



UNIVERSIDADE D  
COIMBRA

Beatriz Lopes de Carvalho

**MODELO DE INTELIGÊNCIA ARTIFICIAL  
PARA INFERIR AUTOMATICAMENTE OS  
ESTADOS MENTAIS DO PACIENTE  
ASSOCIADOS AO *ENGAGEMENT***

Dissertação no âmbito do Mestrado de Engenharia Biomédica, no ramo de Informática Clínica e Bioinformática, orientada pelo Professor Doutor Luís Miguel Machado Lopes Macedo e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Setembro de 2023



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE  
**COIMBRA**

**MODELO DE INTELIGÊNCIA ARTIFICIAL PARA  
INFERIR AUTOMATICAMENTE OS ESTADOS  
MENTAIS DO PACIENTE ASSOCIADOS AO  
*ENGAGEMENT***

Beatriz Lopes de Carvalho

Dissertação apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra  
para obtenção do grau de Mestre em Engenharia Biomédica com especialização em  
Informática Clínica e Bioinformática

**Supervisor: Prof. Dr. Luís Macedo**

Setembro de 2023



Este trabalho foi desenvolvido em colaboração com:

**Departamento de Engenharia Informática**

**Departamento de Engenharia Física**

**Faculdade de Ciências e Tecnologia da Universidade de Coimbra**

*Centre for Informatics and Systems of the University of Coimbra (CISUC)*

*Center for Research in Neuropsychology and Cognitive and Behavioral  
Intervention (CINEICC)*



Esta cópia da tese é fornecida na condição de que quem a consulta reconhece que os direitos de autor são da pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada sem a referência apropriada.

This thesis copy has been provided on the condition that anyone who consults it understands and recognizes that its copyright belongs to its author and that no reference from the thesis or information derived from it may be published without proper acknowledgement.



# Agradecimentos

Gostaria de expressar os meus sinceros agradecimentos a todas as pessoas que contribuíram para o desenvolvimento deste projeto e para o meu crescimento pessoal e académico.

Ao meu orientador, Professor Doutor Luís Macedo, estendo o meu profundo agradecimento pela sua orientação excecional, a sua paciência incansável e disponibilidade constante, que desempenharam um papel crucial no sucesso deste trabalho e na minha formação.

À neuropsicóloga Ana Rita Silva, agradeço imensamente por toda a ajuda e colaboração dedicadas. A sua experiência e contribuição foram fundamentais para o aprimoramento deste estudo.

À minha família, especialmente aos meus pais e à minha irmã, quero expressar a minha gratidão. O apoio constante de todos ao longo desta jornada foi um pilar essencial para mim. Cada passo que dei, cada conquista que alcancei, foi moldada pelo amor que sempre me proporcionaram.

Às minhas amigas Mariana, Leonor e Carolina, quero destacar o valor da nossa amizade e a força que me deram em todos os momentos. Agradeço por me aquecerem o coração com o vosso apoio constante e amor incondicional de uma amizade que sei que é para sempre.

À minha afilhada Maria, agradeço por seres uma fonte de alegria e inspiração na minha vida. Obrigada por seres tão especial.

Aos meus amigos de mestrado, o grupo dos "Palhinhas" e "Desintegrados", agradeço por compartilharmos momentos inesquecíveis em Coimbra, que tornaram esta viagem mais doce e memorável. A nossa amizade é um dos presentes mais preciosos que este percurso me trouxe.

Um agradecimento especial para a *Isecotuna*, a tuna da qual fiz parte desde o início da minha jornada académica. A música, a união e os momentos vividos, trouxeram uma dimensão única aos últimos anos, momentos esses que já deixam saudade.

A todos os que, de uma maneira ou outra, contribuíram para esta conquista, o meu mais sincero obrigado.



# Resumo

Neste estudo é abordado um desafio global na área da saúde: a manutenção do *engagement* dos pacientes com demência, em tratamentos de longo prazo. Reconhecendo a importância crucial do envolvimento do paciente para o sucesso destes tratamentos, procura-se identificar soluções eficazes para monitorizar e alertar sobre flutuações, ao nível de empenho ao longo do tempo. A crescente utilização de técnicas de Inteligência Artificial (IA) oferece uma abordagem promissora para detetar automaticamente essas situações e fornecer intervenções oportunas. Esta pesquisa concentra-se em desenvolver estratégias que alavanquem a IA para identificar padrões de *engagement*, através das expressões faciais, permitindo aos profissionais de saúde agir proativamente e otimizar o cuidado do paciente.

Para isso foi utilizado, inicialmente, um conjunto de dados constituído por 35887 imagens faciais de várias pessoas, retirado do *Kaggle*. Para dividir as imagens em "engaged" e "not\_engaged", foi utilizada a técnica de *clustering* e, em seguida, foi efetuada a divisão em dados de treino, teste e validação. Para formar o modelo de *Machine Learning* (ML), foi utilizada a *Visual Geometry Group (16 layers)* (VGG-16), uma arquitetura *Convolutional Neural Network* (CNN) bem sucedida. Para tornar este estudo o mais próximo da realidade, foi utilizado um conjunto de dados reais, composto por frames de vídeo de adultos com mais de 60 anos que participaram num programa de educação para a saúde do cérebro, para testar o modelo já desenvolvido e treinado.

Os resultados mostram que o modelo teve um bom desempenho quando treinado e testado com o primeiro *dataset*, obtendo valores acima de 90% para as métricas de *F1-Score*, *precision* e *recall*. Em contrapartida, quando testado com o segundo *dataset*, alcançou valores entre 40% e 50% nas métricas referidas.

A análise de conjuntos de dados variados e a colaboração com especialistas, como a neuropsicóloga Ana Rita Silva, enriqueceram a abordagem experimental. A discrepância nas métricas de avaliação entre os diferentes conjuntos de dados, realça a importância da seleção criteriosa de dados de teste. Estes resultados contribuem para o avanço no controlo e intervenção na saúde, direcionando futuros esforços para uma abordagem mais personalizada e eficaz. Com o projeto atual, que testa a colaboração entre humanos e a IA, pode ser possível alargar novos caminhos para o tratamento de dados e fazer previsões sobre a adesão a tratamentos/intervenções em doenças crónicas.

**Palavras-chave** – *Engagement*, Saúde, Demência, Expressões Faciais, Intervenção, Inteligência Artificial, Redes Neurais Convolucionais, *VGG16*, Classificação de Imagens, Detecção Automática, *F1-Score*, *Precision*, *Recall*, Doenças Crônicas

# Abstract

This study addresses a global healthcare challenge: maintaining the engagement of dementia patients in long-term care. Recognising the crucial importance of patient engagement for the success of these treatments, the aim is to identify effective solutions for monitoring and alerting to fluctuations in engagement over time. The increasing use of wearable techniques offers a promising approach to automatically detect such situations and provide timely interventions. This research focuses on developing strategies that leverage Artificial Intelligence to identify patterns of engagement through facial expressions, allowing healthcare professionals to act proactively and optimise patient care.

To this end, a dataset consisting of 3,5887 facial images of various people, taken from *Kaggle*, was initially used. To divide the images into "engaged" and "not engaged", the clustering technique was used, followed by division into training, test and validation data. To form the *Machine Learning* (ML) model, the *Visual Geometry Group (16 layers)* (VGG-16), a successful *Convolutional Neural Network* (CNN) architecture, was used. To make this study as close to reality as possible, a real dataset consisting of video frames of adults over 60 who took part in a brain health education programme was used to test the model already developed and trained.

The results show that the model performed well when trained and tested with the first dataset, obtaining values above 90% for the F1-Score, precision and recall metrics. On the other hand, when tested with the second dataset, it achieved values between 40% and 50% in the aforementioned metrics.

Analysing varied data sets and collaborating with experts such as neuropsychologist Ana Rita Silva enriched the experimental approach. The discrepancy in the evaluation metrics between the different data sets emphasises the importance of careful selection of test data. These results contribute to advances in health monitoring and intervention, directing future efforts towards a more personalised and effective approach. With the current project, which tests collaboration between humans and Artificial Intelligence, it may be possible to broaden new avenues for processing data and making predictions about adherence to treatments/interventions in chronic diseases.

**Keywords** – *Engagement*, Health, Dementia, Facial Expressions, Intervention, Artificial Intelligence, Convolutional Neural Networks, *VGG16*, Image Classification, Automatic Detection, *F1-Score*, *Precision*, *Recall*, Chronic Diseases



# Lista de Acrónimos

**AUC** *Area Under the Curve.* xx, 28, 36, 65

**BBO** *Biogeography-based Optimization.* 38

**CINEICC** *Center for Research in Neuropsychology and Cognitive and Behavioral Intervention.* iii, xxi, 1, 4, 43, 87

**CISUC** *Centre for Informatics and Systems of the University of Coimbra.* iii, 1

**CNN** *Convolutional Neural Network.* ix, xi, xv, xvi, 3, 4, 24, 25, 26, 38, 39, 40, 56

**ECG** *Eletrocardiograma.* 35, 36, 38

**EDA** *Atividade Eletrodérmica.* 35, 36

**EMFACS** *Emotional Facial Action Coding System.* 13

**EMG** *Eletromiografia.* 35

**FACS** *Facial Action Coding System.* 9, 13, 14

**IA** *Inteligência Artificial.* ix, 1, 2, 3, 31, 39, 43, 69, 70

**LSTM** *Long Short Term Memory.* 39

**ML** *Machine Learning.* ix, xi, 2, 40

**PCRN** *Parallelized Convolutional Recurrent Neural Network.* 38

**PSO** *Particle Swarm Optimization.* 38

**ROC** *Receiver Operating Characteristic.* xvi, xx, 28, 65, 66

**SGD** *Stochastic Gradient Descent.* 57

**SMOTE** *Synthetic Minority Over-sampling Technique.* xx, 30, 62, 68

**SVM** *Support Vector Machine.* 38, 39

**TPR** *True Positive Rate.* 27

**VGG-19** *Visual Geometry Group de 19 Camadas.* 25, 40

**VGG-16** *Visual Geometry Group (16 layers).* ix, xi, xv, xvi, xx, 2, 3, 24, 25, 26, 34, 40, 56, 67, 69

**WCSS** *Within-Cluster Sum of Square.* 51



# Lista de Figuras

2.1	Modelo OCC, proposto por <i>Ortony, Clore</i> e <i>Collins</i> [14], com a divisão entre os 3 principais tipos de emoção. . . . .	6
2.2	Modelo proposto por Russell [5] com a disposição circular de 8 conceitos afetivos. . . . .	8
2.3	Imagens das expressões faciais de dois elementos da tribo <i>Fore</i> [15]. . . . .	9
2.4	Alegria . . . . .	10
2.5	Raiva . . . . .	10
2.6	Medo . . . . .	10
2.7	Repugnância . . . . .	10
2.8	Surpresa . . . . .	10
2.9	Tristeza . . . . .	10
2.10	Desprezo . . . . .	10
2.11	As 7 emoções principais segundo Paul Ekman [16]. . . . .	10
2.12	Alegria . . . . .	13
2.13	Raiva . . . . .	13
2.14	Medo . . . . .	13
2.15	Repugnância . . . . .	13
2.16	Ódio . . . . .	13
2.17	Impotência . . . . .	13
2.18	Horror e agonia . . . . .	13
2.19	Alguns desenhos e gravuras que ilustram o livro de Charles Darwin [17]. . . . .	13
2.20	Reconhecimento facial da raiva [16]. . . . .	15
2.21	Reconhecimento facial do desprezo [16]. . . . .	15
2.22	Reconhecimento facial da repugnância [16]. . . . .	16
2.23	Reconhecimento facial da alegria [16]. . . . .	17
2.24	Reconhecimento facial da surpresa [16]. . . . .	18
2.25	Reconhecimento facial do medo [16]. . . . .	19
2.26	Diagrama da Arquitetura <i>Visual Geometry Group (16 layers)</i> (VGG-16): <i>Convolutional Neural Network</i> (CNN) para Classificação de Imagens [27]. . . . .	26
2.27	Esquema do método <i>K-Fold Cross Validation</i> [36]. . . . .	33
4.1	Imagens de cada emoção presente no primeiro <i>dataset</i> . . . . .	44
4.2	<i>Elbow Method</i> para obtenção do número ótimo de <i>clusters</i> . . . . .	51

4.3	(a) Imagens pertencentes ao <i>Cluster</i> 0 (b) Imagens pertencentes ao <i>Cluster</i> 1 (c) Imagens pertencentes ao <i>Cluster</i> 2 (d) Imagens pertencentes ao <i>Cluster</i> 3 (e) Imagens pertencentes ao <i>Cluster</i> 4 . . . . .	53
5.1	Diagrama da Arquitetura VGG-16 utilizada neste estudo: CNN para Classificação de Imagens. . . . .	56
5.2	Resultados de treino e validação do 1 <sup>o</sup> teste. . . . .	59
5.3	Resultados de treino e validação, após a aplicação de <i>K-Fold Cross Validation</i> e <i>Oversampling</i> . . . . .	64
5.4	Curva <i>Receiver Operating Characteristic</i> (ROC) após o segundo teste experimental. . . . .	66
5.5	Matriz de Confusão após o segundo teste experimental. . . . .	66
C.1	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com <i>dropout</i> de 25% e 50%. . . . .	89
C.2	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com <i>dropout</i> de 15% e 25%. . . . .	89
C.3	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com <i>dropout</i> de 5% e 15%. . . . .	89
D.1	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com regularização do peso <i>L1</i> , com 0.001. . . . .	91
D.2	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com regularização do peso <i>L1</i> , com 0.0001. . . . .	91
E.1	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com regularização do peso <i>L2</i> , com 0.001. . . . .	93
E.2	<i>Loss</i> de treino e validação ao longo das <i>epochs</i> , com regularização do peso <i>L2</i> , com 0.0001. . . . .	93

