm, S. Patel, and D. McDuff, "Multi-task temporal shift attention networks for on-device contactless vitals measurement," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 19400–19411, 2020.
- [18] X. Liu, B. Hill, Z. Jiang, S. Patel, and D. McDuff, "EfficientPhys: Enabling simple, fast and accurate camera-based cardiac measurement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.* (WACV), 2023, pp. 5008–5017.
- [19] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [20] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, 2010.

- [21] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [22] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, 2016.
- [23] G. De Haan and A. Van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiol. Meas.*, vol. 35, no. 9, p. 1913, 2014.
- [24] C. S. Pilz, S. Zaunseder, J. Krajewski, and V. Blazek, "Local group invariance for heart rate estimation from face videos in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2018, pp. 1254–1262.
- [25] C. Álvarez Casado and M. Bordallo López, "Face2PPG: An unsupervised pipeline for blood volume pulse extraction from faces," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 11, pp. 5530–5541, Nov. 2023, doi: 10.1109/JBHI.2023.3307942.
- [26] X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, Y. Wang, S. Sengupta, S. Patel, and D. McDuff, "rPPG-Toolbox: Deep Remote PPG Toolbox," \*arXiv preprint\* arXiv:2210.00716, 2022. [Online]. Available: https://arxiv.org/abs/2210.00716
- [27] R. de A. Pereira, J. L. de B. Alves, J. H. da C. Silva, M. da S. Costa, and A. S. Silva, "Validity of a smartphone application and chest strap for recording RR intervals at rest in athletes," *Int. J. Sports Physiol. Perform.*, vol. 15, no. 6, pp. 896–899, 2020, doi: 10.1123/ijspp.2019-0406.
- [28] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Letters*, 2017.
- [29] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in \*Proc. 23rd IEEE Int. Symp. Robot Hum. Interactive Commun. (Ro-Man)\*, Edinburgh, Scotland, UK, 2014, pp. 1056–1062.
- [30] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *arXiv preprint*, Sep. 2017. [Online]. Available: http://publications.idiap.ch/index.php/publications/show/3688. [Accessed: Jul. 7, 2025].

[31] A. M. A. Ramos, \*Análise de marcadores biométricos e sinais vitais através de visão artificial: Ritmo Respiratório\*, trabalho de projeto, Universidade de Coimbra, jul. 2025.