# Lista de Tabelas

2.1	Sintomas e sinais associados à demência. . . . .	21
4.1	Características do primeiro <i>dataset</i> usado. . . . .	44
4.2	Características do segundo <i>dataset</i> usado. . . . .	50
4.3	Número de imagens de cada classe do primeiro <i>dataset</i> usado. . . . .	53
4.4	Número de imagens de cada classe do segundo <i>dataset</i> usado. . . . .	54
5.1	Resultado das métricas do primeiro teste experimental realizado, com o modelo inicial. . . . .	58
5.2	Resultado das métricas do primeiro teste experimental realizado, com o modelo após a aplicação de <i>K-Fold Cross Validation</i> e <i>Oversampling</i> . . . . .	63
5.3	Resultado das métricas do segundo teste experimental realizado, com o modelo após a aplicação de <i>K-Fold Cross Validation</i> e <i>Oversampling</i> . . . . .	65
D.1	Resultado das métricas do primeiro teste experimental realizado, com o modelo após as duas aplicações da regularização do peso <i>L1</i> . . . . .	91
E.1	Resultado das métricas do primeiro teste experimental realizado, com o modelo após as duas aplicações da regularização do peso <i>L2</i> . . . . .	93



# Conteúdo

<b>Lista de Acrónimos</b>	<b>xiii</b>
<b>Lista de Figuras</b>	<b>xv</b>
<b>Lista de Tabelas</b>	<b>xvii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto do Projeto . . . . .	1
1.2 Motivação . . . . .	1
1.3 Objetivos e Metodologia . . . . .	2
1.4 Contribuições Científicas . . . . .	3
1.5 Organização do Documento . . . . .	4
<b>2 Fundamentos Teóricos</b>	<b>5</b>
2.1 Computação Afetiva . . . . .	5
2.1.1 Modelo OCC . . . . .	5
2.1.2 Modelo de Russell . . . . .	7
2.1.3 Modelo de Ekman . . . . .	8
2.1.4 Reconhecimento de emoções através da expressão facial . . . . .	10
2.1.4.1 Emoção da Raiva . . . . .	14
2.1.4.2 Emoção do Desprezo . . . . .	14
2.1.4.3 Emoção da Repugnância . . . . .	15
2.1.4.4 Emoção da Alegria . . . . .	16
2.1.4.5 Emoção da Surpresa . . . . .	17
2.1.4.6 Emoção do Medo . . . . .	18
2.2 Demência . . . . .	19
2.2.1 O que é? . . . . .	19
2.2.2 Sintomas, adversidades e limitações . . . . .	19
2.2.3 Fases e sinais de alerta . . . . .	20
2.3 <i>Engagement</i> . . . . .	20
2.3.1 Níveis de <i>Engagement</i> . . . . .	22
2.3.1.1 Nível Individual . . . . .	22
2.3.1.2 Nível Organizacional . . . . .	22
2.3.1.3 Nível Social . . . . .	22
2.4 Aprendizagem Ativa . . . . .	23

2.4.1	<i>Diversity Sampling</i> . . . . .	23
2.5	Modelo VGG-16 . . . . .	24
2.5.1	Arquitetura . . . . .	25
2.6	Ferramentas de Avaliação . . . . .	26
2.6.0.1	<i>Accuracy</i> . . . . .	27
2.6.0.2	<i>Precision, Recall e F1-score</i> . . . . .	27
2.6.0.3	Matriz de Confusão . . . . .	28
2.6.0.4	Curva ROC e <i>Area Under the Curve</i> (AUC) . . . . .	28
2.7	Irregularidades de Distribuição de Dados - Dados Não Balanceados . . . . .	29
2.7.1	<i>Undersampling</i> . . . . .	30
2.7.2	<i>Oversampling</i> . . . . .	30
2.7.2.1	<i>Synthetic Minority Over-sampling Technique</i> (SMOTE) . . . . .	30
2.8	Técnicas para Redução/Eliminação de <i>Overfitting</i> . . . . .	31
2.8.1	<i>Overfitting</i> . . . . .	31
2.8.2	<i>Dropout</i> . . . . .	31
2.8.3	Regularização de Pesos . . . . .	32
2.8.4	<i>K-Fold Cross Validation</i> . . . . .	33
<b>3</b>	<b>Trabalhos Relacionados</b>	<b>35</b>
<b>4</b>	<b><i>Datasets</i></b>	<b>43</b>
4.1	<i>Dataset 1</i> . . . . .	44
4.2	<i>Dataset 2</i> . . . . .	44
4.3	<i>Dataset de Engagement</i> para Treino . . . . .	50
4.3.1	<i>Clustering e Elbow Method</i> . . . . .	50
4.4	<i>Dataset de Engagement</i> para Teste . . . . .	54
<b>5</b>	<b>Experimentação: Métodos, Resultados e Discussão</b>	<b>55</b>
5.1	Divisão dos Dados . . . . .	55
5.2	Implementação e Configuração do VGG-16 . . . . .	56
5.2.1	Arquitetura Utilizada . . . . .	56
5.2.2	Pré-Processamento dos Dados . . . . .	56
5.2.3	<i>Early Stopping</i> . . . . .	57
5.2.4	Parâmetros de Treino . . . . .	57
5.2.4.1	<i>Adam</i> . . . . .	57
5.2.4.2	<i>Binary Crossentropy</i> . . . . .	57
5.3	Teste Experimental I . . . . .	57
5.3.1	Redução/Eliminação de <i>Overfitting</i> . . . . .	58
5.3.1.1	<i>Dropout</i> . . . . .	58
5.3.1.2	Regularização de Pesos . . . . .	60
5.3.1.2.1	Treino com Regularização <i>L1</i> . . . . .	60
5.3.1.2.2	Treino com Regularização <i>L2</i> . . . . .	61
5.3.1.3	<i>k-Fold Cross Validation</i> . . . . .	61

5.3.2	Tratamento de Dados Não Balanceados . . . . .	62
5.3.2.1	<i>Oversampling</i> . . . . .	62
5.4	Teste Experimental II . . . . .	63
<b>6</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>69</b>
	<b>Referências</b>	<b>71</b>
<b>A</b>	<b>Descrição do Programa Reminder</b>	<b>81</b>
<b>B</b>	<b>Póster apresentado no <i>IV Congresso Internacional do Center for Research in Neuropsychology and Cognitive and Behavioral Intervention (CINEICC)</i></b>	<b>87</b>
<b>C</b>	<b>Dropout para Redução/Eliminação de <i>Overfitting</i></b>	<b>89</b>
<b>D</b>	<b>Regularização de Pesos <i>L1</i> para Redução/Eliminação de <i>Overfitting</i></b>	<b>91</b>
<b>E</b>	<b>Regularização de Pesos <i>L2</i> para Redução/Eliminação de <i>Overfitting</i></b>	<b>93</b>



# Introdução

## 1.1 Contexto do Projeto

A presente tese foi desenvolvida a partir de uma parceria entre a Universidade de Coimbra e o *Center for Research in Neuropsychology and Cognitive and Behavioral Intervention* (CINEICC).

Inicialmente foram realizadas dezanove sessões online do programa *Reminder* [1, 2, 3] (e a última presencial), desenvolvido pela neuropsicóloga Ana Rita Silva, de onde surgiu o *dataset* usado para teste do modelo produzido. Este programa é uma intervenção não farmacológica de reabilitação neuropsicológica projetada para reduzir o risco de demência em pessoas com 60 anos ou mais, onde cada sessão, com duração aproximada de 60 minutos, abrange quatro elementos principais: relaxamento *mindfulness*, educação psicológica, atividades de partilha e tarefas para casa. A descrição pormenorizada do programa, conduzida pela neuropsicóloga responsável, está disponível no Anexo A. A elaboração do trabalho teve lugar nos *Centre for Informatics and Systems of the University of Coimbra* (CISUC).

## 1.2 Motivação

Existem no mundo bastantes problemas de saúde que requerem tratamentos a longo prazo e, por vezes, estes são difíceis de realizar por parte dos doentes. É geralmente reconhecido que o empenho dos doentes desempenha um papel fundamental no sucesso desses tratamentos. Infelizmente, muitas vezes, os doentes perdem a sua motivação, o seu interesse, o chamado *engagement*<sup>1</sup>, durante o período de tratamento. Assim, é muito importante detetar as flutuações do nível de empenho dos pacientes ao longo de toda a duração do tratamento, e produzir alertas quando este se afasta dos níveis normais, para que os profissionais possam atuar a tempo de tentar corrigir a situação. Infelizmente, essas situações não são usualmente comunicadas, ou porque os doentes não são capazes de as detetar ou porque, simplesmente, não o querem fazer. A solução poderá passar pela incorporação de técnicas de Inteligência Artificial (IA) [4] para que estas situações possam ser detetadas de forma automática e rápida [5].

---

<sup>1</sup>Neste trabalho, optou-se por utilizar a palavra *engagement* em inglês devido à sua abrangência e precisão em descrever conceitos que envolvem atenção, envolvimento, motivação e comprometimento. Enquanto que em português seria necessário empregar várias palavras para capturar a abundância de significados abordados pela palavra *engagement*, esta última oferece uma concisão, que torna a comunicação mais fácil e prática.

A ligação entre a IA e a saúde não é nova. Os benefícios da IA no domínio da saúde podem ser encontrados de várias formas [6]. Por exemplo, a IA pode ajudar a apoiar os profissionais de saúde no desempenho das suas funções. Os algoritmos podem analisar dados muito mais rapidamente do que os humanos, sugerir possíveis tratamentos, monitorizar a evolução de um doente e alertar o profissional humano para quaisquer preocupações. Em muitos casos, a IA e um clínico humano poderão trabalhar em conjunto, num processo de interação que se designa por *Collaborative-AI* [7].

Todas estas tarefas são possíveis uma vez que o mundo está a viver a revolução digital, a Era da Informação, com tecnologias como a Internet das Coisas e, por isso, estão disponíveis muitos dados digitais para treinar algoritmos de Aprendizagem Automática com sucesso. Os seres humanos têm dificuldades neste domínio. Estes dados alimentam os algoritmos de Aprendizagem Automática e, assim, é possível classificar e prever condições de saúde com uma precisão considerável [8].

A comunidade de IA está agora a ser capaz de recolher dados de redes sociais, como publicações no *Facebook*, e identificar, por exemplo, emoções. A tarefa envolve compilar mensagens e aplicar algoritmos de classificação de aprendizagem automática, com uma grande dose de técnicas de processamento de linguagem natural, para identificar marcadores linguísticos de emoções específicas. O mesmo processo pode ser utilizado para os marcadores de voz.

Os agentes de IA também podem olhar para fotografias, aplicar técnicas de reconhecimento facial (baseadas em algoritmos de aprendizagem profunda) e reconhecer marcadores visuais específicos nas expressões faciais das pessoas. Todas estas características foram reveladas e são aplicadas pelo *software* de IA. Os mesmos algoritmos que este tipo de inteligência utiliza com sucesso em muitas outras tarefas de classificação podem ser aplicados também no domínio da saúde. Para as máquinas, a saúde mental ou a diabetes, ou os problemas cardiovasculares, ou as estradas dos automóveis sem condutor, estão todos codificados em números. Por isso, sem surpresa, se a IA tem sucesso com os carros autónomos, também se espera um sucesso semelhante com os problemas de saúde.

### 1.3 Objetivos e Metodologia

Este estudo tem como principal objetivo detetar *engagement* em pessoas com demência, através da análise das suas expressões faciais, durante a fase de tratamento, uma vez que existe uma grande dificuldade, por parte dos profissionais de saúde, em conseguir que os seus pacientes mantenham interesse durante esta etapa.

Em primeiro lugar, foi desenvolvido um modelo de IA, capaz de detetar o *engagement* em imagens estáticas. Para este estudo foi utilizado, inicialmente, um *dataset*, retirado do *Kaggle*, constituído por 35887 imagens faciais de várias pessoas. De seguida, para dividir as imagens nas duas classes essenciais, *engaged* e *not engaged*, foi usada a técnica de *clustering* e, posteriormente, foi feita a divisão em dados de treino, teste e validação.

Para formar o modelo de *Machine Learning* (ML), foi aplicada a arquitetura *Visual Geome-*

*try Group (16 layers)* (VGG-16), um tipo de *Convolutional Neural Network* (CNN), constituída por uma camada de entrada, uma camada de saída e várias camadas ocultas que, no seu todo, tornam o VGG-16 o algoritmo com melhores resultados neste tipo de classificação (classificação de imagens) [9, 10].

Uma vez que o modelo desenvolvido sofria de alguns problemas de ajuste excessivo aos dados de treino, devido, possivelmente, a uma irregularidade de distribuição dos mesmos, foram aplicadas variadas técnicas para impedir que este perdesse a capacidade de generalizar para novos dados e, conseqüentemente, conduzisse a um desempenho insatisfatório.

Após o alcance do algoritmo com resultados mais favoráveis, com o objetivo de aproximar este estudo o mais possível da realidade, foram realizadas algumas sessões do programa *Reminder* (programa de educação para a saúde cerebral), de onde surgiu um *dataset* real, composto por frames de vídeos de adultos na casa dos 60 anos. Este novo conjunto de dados foi usado para testar o modelo já desenvolvido e treinado.

Em seguida, o modelo foi avaliado para determinar o seu sucesso na detecção de *engagement*, com base nas expressões faciais dos pacientes com (e sem) demência, através dos traços irradiados pelas suas expressões faciais.

## 1.4 Contribuições Científicas

As contribuições científicas fundamentais originadas deste estudo compreendem:

- um modelo de IA capaz de identificar o nível de *engagement*, através das expressões faciais, de uma maneira genérica;
- um modelo de IA capaz de identificar o nível de *engagement* de pacientes com demência, com base nas suas expressões faciais;
- a singularidade do segundo modelo desenvolvido, pois embora existam diversos modelos capazes de detetar o *engagement* por meio de expressões faciais, é notável que poucos estejam direcionados para a área da medicina, particularmente na neuropsicologia;
- o uso de *engagement* na saúde e o seu impacto, uma vez que possibilita a compreensão das respostas emocionais dos pacientes, inclusive daqueles mais introvertidos. Esse conhecimento, por sua vez, possibilita a adaptação e personalização dos tratamentos, resultando numa maior eficácia no atendimento médico;
- a criação de dois *datasets*: o primeiro desenvolvido com base num *dataset* existente e o segundo, construído com base nos dados da Consulta de Neuropsicologia, com anotação realizada por neuropsicólogos, mas que apresenta desafios significativos relacionados com a proteção de dados e privacidade dos pacientes;
- o potencial prático deste estudo, especialmente num contexto de aplicação real de campo, uma vez que foi testado com um *dataset* real. Após aprimoramentos no modelo, este

constituirá uma ferramenta valiosa para auxiliar profissionais de saúde a abordar o desafio do *engagement* nos seus pacientes.

Os resultados deste estudo foram apresentados em formato de póster (Anexo B) durante o *IV Congresso Internacional do CINEICC*, realizado na Universidade de Coimbra, em março de 2023. Além disso, planeia-se a elaboração de um artigo científico, com o objetivo de uma análise mais profunda do estudo realizado.

### 1.5 Organização do Documento

A restante parte deste presente documento está estruturada da seguinte forma: o Capítulo 2 fornece uma visão geral sobre os conteúdos base necessários para compreensão do estudo descrito nesta tese; no Capítulo 3 são apresentados os principais trabalhos relacionados com o presente trabalho, particularmente no que se refere à deteção de *engagement*, à computação afetiva e ao uso de CNN; o Capítulo 4 retrata a elaboração dos *datasets* utilizados para este estudo; o Capítulo 5 tem como principal objetivo a descrição das diferentes etapas deste trabalho, bem como a exposição e discussão dos resultados obtidos; por último, no Capítulo 6, são produzidas algumas ideias acerca das conclusões e limitações deste trabalho, tais como potenciais orientações para trabalhos futuros.

# Fundamentos Teóricos

## 2.1 Computação Afetiva

As emoções podem ser caracterizadas como sensações físicas ou afetivas sentidas pelo ser humano, provocadas por um determinado estímulo. Estas são sentidas de forma diferente por cada pessoa e, por isso, permitem que uma pessoa reaja de uma determinada forma diante de algum acontecimento, de maneira muito pessoal [11].

Tanto a nível psicológico como fisiológico, o conhecimento, a análise e a interpretação das mesmas tem sido tema de muitos estudos [12]. Sendo o século XXI um marco na evolução da tecnologia, diversos investigadores ambicionam aplicá-la ao estado emocional de cada indivíduo. Surge assim uma nova área de investigação: a Computação Afetiva [13].

A Computação Afetiva refere-se, então, à aprendizagem e desenvolvimento de métodos e dispositivos que são capazes de interpretar, reconhecer, estimular e processar as emoções humanas. Este tipo de investigação traduz uma simbiose entre diversos campos como a engenharia e as ciências da computação, com as ciências cognitivas, psicologia, neurociências, sociologia, ética, entre outros.

Até hoje não é possível quantificar um número exato de emoções sentidas pelos humanos, mas existem diversas teorias e modelos desenvolvidos por psicólogos e/ou investigadores, que serão abordadas nas próximas subsecções [14, 5, 15].

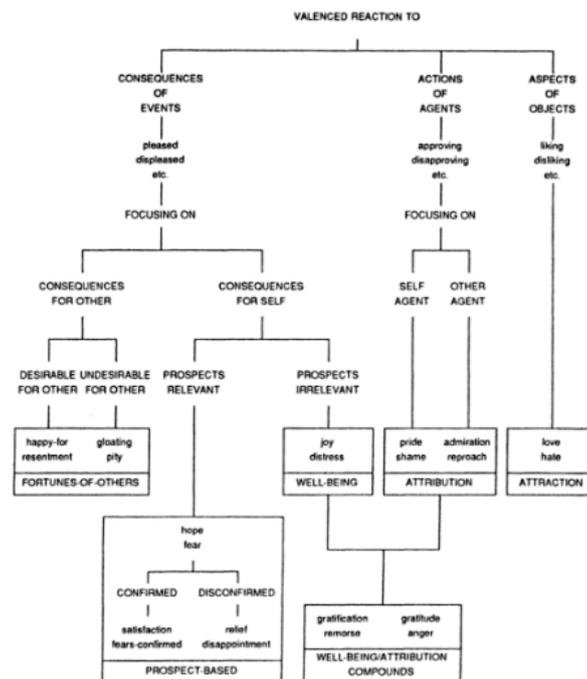
### 2.1.1 Modelo OCC

Baseado na teoria desenvolvida pelos três psicólogos Andrew Ortony, Gerald Clore e Allan Collins, o modelo OCC [14] defende que existem 22 tipos de emoções diferentes, que se distinguem, psicologicamente, pelas situações significativas que representam.

Na opinião de Ortony, Clore e Collins, as emoções que se focam em eventos podem ser diferenciadas das que se focam em ações e das que se concentram em objetos. Este modelo é então baseado no princípio da diferenciação entre comportamentos de valência positivos e negativos. Isto é, são atribuídas variáveis de maneira a determinar se o evento proporciona sensações positivas ou negativas para o sujeito em causa. Como descrito acima, esta ideologia reflete uma divisão em três categorias principais diferentes: emoções baseadas em eventos, emoções baseadas

em ações e emoções baseadas em objetos.

Fatores que dizem respeito ao próprio estado do indivíduo em causa (por exemplo, o estar triste), aos resultados de outrem (por exemplo, a piedade), aos resultados em termos de perspectiva (por exemplo, o medo) ou mesmo aos resultados já conhecidos (por exemplo, a tristeza), estão ligados às emoções que se focam em eventos. Como representado no esquema da 2.1, as emoções envolvidas com outros indivíduos são *Happy-For*, *Pity*, *Gloating* e *Ressentment*, enquanto que *Joy*, *Distress*, *Satisfaction*, *Fears-Confirmed*, *Relief* e *Desapointment*<sup>1</sup> correspondem a emoções relacionadas com o próprio sujeito. No que diz respeito às emoções que se concentram nas ações, estas abrangem, não só avaliações das ações como louváveis ou censuráveis (por exemplo, orgulho e vergonha, respetivamente), mas também se, dentro deste tipo de classificação, as ações representam o sujeito em questão (por exemplo, orgulho e vergonha) ou outra pessoa que esteja “do lado de fora” (por exemplo, admiração e censura). Este conjunto de emoções está representado no esquema da Figura 2.1 pelas emoções: *Admiration* e *Reproach*, no que diz respeito às emoções envolvidas com outros indivíduos e *Pride* e *Shame*, relacionadas com os próprios sujeitos. Em contraste com as emoções ligadas aos eventos e às ações, existem as emoções centradas nos objetos. Estas focam-se no quão cativante será o objeto, expressando um sentimento de, por exemplo, felicidade se o objeto fascinar e interessar, ou manifestando a emoção da repugnância, por exemplo, se o objeto for desinteressante e dececionante. Este tipo de classe é definida apenas por duas emoções: *Love* e *Hate*.



**Figura 2.1:** Modelo OCC, proposto por Ortony, Clore e Collins [14], com a divisão entre os 3 principais tipos de emoção.

<sup>1</sup>Optou-se por utilizar os termos em inglês, diretamente dos autores originais, a fim de preservar a integridade dos conceitos subjacentes, uma vez que as traduções diretas nem sempre capturam adequadamente a complexidade e nuances das emoções e sentimentos descritos.

Este modelo baseia-se, assim, em várias áreas de estudo como objetivos, desejos, expectativas, normas, preceitos, princípios, gostos e atitudes, ou seja, na componente cognitivo-interpretativa.

O OCC apoia a ideia de que a emoção começa com avaliações e não com *emotion words*. Isto é, cada pessoa avalia o mundo consoante as suas preocupações, variando estas de pessoa para pessoa, e relacionam-se com distintos tipos de representações, ou seja, diferentes objetivos, padrões e gostos. Estas representações podem ser classificadas em três tipos: motivacionais (se estiverem relacionadas com os objetivos), morais (se estiverem relacionadas com os padrões) e estéticas (se estiverem relacionadas com os gostos) [14].

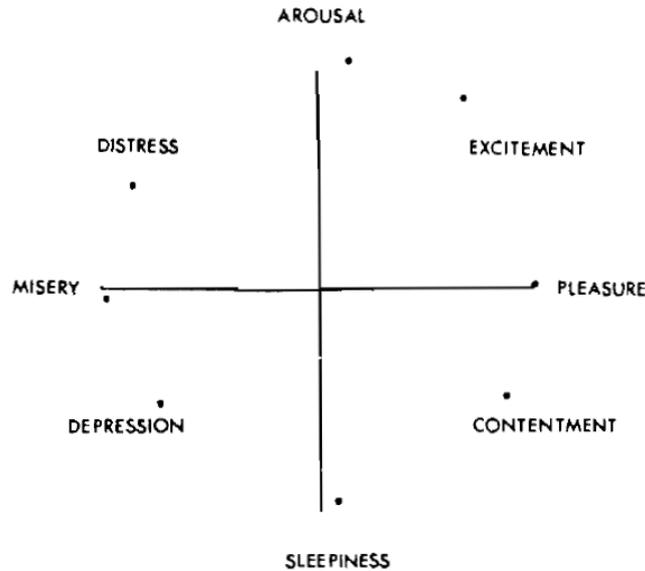
### 2.1.2 Modelo de Russell

Em contrapartida, alguns cientistas defendem que existem apenas duas propriedades que caracterizam as emoções, em qualquer situação: valência e excitação. Valência refere-se à maneira como o sujeito avalia uma determinada situação, conferindo-lhe o grau de agradável ou desagradável. Excitação diz respeito ao estado de entusiasmo ou ativação que um sujeito pode sentir em relação a um certo estímulo. Para além destes dois conceitos, acreditam que os contrastes entre as duas emoções negativas (baixa valência) e altamente “agitadas” (alta excitação) da raiva e do medo, surgem da interpretação de cada indivíduo sobre a realidade do acontecimento, e não de sentimentos emocionais.

Após algumas análises baseadas em estudos matemáticos, Russell [5] conclui que existem algumas emoções que se verificam em conjunto. Por exemplo, as emoções negativas do medo e da tristeza e as emoções positivas da alegria e da admiração. Isto pode induzir a ideia de que as emoções de valência semelhante tendem a acontecer em simultâneo, assim como as emoções idênticas em excitação.

Para esclarecer e clarificar melhor esta sua ideia, Russell criou um esquema em círculo, apresentado na Figura 2.2.

Observando-o é possível visualizar 8 variáveis que se localizam sobre um círculo, numa superfície bidimensional, de aparência semelhante aos pontos de uma bússola. A dimensão horizontal (este-oeste) representa, metaforicamente, o desgosto-prazer, e a dimensão vertical (norte-sul) a despertar/estimular-dormir/sossegar. Quanto às restantes quatro variáveis, estas não representam dimensões independentes, no entanto contribuem para a definição dos quadrantes espaciais. A depressão encontra-se num ponto sudoeste, estando esta entre o desgosto e o sossego. O oposto da depressão, a excitação, é definida como a queda num ponto no nordeste, ou seja, a conjugação ideal do alto prazer e da alta estimulação. O sofrimento e o seu oposto, o contentamento, formam uma dimensão bipolar semelhante, estando em pontos completamente opostos (noroeste e sudeste, respetivamente).



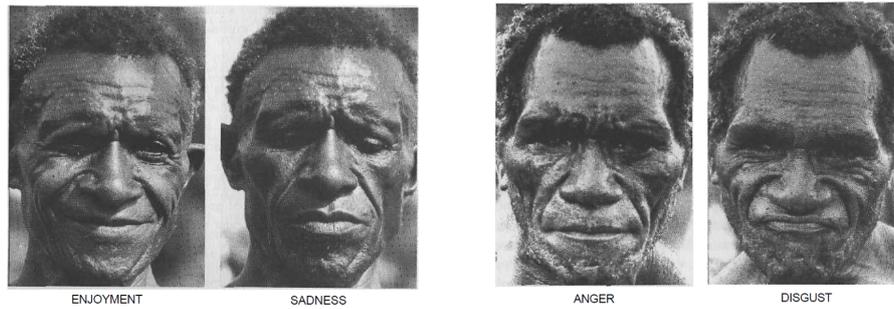
**Figura 2.2:** Modelo proposto por Russell [5] com a disposição circular de 8 conceitos afetivos.

### 2.1.3 Modelo de Ekman

Paul Ekman, considerado um dos mais notáveis psicólogos do século XX, dedicou a maior parte do seu tempo ao estudo das emoções e expressões faciais. Durante 40 anos viajou por diversos lugares tais como Papua-Nova Guiné, Estados Unidos, Japão, Brasil, Argentina, Indonésia e ex-União Soviética, lugares estes que conduziram as suas ideias a respeito da natureza da emoção, pois muito do que desenvolveu foi influenciado pelos seus estudos interculturais da expressão facial [16].

De entre os muitos estudos que realizou, um dos mais importantes desencadeou o fenómeno hoje conhecido como as “regras de exibição” ou os “costumes”. Esta investigação foi desenvolvida através da comparação da expressão emocional entre os cidadãos dos EUA e Japão, onde Ekman verificou que ambos exibiam as mesmas expressões faciais quando se encontravam sozinhos a visualizar os vídeos do estudo, mas os japoneses exibiam um sorriso enquanto mascaravam as emoções negativas do medo e do nojo. Estas “regras de exibição” referem-se às normas informais de uma cultura ou grupo social que diferenciam a maneira como o sujeito se deve expressar. Estes padrões culturais são, geralmente, instruídos numa idade jovem, através de interações e socializações com outras pessoas. Estas são, assim, compreendidas de forma distinta de cultura para cultura, de pessoa para pessoa, e são frequentemente usadas como forma de proteção da própria imagem ou da imagem de outrem.

Posteriormente, viajou para a Papua Nova-Guiné com o objetivo de desenvolver um outro estudo. Durante dias, fotografou as faces emocionais de uma tribo socialmente isolada (os Fore) e chegou à conclusão que existiam seis emoções básicas universais. Assim Ekman desenvolveu a teoria Neuro-Cultural das Emoções, que compreende aspetos universais e inatos, relativos não só ao funcionamento cerebral, mas também ao poder da cultura, em relação à flexibilidade e abertura de uma expressão emocional.



(a) Emoções de divertimento e tristeza, respetivamente.

(b) Emoções de raiva e repugnância, respetivamente.

**Figura 2.3:** Imagens das expressões faciais de dois elementos da tribo *Fore* [15].

Nos anos 70, Ekman desenvolveu uma ferramenta científica de medição da expressão facial *Facial Action Coding System* (FACS), que permite descrever todos os movimentos faciais em componentes individuais da ação muscular. Por outras palavras, possibilita a medição de qualquer expressão facial produzida pelo ser humano.

Na década de 80, após alguns estudos e pesquisas, a emoção do “desprezo” foi acrescentada à lista das emoções básicas universais.

Deste modo, e a partir da evolução na área das emoções feita por Paul Ekman, a maioria dos psicólogos já acredita que existirão sete emoções principais:

- A **alegria** 2.4, uma emoção positiva que provoca bem-estar e satisfação. Esta desencadeia sentimentos positivos e potencia a inibição de pensamentos negativos;
- A **raiva** 2.5, considerada a emoção mais perigosa, é muitas vezes associada ao medo e ao nojo, podendo conduzir ao desejo de prejudicar algo ou alguém;
- O **medo** 2.6, que apesar de ser considerado uma emoção negativa associada ao perigo e à ameaça, é essencial para a sobrevivência do ser humano, pois pode assumir um papel protetor ou defensor, conduzindo e ensinando a pessoa a reagir perante alguma situação deste tipo;
- A **repugnância** 2.7, que se manifesta perante situações repulsivas e é considerada uma emoção negativa. Pode estar relacionada com algo prejudicial ou desagradável para o sujeito em si, a nível físico e/ou psicológico;
- A **surpresa** 2.8, a emoção mais breve de todas, com uma duração máxima de apenas alguns segundos, sendo provocada por situações inesperadas;
- A **tristeza** 2.9, que pode ser vivida em diversas situações, como luto, insucesso profissional, perda de autoestima, rejeição, etc., tornando esta emoção bastante flexível, pois pode ser vivenciada, de forma mais ou menos intensa, em situações distintas;
- O **desprezo** 2.10, considerado um sentimento forte, traduz-se numa persistente desvalorização de uma pessoa ou instituição.



**Figura 2.4:** Alegria



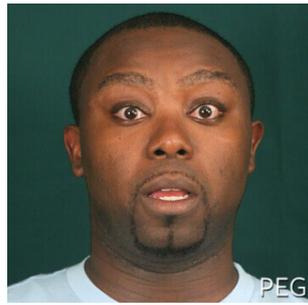
**Figura 2.5:** Raiva



**Figura 2.6:** Medo



**Figura 2.7:**  
pugnância



Re- **Figura 2.8:** Surpresa



**Figura 2.9:** Tristeza



**Figura 2.10:** Desprezo

**Figura 2.11:** As 7 emoções principais segundo Paul Ekman [16].

#### 2.1.4 Reconhecimento de emoções através da expressão facial

O reconhecimento de emoções favorece o comportamento social humano e, conseqüentemente, a sua adaptação. A percepção emocional, definida como a capacidade de identificar emoções em si mesmo e nos outros, pode ser identificada através das expressões faciais.

As expressões faciais representam o movimento de um ou mais músculos da face, convertendo-o num sinal de emoção. Estas são consideradas um dos principais meios de comunicação (comunicação não-verbal) entre os seres humanos, embora também sejam utilizadas por outros mamíferos.

Existem dois tipos de expressões faciais: as **voluntárias** e as **involuntárias**, e os mecanismos neuronais responsáveis pelo controlo dessas expressões diferem de tipo para tipo. As **expressões faciais voluntárias** são, na maioria das vezes, socialmente condicionadas e conduzidas por instruções enviadas ao cérebro. Por outro lado, as **expressões faciais involuntárias**

são classificadas como inatas e seguem um percurso subcortical no cérebro, como por exemplo, a expressão do choro.

Segundo Ekman, e apesar da codificação das sete emoções básicas universais, o rosto está preparado para transmitir mais de dez mil expressões e, quando se verifica uma alteração do rosto, o conjunto de sentimentos e emoções vai também alterar-se [16].

Embora Ekman tenha sido um dos grandes pioneiros das descobertas feitas em relação às emoções e expressões faciais, os estudos deste tipo de temática foram iniciados pelo naturalista britânico Charles Darwin. Na primavera de 1872, dedicou-se à escrita de um dos seus livros denominado de *The Expression of Emotions in Man and Animals*, onde descreve, detalhadamente, as diferentes expressões manifestadas pelos animais e pelos humanos mediante distintas situações comportamentais que observara, acompanhadas de Figuras, gravuras, desenhos e fotografias usadas para sustentar e enriquecer os seus argumentos. Aparentemente, o motivo que levou Darwin a querer iniciar esta investigação, deve-se ao facto de este ter lido a obra do anatomista Charles Bell, que reflete a ideia de que determinados músculos do homem existiam apenas para a expressão das emoções. Como Darwin não concordava com esta teoria, resolveu analisá-la [17].

Assim, com este estudo, Darwin provou que existem diversas expressões faciais que se manifestam de forma semelhante nos animais e nos humanos. Por exemplo, as expressões de mau-humor e a boa disposição de cães e cavalos; a tendência a manifestar vingança por parte de certos animais; a fúria e raiva arquitetadas por diversos animais; o simples amor e o carinho de um cão para com o seu dono no sofrimento relacionado com a morte; o sentimento de afeto materna das fêmeas de todas as espécies; a aflição dos macacos fêmeas aquando da perda dos seus filhos; o acolhimento de macacos órfãos por outros de diferentes bandos; a bondade de certas fêmeas de babuínos ao adotarem macacos de outras espécies, entre outras observações que revelam a aproximação dos humanos com os animais, no que toca à manifestação de emoções por meio de expressões faciais e, em específico, da movimentação dos músculos da face, nomeadamente das pálpebras franzidas e dos cantos da boca repuxados para trás (observados na expressão do riso e do choro de alguns macacos) [17].

No entanto, e embora tenha conseguido retirar algumas conclusões acerca das expressões faciais nos animais e nos homens, o naturalista inglês reconhecia que persistia uma grande dificuldade em conseguir estudar a origem dos hábitos de expressão dos humanos. Esta adversidade existia pois os movimentos dos músculos eram demasiado curtos, apresentando uma duração reduzida, o que dificultava a análise dos mesmos, e, também devido a isso, era complicado de detetar até que ponto as mudanças de traços faciais destacavam realmente certas emoções e estados de espírito. De maneira a ultrapassar essas dificuldades, sugeriu a observação de crianças, uma vez que estas exibem diversas emoções e mostram os seus sentimentos de forma descontrolada. Propôs também a análise das expressões de tribos que não tenham tido muito contacto com outras comunidades. Com estes procedimentos, considerou que se reduziria a hipótese de que o observador fosse guiado pela imaginação, proporcionando uma base mais segura para se conseguir descobrir a causa dos movimentos faciais, de uma forma mais geral [17].

À semelhança daquilo que Ekman estudou posteriormente, também *Darwin* dedicou algum tempo da sua vida ao estudo das comunidades nativas, de maneira a completar a sua análise das expressões faciais que começara anteriormente. Para isso, divulgou um folheto intitulado de “*Queries about expression*”, composto por 16 questões e enviou para pessoas que estavam em contacto com povos primitivos, em diferentes regiões do mundo. As questões foram as seguintes:

1. *Exprime-se a surpresa pelo arregalar dos olhos e da boca e pela elevação das sobrancelhas?*
2. *A vergonha produz enrubescimento, quando a cor da pele nos permite percebê-lo? Se sim, até onde desce pelo corpo?*
3. *Quando um homem está indignado ou desafiador, ele franze o cenho, mantém cabeça e corpo erguidos, apruma os ombros e cerra os punhos?*
4. *Quando se concentra ou tenta resolver algum problema, ele franze o cenho ou enruga a pele abaixo das pálpebras inferiores?*
5. *Quando abatido, desce os cantos da boca e eleva a extremidade interna das sobrancelhas pela ação desse músculo que os franceses apelidaram de “músculos de sofrimento”? Nesse estado, as sobrancelhas fazem-se levemente oblíquas, com um pequeno inchaço em sua extremidade medial; e o meio da testa fica enrugado, não toda a sua extensão, como quando se elevam as sobrancelhas exprimindo surpresa?*
6. *Quando satisfeito, brilham seus olhos, enruga-se a pele em volta destes e retraem-se os cantos da boca?*
7. *Quando um homem olha para outro com desprezo ou ironia, ergue-se o canto do lábio superior por sobre o canino do lado pelo qual ele o está encarando?*
8. *Pode uma expressão de obstinação e tenacidade ser reconhecida principalmente pela boca firmemente fechada, pelo cenho baixo e pelas sobrancelhas levemente franzidas?*
9. *O desdém é exprimido por uma leve protrusão dos lábios e discreta expiração com o nariz empinado?*
10. *Manifesta-se o nojo virando-se o lábio inferior para baixo e elevando-se levemente o lábio superior com uma súbita expiração, como um vomitar incipiente ou cuspir?*
11. *O medo extremo é expresso aproximadamente da mesma maneira que o fazem os europeus?*
12. *O riso pode chegar ao extremo de fazer com que lacrimejem os olhos?*
13. *Quando um homem quer demonstrar que não pode impedir algo ou que ele mesmo não consegue fazer alguma coisa, ele encolhe os ombros, vira para dentro os cotovelos e estende as mãos para fora com as palmas abertas; e as sobrancelhas são erguidas?*
14. *As crianças, quando emburradas, fazem bico ou protraem fortemente os lábios?*
15. *Expressões de culpa, malícia ou ciúme podem ser reconhecidas, ainda que eu não consiga defini-las?*

16. *Balança-se a cabeça verticalmente na afirmação e horizontalmente na negação*

Após a análise das respostas, levantou a hipótese de que algumas ações humanas, realizadas anteriormente de forma consciente, terem sido convertidas em ações reflexas, ou seja, em respostas rápidas e involuntárias, devido à força do hábito. Darwin acredita que a teoria que explica estes fenómenos é a da herança de caracteres adquiridos pelo uso e desuso. Estas ações passaram a ser executadas de forma inconsciente, através do funcionamento das células nervosas sensitivas que estimulam as células motoras. Defende que, por exemplo, o espirro e a tosse, foram adquiridos como ações reflexas através do hábito que outrora havia de expelir, de forma violenta, partículas incomodativas das vias aéreas e respiratórias [17].



**Figura**  
Alegria



**Figura**  
Raiva



**Figura**  
Medo



**Figura**  
Repugnância



**Figura**  
Ódio



**Figura**  
Impotência



**Figura**  
Horror e agonia

**Figura 2.19:** Alguns desenhos e gravuras que ilustram o livro de Charles Darwin [17].

Alguns anos após as descobertas de Darwin, Ekman, um apaixonado pela teoria evolucionista do naturalista britânico, continuou os seus estudos, conseguindo desenvolver- como dito anteriormente- um sistema de codificação facial, o FACS. O psicólogo criou este modelo pois achou fundamental estudar a fundo a fisiologia do rosto humano, uma vez que considerava que a expressão facial das emoções primárias era universal. Assim, baseando-se nos estudos do anatomista sueco Carl-Herman Hjortsjö, Ekman e Wallace Friesen, um colega, analisaram, durante 8 anos, a estrutura detalhada das unidades musculares responsáveis por todas as expressões faciais observáveis. Em 1978 nasceu então o FACS, o instrumento de observação do rosto humano mais completo do mundo. Posteriormente, foi também desenvolvido o *Emotional Facial Action Coding System* (EMFACS), uma versão mais reduzida do original FACS, com o reconhecimento de apenas as 7 emoções principais, explicadas acima. Estas duas ferramentas são ainda hoje

bastante usadas pela comunidade científica, uma vez que permitem uma codificação rigorosa de todas as movimentações dos músculos, durante as expressões faciais demonstradas. Estudos de marketing, investigações clínicas, criminais e a área cinematográfica também fazem uso deste tipo de técnica. Por exemplo, os filmes *Shrek* e *Toy Story* foram os primeiros a utilizarem o FACS, o que lhes permitiu conseguir apresentar as expressões faciais das personagens de uma forma mais realista [15].

Este sistema é composto por cinco códigos que correspondem aos comportamentos faciais e às alterações de aparência. São eles *Action Units (AUs)*, *Action Descriptors (ADs)*, *Moviments*, *Gross Behaviors Codes* e *Visibility Codes*. Aquando da análise facial por parte do FACS, a face humana é separada em face superior e face inferior e o período da expressão facial deve ser compreendido em 3 momentos distintos: *onset*, o início, *apex*, o ponto máximo, e *offset*, o término. Após estas divisões, inicia-se a codificação, isto é, a análise e a transcrição da expressão facial nos códigos FACS. Esses códigos, por sua vez, também são classificados em 5 níveis de intensidade: nível A, de intensidade baixa, nível B, de intensidade leve, nível C, de moderada intensidade, nível D, de intensidade forte e, por fim, nível E, correspondente à intensidade máxima [15].

### **Emoção da Raiva**

A raiva pertence ao conjunto das sete emoções principais, segundo Paul Ekman. Esta surge, por exemplo, quando há impedimento para perseguir um objetivo e/ou quando existe a manifestação de algum comportamento ligado à injustiça. No seu extremo, esta emoção pode ser considerada uma das mais perigosas, uma vez que apresenta uma forte conexão com a violência. Quando esta emoção está presente, é possível vivenciar outros sentimentos como o medo (de causar danos, à pessoa em causa ou a outrem) ou repugnância (pelo sujeito ou objeto que está a bloquear a pessoa em causa de seguir o seu objetivo) [16].

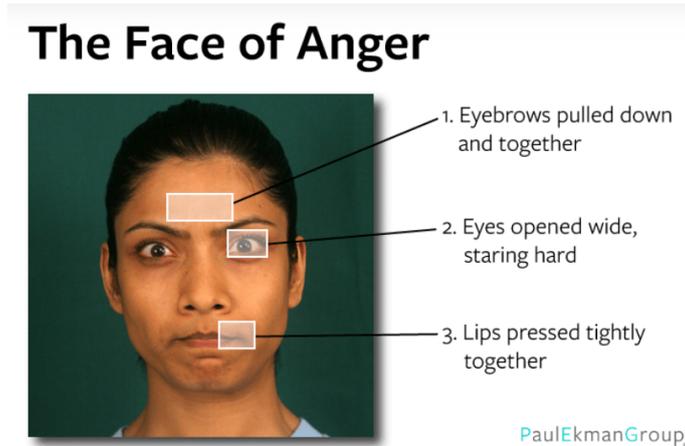
No momento em que esta emoção se manifesta, as sobrancelhas descem e ficam mais juntas uma da outra, os olhos brilham e os cantos dos lábios ficam mais estreitos.

Para além da expressão facial, a raiva pode ser identificada por outros fatores: pela expressão vocal, uma vez que, se não for controlada, pode gerar um rugido ou grito; pela postura, pois a maioria das pessoas inclina a cabeça e o queixo para a frente e enche o peito de ar para parecer maior e por sensações típicas, tais como, a sensação de calor, que confere um tom mais rosado à face, suor, cerrar a mandíbula e os punhos e tensão muscular.

### **Emoção do Desprezo**

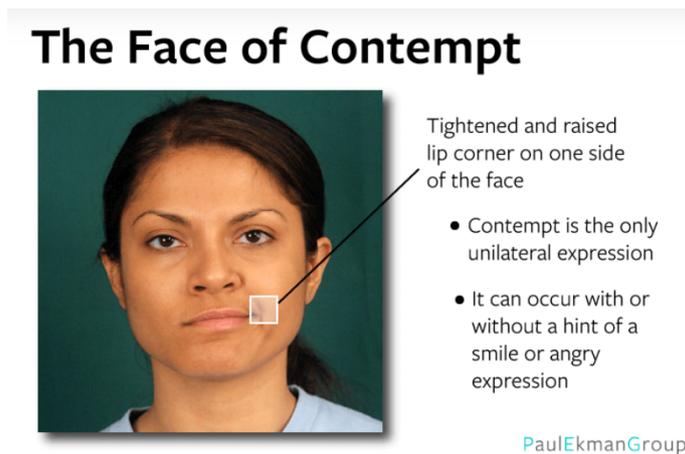
A emoção do desprezo é a única emoção que, quando sentida, apresenta uma expressão facial unilateral, ou seja, a expressão não é simétrica. Esta pode-se manifestar com ou sem um sorriso ou expressão de raiva [16].

O sinal mais comum ligado à expressão vocal do desprezo é o tom de voz presunçoso e, por vezes, até arrogante. Em relação às sensações associadas a este sentimento, estas podem



**Figura 2.20:** Reconhecimento facial da raiva [16].

ser bastante distintas, dependendo da situação em que o sujeito se encontra. Às vezes, pode ser sentida revelando sensações idênticas à raiva (tensão, calor, etc) ou mostrando sensações de alegria (excitação, agitação, etc). Quanto à postura, é habitual erguer o peito, com postura erguida, olhar “empinado” e revirar os olhos.



**Figura 2.21:** Reconhecimento facial do desprezo [16].

### Emoção da Repugnância

A emoção da repugnância surge como um sentimento de aversão a algo ofensivo. Esta emoção manifesta-se quando há uma sensação de repulsa por alguma coisa, que é possível de perceber fisicamente, pela visão, olfato, tato, audição e paladar, por ações ou aparências e até mesmo por ideias e ideologias [16].

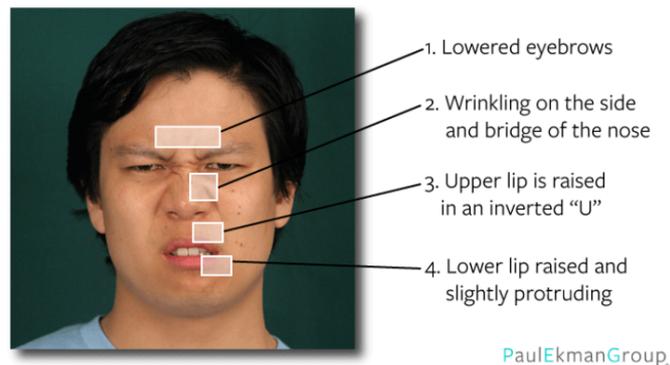
Este sentimento possui uma vasta gama de estados com variadas intensidades, que vão desde uma leve aversão até uma aversão intensa. Independentemente dos níveis de intensidade, todos os estados de repugnância são provocados pela sensação de que algo é repulsivo e até tóxico. Alguns estímulos que desencadeiam este sentimento são universais, tais como encontrar fluidos corporais ou a falta de higiene a vários níveis, ou culturais e individuais, como certos

tipos de comida ou repulsa interpessoal e social.

Este sentimento apresenta um grande desenvolvimento ao logo da vida. As crianças e os adolescentes possuem, muitas vezes, fascínio por diversas coisas e situações ligadas à repugnância, achando as mesmas engraçadas, provocando, por vezes, o riso. Isto acontece porque as crianças não se incomodam com a maioria das coisas que os adultos acham repugnantes, uma vez que ainda não possuem a capacidade cognitiva necessária para certas formas de repulsa apreendidas.

Para o reconhecimento facial desta emoção, a forma mais fácil é a de observar o tão conhecido franzir do nariz.

### The Face of Disgust



**Figura 2.22:** Reconhecimento facial da repugnância [16].

### Emoção da Alegria

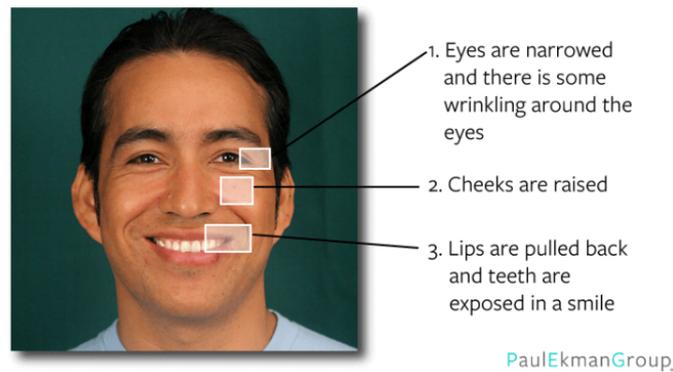
A alegria é a emoção mais desejável das sete emoções universais. Esta, geralmente resultante da conexão ou prazer sensorial, refere-se ao bem-estar, à felicidade, entre outros, e é usada para descrever um conjunto de estados de prazer relacionados, que vão desde a paz até ao êxtase [16].

Este sentimento advém de variados estímulos, entre os quais, o prazer derivado de um dos cinco sentidos (tato, paladar, olfato, visão e audição), de presenciar ou colaborar em atos de gentileza e compaixão, de testemunhar ou participar em algo engraçado, divertido, bonito e/ou surpreendente, etc.

Relativamente aos traços faciais que apontam para esta emoção, o sorriso é a expressão facial universal que melhor a sinaliza. Porém, existem vários sorrisos que são revelados quando não se está perante a emoção da felicidade. Por exemplo, quando há a expressão de um sorriso que serve de máscara para esconder uma outra emoção. O verdadeiro sorriso de prazer denomina-se de “sorriso de Duchenne” e pode ser identificado pela observação dos conhecidos “pés de galinha”.

Para além dos traços observáveis na face, esta emoção também pode ser identificada por expressões vocais, tais como guinchos de alegria, gritos excitados e suspiros de contentamento. Estas características são, geralmente, acompanhadas de algumas sensações, nomeadamente as

## The Face of Happiness



**Figura 2.23:** Reconhecimento facial da alegria [16].

sensações de leveza, energia, vibração e calor. Quanto à postura relacionada com a alegria, esta depende do estado e nível de felicidade, podendo ir de imóvel e relaxada a erguida e elevada.

Esta emoção é considerada bastante importante, pois desempenha um papel considerável em garantir que nenhum indivíduo constitui uma ameaça para os demais, assegurando amizade. Estudos revelam que as emoções consideradas positivas e agradáveis, normalmente, motivam o ser humano a realizar ações e a ter comportamentos classificados como bons, tanto para o sujeito em questão, como para a sobrevivência da nossa espécie.

### Emoção da Surpresa

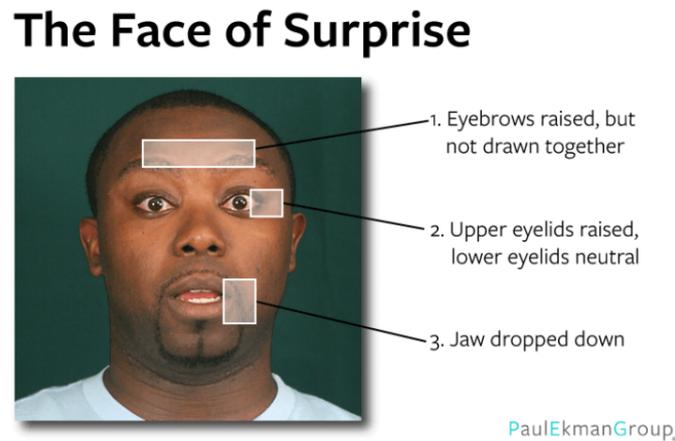
A emoção que é abordada nesta subsecção nasce quando surgem eventos inesperados e súbitos. Caracterizada como a emoção mais breve, a sua função é concentrar a atenção de cada indivíduo na percepção daquilo que está a acontecer e na determinação do nível de perigo ou segurança [16].

A surpresa é considerada a mais efémera das sete emoções universais pois dura, no máximo, alguns segundos. Esta termina quando surge a percepção daquilo que está a acontecer. A partir desse momento, este sentimento pode ser convertido noutras emoções, como medo, alívio, raiva, desgosto, etc, consoante a causa da surpresa. Pode surgir pela audição de sons altos (por exemplo, um balão a rebentar ou uma mota a arrancar) ou movimentos inesperados (por exemplo, alguém se aproximar de repente).

Quanto ao reconhecimento facial, o medo e a surpresa são duas das emoções cujas expressões faciais são mais facilmente confundidas, pois espelham as mesmas características principais: olhos, boca e sobrancelhas. No entanto, existem algumas diferenças. Na surpresa, as pálpebras superiores e os maxilares ficam mais relaxados, enquanto que as sobrancelhas se levantam, apresentando uma forma mais curvilínea do que na expressão do medo.

Para muitos psicólogos, como Andrew Ortony, Gerald Clore, Allan Collins [14] e Rainer Reisenzein [18], a surpresa não é considerada uma emoção pois, no seu ponto de vista, uma emoção deve ter valência intrínseca, ou seja, deve ser positiva ou negativa, consoante a situação

e o evento que a desencadeia. No entanto, uma vez que o caso em estudo nesta tese se baseia na teoria de Paul Ekman [16], considerar-se-á a surpresa como uma emoção, sendo esta neutra, dado que não apresenta conotação positiva ou negativa.



**Figura 2.24:** Reconhecimento facial da surpresa [16].

### Emoção do Medo

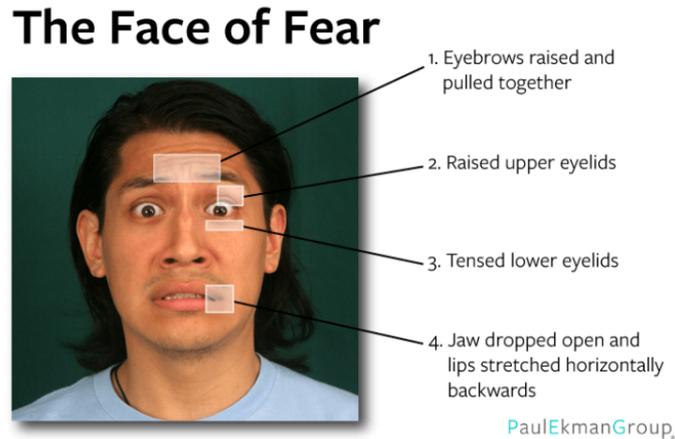
O medo, tradicionalmente considerado um sentimento com uma conotação negativa, desempenha um papel bastante importante na vida no ser humano, pois ajuda a manter a segurança e a lidar com o potencial perigo. Este surge aquando de ameaças de dano, seja ele físico, emocional ou psicológico, real ou imaginário [16].

Esta emoção pode ser fragmentada em 3 fatores: intensidade, que revela o quão severo é o dano resultante da ameaça, momento, que caracteriza o dano como imediato ou iminente e o confronto, que aponta para as ações que podem ser tomadas para reduzir ou eliminar a ameaça. O medo pode ser diminuído ou intensificado, dependendo da capacidade que o sujeito possui para lidar, ou não, com a ameaça. Este surge muitas vezes após a surpresa e oscila com a experiência da raiva.

Embora a sensação de medo difira de pessoa para pessoa, é possível ganhar medo de quase tudo. São exemplos de estímulos desta emoção a escuridão ou perda de visibilidade, alturas, rejeição e interação social, alguns animais (sobras, aranhas, roedores) e, principalmente, a morte.

Como relatado na subsecção 2.1.4.5, a expressão facial do medo é muitas vezes confundida com a da surpresa. Na expressão do medo, as sobrancelhas são mais retas e horizontais, enquanto que na da surpresa são mais levantadas e curvilíneas. Quanto à pálpebra superior, esta encontra-se mais levantada na presença do medo, mostrando a esclera, a parte branca do olho. No que diz respeito aos lábios, estes aparecem mais tensos e esticados.

A voz apresenta, geralmente, um tom mais tenso e alto, podendo também surgir gritos. As sensações mais usuais passam pela sensação de frio e falta de ar, suor, tremores e contração de alguns músculos, especialmente dos braços e das pernas. Quanto à postura, esta pode ser mobilizadora ou imobilizadora, deixando a pessoa parada ou com vontade de se afastar.



**Figura 2.25:** Reconhecimento facial do medo [16].

Embora esta emoção seja considerada negativa, e algumas pessoas a achem quase intolerável, há quem sinta prazer em sentir medo e procure vivenciar esta emoção, assistindo, por exemplo, a filmes de terror.

## 2.2 Demência

### 2.2.1 O que é?

A demência define-se como uma síndrome provocada por uma doença cerebral, habitualmente crónica e gradual, descrita por uma decadência cognitiva relevante, essencialmente ao nível de funções mentais, tais como o pensamento, orientação, linguagem, cálculo, julgamento e, a mais comum, a memória. Nestas circunstâncias, o sujeito em causa perde, usualmente, a autonomia e a independência, o que leva a consequências consideráveis na vida de familiares e cuidadores [19].

Existem diferentes formas de demência que decorrem de manifestações clínicas distintas, sendo as mais usuais a doença de Alzheimer (atingindo 60/70% dos casos), a demência de corpos de Lewy, a demência frontotemporal e a demência vascular.

Esta doença constitui uma das principais causas de incapacidade no conjunto da população mais envelhecida, embora esta não afete apenas pessoas idosas. Atualmente, existe uma grande falta de conhecimento acerca das características desta tão falada síndrome, o que origina, muitas das vezes, a estigmatização do paciente e o aparecimento de obstáculos na busca de apoio clínico.

### 2.2.2 Sintomas, adversidades e limitações

Consoante as características de cada indivíduo e a doença de base, a demência consegue afetar cada pessoa de maneira distinta. Isto é, o dano cerebral não define sozinho a maneira como a demência afeta alguém, pois existem outros fatores que influenciam neste aspeto, como por exemplo, as experiências da vida, a personalidade do sujeito em questão e o ambiente onde este se insere [19].

Em relação aos sintomas, estes podem variar desde uma simples e ínfima perda de memória a consideráveis dificuldades na comunicação e tomada de decisões coerentes e racionais, que podem conduzir à impossibilidade de realização de tarefas diárias sem apoio de terceiros. De um modo geral, esta doença está relacionada com um declínio gradual da cognição, acompanhado de indicativos psicológicos, tais como ansiedade, delírio, depressão e alucinações, e mudanças a nível do comportamento, como agressão, apatia e agitação.

### 2.2.3 Fases e sinais de alerta

Frequentemente, esta doença pode ser dividida em três fases, consoante o desenvolvimento da mesma. Todavia, os sintomas e sinais de alerta podem variar de pessoa para pessoa [19].

Numa **fase inicial**, os indicativos de demência são usualmente desvalorizados, passando, muitas vezes, despercebidos. Isto acontece pois os problemas de memória (impedimento na realização de novas atividades e dificuldade em recordar acontecimentos recentes) são tipicamente associados ao respetivo processo de envelhecimento. Ainda que esta fase seja considerada ligeira, podem manifestar-se dificuldades ao nível da orientação espacial (como por exemplo, perder-se numa rua que é já conhecida da pessoa em questão), orientação temporal (como por exemplo, confundir o dia da semana) e na toma de decisões.

À medida que a demência avança para a **fase intermédia (ou moderada)**, os sinais de existência da mesma tornam-se mais claros e evidentes. Podem então surgir dificuldades ao nível da resolução de problemas, comunicação, reconhecimento dos outros, recordação do nome de pessoas próximas e negligência com higiene ou alimentação.

Quando existe uma grande dificuldade no reconhecimento dos familiares e amigos próximos, surgimento de comportamentos perturbadores e angustiantes para o indivíduo em questão mas, principalmente, para as pessoas à sua volta, dependência diária total e dificuldade em andar, pode confirmar-se a presença da última fase desta doença, a **fase avançada**. Esta é caracterizada, essencialmente, por um desgaste severo das funções mentais, inatividade e dependência.

Na tabela 2.1 estão resumidos alguns dos sintomas e sinais mais frequentes em cada uma das fases desta doença.

## 2.3 *Engagement*

O *engagement* [20] pode definir-se como um conceito relacional pluridimensional, composto por atributos ligados à psicologia e características comportamentais de conexão, interação, participação e envolvimento, destinados a atingir resultados a nível individual, organizacional ou social.

O *engagement* é bastante utilizado no mundo empresarial, especialmente no relacionamento com o cliente. Este é chamado de *Customer Engagement* e é definido pelo conjunto de processos e sistemas, utilizados com a finalidade de adquirir informações que permitam a adequada administração da relação com o público-alvo, de maneira a promover a atração de novos cli-

<b>Fase Inicial</b>	<b>Fase Moderada</b>
Apatia e menor vivacidade	Esquecimento fácil de acontecimentos recentes
Menor interesse nos passatempos e atividades	Confusão em relação a tempo e espaço
Hesitação em fazer novas coisas	Esquecimento do nome de pessoa próximas
Incapacidade em adaptar-se às mudanças	Esquecimento de panelas no fogão ligado
Discernimento reduzido	Comportamentos inadequados
Mais tempo para compreensão ou realização de trabalhos de rotina	Ver ou ouvir coisas que não existem
Maior egocentrismo	Negligência com higiene ou alimentação
Esquecimento de detalhes e acontecimentos recentes	Angústia, frustração
Maior irritabilidade	-
Perda de objetos	-

<b>Fase Avançada</b>
Incapacidade para se lembrar de situações ocorridas há poucos minutos
Baixa capacidade de compreensão e utilização da linguagem
Incontinência
Não reconhecimento de amigos e familiares
Dependência para comer e para higiene
Incapacidade de reconhecer objetos do cotidiano
Inquietude
Agressividade
Dificuldade em andar
Movimentos incontrolados

**Tabela 2.1:** Sintomas e sinais associados à demência.

entes. No que diz respeito à parte cognitiva, o *engagement* é considerado relevante, uma vez que poderá fornecer conclusões importantes acerca do desempenho dos trabalhadores, como por exemplo, que os funcionários empenhados desempenham um melhor papel no seu trabalho ou que os consumidores empenhados mais facilmente aprovam um produto [20].

### 2.3.1 Níveis de *Engagement*

#### Nível Individual

O *engagement* a nível individual pode dividir-se em três domínios: cognitivo, afetivo e comportamental.

O *engagement* cognitivo está ligado ao esforço e dedicação para compreender e adquirir competências complexas. Baseia-se no compromisso de um indivíduo no que toca à atenção e desenvolvimento da compreensão sobre um tema ou ideia. Um sujeito que mostre elevados níveis de conhecimento, atenção e interesse (pessoal ou situacional) num determinado tópico, apresenta ideais de *engagement* cognitivo.

Todas as reações emocionais, tanto positivas como negativas, tais como o medo, a raiva, o apoio, etc, pertencem ao *engagement* afetivo. Geralmente este tipo de *engagement* conduz a estados de motivação, interesse ou preocupação, tendo em conta o tipo de reação.

O último domínio de *engagement* individual a ser abordado é o *engagement* comportamental. Este nível é frequentemente equiparado à interação ou à experiência única, como colocar um simples “gosto” no *Facebook* ou visitar uma página *Web*. Assim, este compromisso incorpora conceitos de interação, ação, colaboração e participação (intencional e não intencional)[20].

#### Nível Organizacional

A nível organizacional o *engagement* é considerado como um meio para permitir alcançar diversos objetivos, relacionais e organizacionais, promovendo resultados vinculados a objetivos financeiros e sociais.

#### Nível Social

O *engagement*, num contexto social, simboliza o envolvimento baseado na organização, no grupo e na comunidade. As diversas interpretações acerca deste tipo de compromisso concentram-se na ideia de que este representa um processo, onde o conceito é formulado através da comunicação e da ação.

Este nível de *engagement* pode ser definido como “um estado coletivo de envolvimento que pode ser representado por formas comportamentais (ação coletiva, participação de grupo), cognitivas (conhecimento partilhado) e afetivas (orientação, intenção e experiência) e é o resultado de um sistema ou processos socialmente situados de forma dinâmica.”[20]

No que diz respeito às medidas comportamentais, a ação coletiva é estimulada por valores partilhados e pela identidade comunitária acerca de uma determinada temática de envolvimento.

Esta visão partilhada no meio do cenário social, advém de processos coletivos desenvolvidos pelos pontos de vista individuais. Já a participação envolve a comunidade, como membros ativos, no desenvolvimento coletivo de significados e ajuste de soluções, através do diálogo com a organização focal.

Relativamente às formas afetivas, a orientação define-se como uma ideologia coletiva de crenças e comportamentos, que conduz a um condensação de significados e valores, que progridem do diálogo, interação e ligação com variadas perspectivas das partes interessadas. A intenção, determinada por uma atitude previamente existente, reflete motivações intrínsecas e um desejo individual de efetuar uma ação. A experiência representa uma ligação ou interação, antecedentes e/ou uma consequência de um compromisso.

## 2.4 Aprendizagem Ativa

A Aprendizagem Ativa, do inglês *Active Learning*, é uma abordagem no campo do *Machine Learning* em que o modelo aprende ativamente, escolhendo os exemplos de treino mais úteis. Este treino ativo permite que o modelo solicite rótulos adicionais para exemplos mais difíceis ou incertos, ao contrário do treino tradicional, onde todos os exemplos de treino são fornecidos de uma só vez.

O objetivo deste tipo de aprendizagem é maximizar o desempenho do modelo, usando o mínimo de exemplos de treino rotulados possíveis. Isto é particularmente útil em situações em que o processo de classificação de dados é caro ou demorado, pois permite que o modelo seja treinado com eficiência, concentrando-se apenas nos exemplos mais pertinentes [21, 22].

O *Active Learning* oferece uma variedade de métodos de amostragem para selecionar exemplos informativos. Alguns exemplos de métodos usados frequentemente incluem:

- **Incerteza:** seleciona exemplos em que as previsões do modelo não são confiáveis.
- **Diversidade:** para cobrir um espectro mais amplo de informações, seleciona exemplos que representam diferentes regiões ou conceitos no espaço de características.
- **Informação esperada:** considera os custos ou benefícios associados à rotulagem de exemplos particulares. Por exemplo, é possível escolher exemplos que resultarão numa redução significativa na incerteza do modelo.

Esta técnica pode ser usada em diversos contextos, incluindo deteção de objetos, classificação, segmentação de imagens e processamento de linguagem natural. O *Active Learning* tem a capacidade de aumentar significativamente o desempenho do modelo, mesmo quando os exemplos de treino são escolhidos de forma inteligente [21].

### 2.4.1 *Diversity Sampling*

O *Diversity Sampling* é uma abordagem utilizada no *Active Learning* para selecionar as instâncias de um conjunto de dados que são mais informativas e representativas. Esta abordagem

leva em consideração a diversidade das amostras selecionadas, enquanto que o *Active Learning* se concentra apenas na escolha das amostras mais relevantes para a classificação. Por outras palavras, este método procura uma seleção variada, que abrange várias áreas e aspetos do espaço de características, em vez de selecionar apenas amostras mais difíceis ou incertas para o modelo. Isto evita a tendência de se concentrar apenas numa parte do conjunto de dados e permite que o modelo aprenda de forma abrangente.

O *Diversity Sampling* é útil quando o conjunto de dados contém uma grande quantidade de exemplos semelhantes e é desejável obter uma representação mais abrangente do espaço de características. Ao diversificar a amostra de treino, esta abordagem pode fazer com que o processo de classificação seja mais eficiente e eficaz, o que pode resultar num modelo mais forte e generalizável [22, 23].

No âmbito deste estudo, aplicou-se a técnica de *Active Learning*, mais especificamente o *Diversity Sampling*. Esta abordagem assume um papel crucial no processo de obtenção de dados rotulados para treinar e aprimorar o modelo.

Após ter sido recolhido um conjunto de imagens das expressões faciais dos pacientes durante os tratamentos, uma parte dessas imagens foi submetida à avaliação de um perito, responsável por atribuir a cada uma delas a *label* de “engaged” ou “not engaged”. Estas avaliações foram comparadas com as previsões geradas pelo modelo desenvolvido e, desta forma, pretende-se avaliar o desempenho e a capacidade do algoritmo em fornecer informações semelhantes às do perito.

O uso do *Diversity Sampling* neste contexto, garante uma seleção diversificada de amostras para aplicação de rótulos (*labels*), evitando tendências e permitindo que o modelo seja treinado de forma abrangente e representativa dos diferentes aspetos do *engagement*, durante os tratamentos.

### 2.5 Modelo *Visual Geometry Group (16 layers) (VGG-16)*

Na presente subsecção, será apresentada, em detalhe, a arquitetura VGG-16, uma *Convolutional Neural Network* (CNN) amplamente utilizada no campo da visão computacional. No âmbito deste estudo, a escolha deste modelo foi motivada pela sua comprovada eficácia na extração de características visuais complexas a partir de imagens [9, 10, 24]. Considerando a natureza do problema abordado, a deteção de *engagement* em pacientes com demência por meio da análise das suas expressões faciais, a VGG-16 demonstra uma opção robusta e confiável. Através da sua estrutura de camadas convolucionais profundas e arquitetura bem definida, esta estrutura é capaz de capturar pormenores nas expressões faciais e aprender representações de alto nível, permitindo uma deteção mais precisa e sensível ao contexto emocional dos pacientes.

Praderdorfer e Kampel [9] estudaram a comparação entre a arquitetura VGG, a *Inception* e a *ResNet*, para o problema do reconhecimento facial. Embora todas tenham contribuído significativamente para o avanço do campo da visão computacional, alcançando resultados impressionantes em tarefas como deteção de objetos, segmentação semântica mas principalmente,

classificação de imagens, a que alcançou melhores resultados foi a arquitetura *VGG*, com exatidão de 72,7%, seguida da *ResNet* com 72,4% e a *Inception* com 71,6%.

Também Abdulsattar e Hussain [10] analisaram o desenvolvimento de algumas arquiteturas populares no reconhecimento facial como *MobileNet*, VGG-16, *Visual Geometry Group* de 19 Camadas (VGG-19), *DenseNet121*, *Inception-v3* e *Xception*, chegando à conclusão que o modelo VGG-16 apresenta maior precisão, de 64,46%.

### 2.5.1 Arquitetura

A arquitetura VGG-16 é um modelo de CNN amplamente utilizado na área de visão computacional. Foi proposto pela *Visual Geometry Group (VGG)* da Universidade de Oxford e foi um dos primeiros modelos a alcançar um bom desempenho no tema do processamento de imagem [25, 26].

Este algoritmo de detecção e classificação de objetos, pode classificar 1000 imagens de 1000 categorias diferentes com uma precisão de 92,7%. É fácil de usar com *Transfer Learning* e é um dos algoritmos mais populares para classificação de imagens.

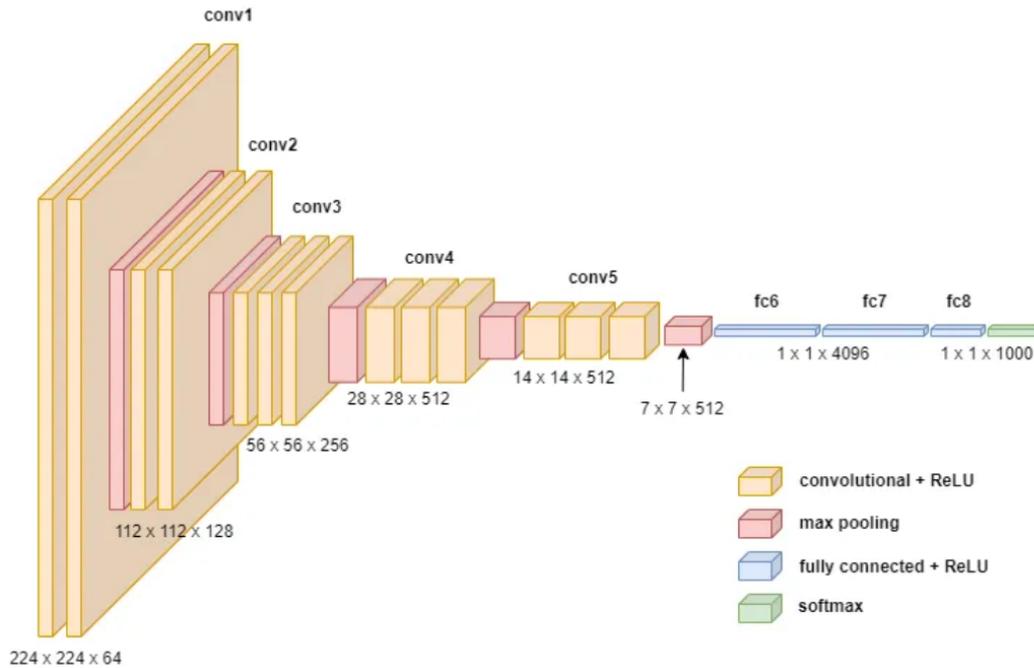
A VGG-16 é composta por 13 camadas de convolução, 5 de *pooling*, seguidas de 3 camadas totalmente conectadas. No total são 21 camadas, mas apenas 16 são camadas de peso, com parâmetros ajustáveis e treináveis. A arquitetura utiliza filtros de convolução pequenos (3x3 pixels) com um *stride* de 1 *pixel* e mantém um tamanho de filtro constante em todas as camadas convolucionais. A camada de *pooling* é realizada com filtros de *max pooling* de 2x2 pixels com um *stride* de 2 pixels [25, 26].

A estrutura detalhada da VGG-16 é a seguinte:

- **Camadas de convolução:** A rede começa com duas camadas convolucionais, cada uma seguida por uma função de ativação ReLU (*Rectified Linear Unit*) para introduzir a não-linearidade.
- **Camadas de *pooling*:** Após cada par de camadas convolucionais, há uma camada de *pooling* que reduz a resolução espacial para metade.
- **Camadas totalmente conectadas:** Após as camadas de convolução e *pooling*, há três camadas totalmente conectadas, cada uma seguida por uma função de ativação ReLU. A última camada totalmente conectada é uma camada *softmax*, que é usada para a classificação [25, 26].

A Figura 2.26 apresenta a arquitetura total da rede neuronal da VGG-16. Cada caixa representa uma camada específica com informações sobre o tipo de camada, tamanho do filtro, número de filtros e funções de ativação utilizadas.

A VGG-16 é conhecida por ter uma arquitetura profunda, o que significa que possui um grande número de parâmetros. Esta foi treinada inicialmente no conjunto de dados *ImageNet*, que possui milhões de imagens rotuladas em várias categorias. A VGG-16 alcançou resulta-



**Figura 2.26:** Diagrama da Arquitetura VGG-16: CNN para Classificação de Imagens [27].

dos impressionantes em competições de reconhecimento de imagens, superando muitos métodos anteriores [28].

No entanto, existem algumas desvantagens como o elevado poder computacional requerido para treino e uso, devido ao elevado peso.

Apesar disso, a VGG-16 ainda é amplamente estudada e usada como base para muitas aplicações de visão computacional e *transfer learning*, onde os conhecimentos aprendidos pela rede em grandes conjuntos de dados são aproveitados para tarefas específicas em conjuntos de dados menores.

## 2.6 Ferramentas de Avaliação

Ao realizar uma classificação binária, é importante entender como avaliar a precisão e o desempenho do modelo. Nesta subsecção serão abordados os métodos e ferramentas de avaliação selecionados para avaliar este estudo. Existem quatro resultados principais que podem ocorrer ao comparar as previsões do modelo com os valores reais. Esses resultados são comumente conhecidos como: **Verdadeiros Positivos (TP)** (amostras corretamente previstas como positivas), **Falsos Positivos (FP)** (amostras previstas como positivas, mas que na realidade são negativas), **Verdadeiros Negativos (TN)** (amostras corretamente previstas como negativas) e **Falsos Negativos (FN)** (amostras previstas como negativas, mas que na realidade são positivas).

### *Accuracy*

A métrica da *accuracy* (Equação 2.1) foi adotada para avaliar o desempenho do modelo, ao longo das *epochs*, durante o treino. Esta mede a proporção de amostras corretamente classificadas em relação ao total de amostras.

A *accuracy* é representada por um valor que varia entre 0 e 1: o valor máximo de 1 indica uma precisão perfeita, o que significa que todos os exemplos foram classificados corretamente, enquanto o valor mínimo de 0 indica uma precisão nula, o que significa que nenhum exemplo foi classificado corretamente.

Apesar da *accuracy* de um modelo ser uma das métricas mais populares para avaliar o desempenho do classificador, é importante lembrar que esta medida pode ser enganosa em algumas situações, especialmente quando os dados são não balanceados, ou seja, quando há uma grande disparidade na distribuição das classes. Quando isto acontece, a *accuracy* pode não fornecer uma compreensão completa do desempenho do modelo, pois o modelo pode ser influenciado pela classe dominante.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1)$$

### *Precision, Recall e F1-score*

Em problemas de classificação, a *precision* (Equação 2.2) é uma métrica de avaliação que mede a proporção de exemplos que foram classificados como positivos corretamente, em relação ao total de exemplos que foram classificados como positivos pelo modelo. Por outras palavras, esta medida refere-se à capacidade do modelo de fazer previsões quando este prevê uma classe como positiva. Quanto maior a *precision*, menor a taxa de falsos positivos.

$$precision = \frac{TP}{TP + FP} \quad (2.2)$$

Por outro lado, a *recall* (Equação 2.3), também chamada de taxa de *sensitivity* ou *True Positive Rate* (TPR), é uma métrica de avaliação que calcula a percentagem de exemplos positivos corretamente identificados em comparação com o total de exemplos positivos.

$$recall = \frac{TP}{TP + FN} \quad (2.3)$$

A *f1-score* (Equação 2.4) é uma métrica de avaliação que combina a *recall* e a *precision* num valor único, para fornecer uma avaliação mais ampla do desempenho do modelo de classificação. É útil em situações em que as classes não estão equilibradas ou quando é necessária uma métrica distinta, que leve em consideração tanto os verdadeiros positivos como os verdadeiros negativos.

$$F_1 = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \quad (2.4)$$

### Matriz de Confusão

A **matriz de confusão** é uma ferramenta fundamental na avaliação de modelos de classificação. Cada célula representa uma classe específica: uma para os casos positivos e outra para os casos negativos. As categorias da matriz são os TP, os FP, os TN e os FN, referidas e explicadas acima.

Ao examinar a matriz de confusão, é possível avaliar o desempenho do modelo em cada uma destas categorias, o que fornece uma ajuda na identificação das previsões que o modelo acerta e erra. Com base nestas informações, é possível tomar decisões informadas, de maneira a otimizar o sistema, ajustar parâmetros e garantir um diagnóstico preciso [29].

### Curva *Receiver Operating Characteristic* (ROC) e *Area Under the Curve* (AUC)

A Curva ROC [29] é uma representação gráfica que expõe a capacidade de um modelo de classificação distinguir entre classes positivas e negativas. Esta é construída traçando a Taxa de Verdadeiros Positivos (TPR) no eixo Y e a Taxa de Falsos Positivos (FPR) no eixo X, para diferentes valores de limiar. A TPR é também conhecida como recall, e a FPR é calculada como a especificidade (o quociente entre os verdadeiros negativos e a soma dos verdadeiros negativos com os falsos positivos). A curva ROC demonstra como o desempenho do modelo varia em diferentes pontos de corte, ajudando a escolher um limiar que equilibre a taxa de falsos positivos e verdadeiros positivos, de acordo com o contexto do problema.

A AUC é uma métrica numérica que simboliza a curva ROC num único valor. Esta mede a capacidade global do modelo de classificar corretamente exemplos positivos em relação a exemplos negativos, independentemente do limiar de decisão escolhido. O valor da AUC varia de 0 a 1, onde um valor de 0,5 indica uma classificação aleatória (sem discriminação) e um valor próximo de 1 uma classificação perfeita. Um valor próximo a 0 indica uma classificação invertida, ou seja, a classificação errada da classe positiva e negativa.

Quando a curva ROC está próxima da diagonal (linha de 45 graus), o modelo não possui discriminação, ou seja, não consegue diferenciar entre classes positivas e negativas de maneira eficaz. Em contrapartida, quando a curva ROC se afasta da diagonal, o modelo demonstra uma maior capacidade de discriminação. Desta forma, um valor de AUC próximo de 0,5 sugere que o modelo não oferece melhor desempenho do que uma escolha aleatória, enquanto um valor de AUC abaixo de 0,5 indica que o modelo está a manifestar um pior desempenho do que uma escolha aleatória [29, 30].

## 2.7 Irregularidades de Distribuição de Dados - Dados Não Balanceados

Ao explorar a relevância de modelos de *Machine Learning* em cenários complexos, surge a necessidade de lidar com um problema comum como o não balanceamento dos dados [31]. Quando um conjunto de dados apresenta uma distribuição desproporcional entre as suas classes, pode levar à formação de modelos que não conseguem capturar a verdadeira natureza dos dados e, conseqüentemente, produzem previsões imprecisas e desajustadas.

O não balanceamento dos dados pode resultar num treino inadequado do modelo, com um destaque desproporcional na classe maioritária, deixando de parte a classe minoritária e diminuindo a sua capacidade de representação. Este desequilíbrio pode levar a uma *loss* de treino que minimiza os erros predominantes da classe maioritária, mas falha em capturar padrões importantes da classe minoritária.

Este problema pode afetar o desempenho do modelo de várias maneiras. Este pode tornar-se tendencioso em relação à classe maioritária, uma vez que possui mais exemplos dessa classe para aprender durante o treino. Neste tipo de conjuntos, o modelo pode ter alta precisão na classe maioritária, mas baixa precisão na classe minoritária. Isto pode acontecer porque o modelo tende a prever a classe maioritária para a maioria dos exemplos. Este poderá também produzir uma alta taxa de falsos positivos ou negativos em relação à classe minoritária, ou seja, irá classificar de forma incorreta bastantes exemplos dessa classe como pertencentes a outra classe.

Existem algumas métricas comumente usadas para quantificar o nível do não balanceamento dos dados. Uma delas é o *Imbalance Ratio*.

Sendo  $n_{maj}$  e  $n_{min}$  o número de amostras na classe maioritária e minoritária, respetivamente, o cálculo desta métrica está apresentado na Equação 2.5. Esse valor representa a quantidade de amostras da classe maioritária, para cada amostra da classe minoritária. Por exemplo, se  $IR = 10$ , ou seja, se para cada amostra da classe minoritária existem 10 amostras da classe maioritária, este *dataset* é considerado não balanceado se  $IR > 2$  [32].

$$IR = \frac{n_{maj}}{n_{min}} \quad (2.5)$$

Para lidar com dados não balanceados, podem ser aplicadas várias estratégias. As mais comuns são a técnica do *oversampling* da classe minoritária e do *undersampling* da classe maioritária. O objetivo da aplicação destes métodos é garantir que o modelo se torne capaz de aprender adequadamente a partir das classes minoritárias, e não se torne transversal em relação à classe maioritária.

### 2.7.1 *Undersampling*

O *undersampling* procura resolver o problema do não balanceamento dos dados, reduzindo a quantidade de exemplos da classe majoritária de modo a igualá-la à classe minoritária, criando um conjunto de dados balanceado. Esta técnica ajuda o modelo a focar-se igualmente em ambas as classes, permitindo uma representação mais justa e abrangente dos dados.

Porém, esta diminuição do número de dados poderá levar a que se perca informação, uma vez que se podem estar a descartar exemplos que podem conter *insights* valiosos sobre a variabilidade e complexidade dessa classe. Essa redução drástica de informações pode levar a um modelo menos abrangente e pode até mesmo causar o surgimento de distorções nas previsões [33, 34].

### 2.7.2 *Oversampling*

*Oversampling* é uma técnica de balanceamento de dados usada para lidar com conjuntos de dados não balanceados, onde uma ou mais classes possuem um número significativamente menor de amostras em relação a outras classes. O objetivo deste método é aumentar o número de amostras nas classes minoritárias, para que o modelo de *Machine Learning* seja capaz de aprender de maneira mais precisa a partir dessas classes e evitar a tendência em direção à classe majoritária.

Esta metodologia deve ser aplicada apenas no conjunto de treino, afim de evitar que dados duplicados ou sintéticos influenciem a avaliação do modelo no conjunto de validação/teste. Se o número de exemplos gerados for grande, em alguns casos, o *oversampling* pode aumentar o risco de *overfitting*.

Existem diversas abordagens para realizar o *oversampling*. Um dos métodos mais comuns é a técnica chamada *Synthetic Minority Over-sampling Technique* (SMOTE). Esta produz novas amostras sintéticas para a classe minoritária, criando instâncias “artificiais” intermediárias entre os exemplos existentes.

#### SMOTE

Como descrito acima, o SMOTE é uma técnica comum, usada para implementar o *oversampling*.

Para cada exemplo na classe minoritária, o SMOTE escolhe, de maneira aleatória, um número dos seus “vizinhos” mais próximos. Para cada vizinho elegido, o SMOTE calcula a diferença entre o exemplo atual e o vizinho. De seguida, multiplica essa diferença por um valor aleatório entre 0 e 1 e adiciona a diferença ponderada ao exemplo atual, de maneira a formar uma nova amostra artificial. Reproduz este processo diversas vezes, com o objetivo de criar novos exemplos sintéticos.

## 2.8 Técnicas para Redução/Eliminação de *Overfitting*

### 2.8.1 *Overfitting*

O *overfitting* é um problema comum em *Machine Learning*, onde um modelo se ajusta excessivamente aos dados de treino e perde a capacidade de generalizar para novos dados. Isto pode levar a um desempenho inferior em dados não vistos, causando uma discrepância entre a performance nos dados de treino e nos dados de teste ou validação. Por outras palavras, o *overfitting* ocorre quando o modelo “decora” os dados de treino em vez de aprender padrões relevantes, resultando numa pior capacidade de generalização para novos cenários [35].

O desenvolvimento de modelos de Inteligência Artificial (IA) para inferir informações valiosas a partir de dados biomédicos é uma tarefa complexa e desafiadora. À medida que se avança em direção à criação de sistemas mais sofisticados, surge uma preocupação fundamental: o *overfitting*.

Como dito acima, o *overfitting* ocorre quando um modelo é treinado de forma excessivamente detalhada nos dados de treino específicos, tornando-o altamente adaptado a esses dados, mas com dificuldades de generalização para novos dados não vistos. Esta limitação pode levar a conclusões equivocadas e a uma perda significativa da eficácia/capacidade quando aplicado a cenários reais [35].

Existem diversas técnicas eficazes para combater o *overfitting* e garantir a generalização adequada do modelo. Algumas delas incluem o uso do *dropout*, a regularização de pesos  $L1$  e  $L2$ , o *k-fold cross validation*, entre outras.

### 2.8.2 *Dropout*

O *dropout* é uma técnica popular de regularização, utilizada em redes neuronais para combater o *overfitting* [35].

A ideia principal do *dropout* consiste em desligar aleatoriamente um conjunto de unidades (neurónios) em cada camada, tornando-os temporariamente inativos, durante o treino da rede neuronal. Esta desativação temporária simula o treino de várias redes neuronais menores, e cada uma delas aprende uma parte diferente da tarefa. No final do treino, todos os neurónios são ativados novamente. Esta aleatoriedade força a rede neuronal a não depender excessivamente de nenhum neurónio específico e impede que as unidades “decorem” os padrões nos dados de treino. Assim, a rede torna-se mais robusta e generaliza melhor para novos dados [35].

A percentagem de desativação controla o grau de regularização aplicado e pode ser ajustada com base nas características específicas do problema e do modelo em questão.

A aplicação desta técnica em *hidden* e *dense layers* serve o mesmo propósito, no entanto, a diferença reside nas características dos neurónios em cada tipo de camada. Em *hidden layers*, o *dropout* é aplicado para evitar que neurónios específicos se tornem muito dependentes de outros neurónios próximos durante o treino, garantindo que o conhecimento seja distribuído de forma

mais equilibrada pela rede. Assim, é mais provável de evitar que a rede memorize os dados de treino e, em vez disso, aprenda padrões mais gerais. Já em *dense layers* (ou *fully connected layers*), os neurónios são altamente interconectados, o que pode levar a um maior risco de *overfitting*. Portanto, é comum aplicar uma taxa de *dropout* mais elevada nestas camadas, para promover a independência dos neurónios entre si.

### 2.8.3 Regularização de Pesos

A **regularização de pesos** ( $L1$  e  $L2$ ) é uma técnica utilizada para combater o *overfitting* em modelos de *Machine Learning*, principalmente em redes neuronais. O objetivo da regularização é evitar que os pesos dos neurónios se tornem muito grandes, o que pode levar a um ajuste excessivo aos dados de treino. O parâmetro de regularização é um hiperparâmetro que controla o quanto a penalidade é aplicada. Um valor maior de regularização implica uma penalidade mais forte, o que pode levar ao *underfitting*, enquanto que um valor menor pode permitir um ajuste mais flexível do modelo, mas também aumenta o risco de *overfitting*.

Existem duas formas comuns de regularização de pesos:  **$L1$  (Lasso)** e  **$L2$  (Ridge)**. Na regularização  $L1$  (Lasso), é adicionada uma penalidade à função de perda do modelo, proporcional ao valor absoluto dos pesos. Isto leva à criação de uma “função de penalidade L1” que penaliza os pesos com valores muito grandes, tornando-os próximos de zero. Como resultado, alguns dos pesos podem ser exatamente iguais a zero, tornando o modelo mais disperso, o que significa que este selecionará apenas os recursos mais importantes para a tarefa.

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M \left( y_i - \sum_{j=0}^p w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^p |w_j| \quad (2.6)$$

Na regularização  $L2$  (Ridge), é também adicionada uma penalidade à função de perda do modelo, mas proporcional ao quadrado dos pesos. Isto cria uma “função de penalidade L2” que penaliza os pesos com valores grandes, mas não os coloca completamente a zero. Estes pesos são reduzidos, mas não eliminados, o que evita o excesso de ajuste.

$$\sum_{i=1}^M (y_i - \hat{y}_i)^2 = \sum_{i=1}^M \left( y_i - \sum_{j=0}^p w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^p w_j^2 \quad (2.7)$$

Ambas as técnicas de regularização ajudam a controlar a complexidade do modelo e a evitar o *overfitting*. A diferença entre elas está no tipo de penalidade aplicada aos pesos. Ao adicionar termos de regularização à função de perda do modelo, o processo de aprendizagem considera,

tanto a minimização da função de perda como a redução dos pesos, o que incentiva o modelo a encontrar soluções mais gerais, melhorando a sua capacidade de generalização para novos dados.

#### 2.8.4 *K-Fold Cross Validation*

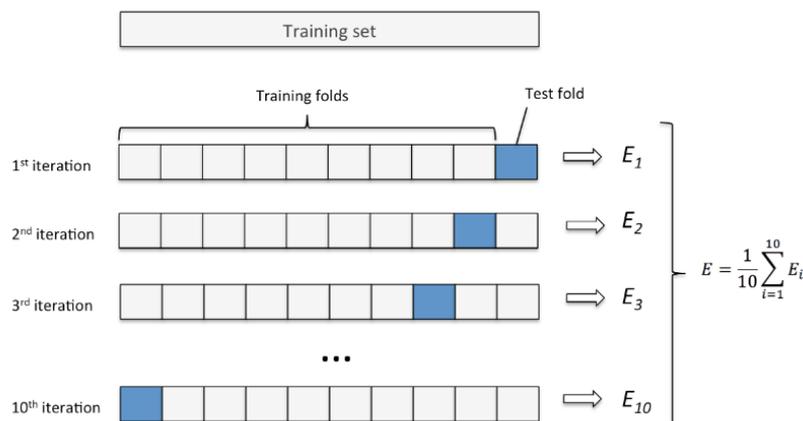
O *K-Fold Cross Validation* é uma técnica usada para avaliar a performance de um modelo de *Machine Learning* de forma mais robusta e precisa.

Primeiro, é feita uma divisão do conjunto de dados. Este é dividido em  $K$  partes (as chamadas *folds*) de tamanho aproximadamente igual. De seguida, o modelo é treinado e validado  $K$  vezes.

O valor de  $K$  é selecionado de modo a garantir que cada conjunto de treino/validação/teste contenha amostras de dados suficientes para representar estatisticamente o conjunto de dados completo.

Em cada iteração, um dos *folds* é usado como conjunto de validação e os  $K-1$  *folds* restantes são usados como conjunto de treino. Em cada iteração, o modelo é avaliado usando uma métrica de performance, como a *accuracy*, *F1-score* ou *loss*. No final das  $K$  iterações, as métricas de performance são agregadas, geralmente calculando-se a média ou a mediana dos resultados obtidos.

Na Figura 2.27 está representado o esquema do método em questão, onde o  $E$  representa a performance do modelo, para um exemplo com  $K = 10$ .



**Figura 2.27:** Esquema do método *K-Fold Cross Validation* [36].

O uso do *K-Fold Cross Validation* ajuda a reduzir o *bias* (diferença entre a média das previsões do modelo e o valor real) e a variância (sensibilidade do modelo a pequenas variações nos dados de treino) da avaliação do modelo, uma vez que utiliza diferentes partes dos dados, tanto para treino como para validação em cada iteração, garantindo a avaliação do modelo em diferentes conjuntos de dados.

O resultado final do *K-Fold Cross Validation* é uma avaliação mais precisa e confiável da performance do modelo, pois considera diferentes divisões dos dados para treino e validação.

Esta técnica é amplamente utilizada pois ajuda a identificar possíveis problemas de *overfitting* e *underfitting* [37, 38].

No presente capítulo, é abordada uma ampla gama de fundamentos teóricos relacionados com a temática central desta pesquisa. São exploradas as principais teorias e modelos, desenvolvidos por psicólogos e investigadores, sobre emoções e computação afetiva, com ênfase no modelo de Ekman [15], como base conceitual, reconhecimento de emoções através de expressões faciais e os traços distintivos para identificar cada emoção, em específico. Também é caracterizada a doença da demência, nas suas diferentes fases e sintomas, bem como o conceito de *engagement* e respetivos níveis. Além disso, é descrita, em detalhe, a arquitetura usada para treinar o modelo, a VGG-16, e são discutidas as ferramentas de avaliação e estratégias de processamento de dados para enfrentar desafios como desequilíbrio de dados e *overfitting*. Esta base teórica servirá como fundamento para os estudos empíricos e análises subsequentes realizadas neste trabalho, proporcionando uma compreensão abrangente e fundamentada para as investigações futuras.

## Trabalhos Relacionados

O *engagement* tem sido uma ferramenta bastante utilizada para descrever diversas realidades. Não é surpresa que, nos últimos anos, tenham surgido vários estudos que ligam o *engagement* a múltiplas áreas, tais como saúde, educação e até mesmo marketing, que visam à diminuição e até resolução de algumas adversidades presentes nestes domínios.

Para medir o envolvimento cognitivo, Li [39] adotou uma perspectiva analítica, fornecendo informação acerca dos instrumentos e técnicas usadas, com o objetivo de ajudar os profissionais destas áreas a melhorar os seus métodos de medição. Para isto, começou pela leitura de 52 artigos retirando sempre, no final, a resposta à questão “Como é que os autores captam a construção do *engagement* cognitivo?”. Verificou que este pode ser detetado através de métodos observacionais, reconhecido por indicadores comportamentais e linguísticos específicos; entrevistas, com o uso de estratégias cognitivas e metacognitivas estruturadas para obter informações sobre o uso de táticas utilizadas, por exemplo, pelos alunos; amostragem de experiência, que envolve o uso de dispositivos eletrónicos ou digitais de maneira a interromper os alunos e perceber os pensamentos e sentimentos naquele exato momento; rastreamento ocular, que funciona pela deteção de posições e movimentos dos olhos (duração da fixação, aumento da pupila, entre outros) e medidas fisiológicas, com o objetivo de medir os sinais elétricos produzidos na pele, no cérebro e nos músculos, através do uso de Atividade Eletrodérmica (EDA), Eletrocardiograma (ECG) e Eletromiografia (EMG), respetivamente.

Também Azevedo [40] desenvolveu um estudo acerca do *engagement* e aprendizagem na ciência. Este resume alguns métodos e ferramentas disponíveis para investigadores em cada categoria diferente, indicando se são adequados para medir processos cognitivos, metacognitivos, afetivos ou motivacionais. Azevedo defende que o tipo de *engagement* pode variar desde o nível micro, ou seja, um indivíduo no momento, tarefa, e atividade de aprendizagem, até ao nível macro, como por exemplo, um grupo de alunos numa turma, curso ou escola. Assim esta medição, no nível mais pequeno, pode incluir índices fisiológicos e psicológicos de imagens do cérebro, rastreio ocular, tempo de resposta, e atribuição de atenção. Análises do discurso, observações, classificações e outras análises contextuais de aprendizagem, escolarização e práticas entre e dentro dos grupos, constituem algumas das medições de *engagement* no nível macro. De maneira a descobrir quais os melhores métodos para medir a concentração antes, durante e após a aprendizagem, o autor submeteu um grupo de alunos a uma sessão de 120 minutos, englobando uma aula de 90 minutos da disciplina de Ciências. Métodos como gravações de vídeo,

expressões faciais, questionários, discursos, auto-relatórios, sensores fisiológicos, etc, são usados como ferramentas de *input* para esta investigação, ajudando a perceber quais serão as melhores técnicas para detetar *engagement* na cognição, metacognição, afeto e motivação. Azevedo concluiu que todos os métodos para medir a cognição são adequados, à exceção das expressões faciais. Quanto à metacognição, apenas metade das ferramentas usadas são apropriadas para este tipo de medição, sendo exemplo de um bom método as gravações de vídeo e áudio e mau método também as expressões faciais. Para aferição do afeto, apenas as expressões faciais, questionários, discursos e sensores fisiológicos apresentam bons processos para esta deteção. Por último, somente os auto-relatórios, questionários e discursos, demonstram ser idealmente adequados para capturar e medir processos relacionados com a motivação.

O trabalho de D’Mello et al. [41], para além de referir todas as medidas e técnicas faladas acima, foca o seu estudo em AAA (*The Advanced, Analytic, Automated Measurement Approach*). AAA trata-se de uma medida que permite avaliar, continuamente, componentes de *engagement* sem a necessidade de envolvimento humano. Esta abordagem confere vantagens em relação a todas as outras, uma vez que não é afetada por lapsos momentâneos de atenção ou fadiga, reduz o tempo e o esforço e é, por isso, considerada mais objetiva. Esta medição é baseada na deteção de algumas ativações dos músculos faciais, como por exemplo a elevação interna da sobrancelha ou o franzimento dos lábios, e técnicas no domínio da fala, usadas para extrair características paralinguísticas, como tom e amplitude de um sinal de áudio. Seguidamente, estes dados foram processados (*denoising*, filtragem, suavização, etc) e, com ajuda de *Machine Learning*, o modelo utiliza-os para gerar estimativas automáticas de estados mentais. Neste artigo, D’Mello et al. reúne diversos *case studies* que utilizam diferentes métodos de *Supervised Learning*. A melhor *performance* foi conseguida por Kappor e Picard [42] que, através de *Gaussian Mixture Models*, analisando expressões faciais e movimentos/postura corporais, atingiram uma *accuracy* de 87%.

Monkaresi, Bosch, Calvo e Mello [43] colocaram estas técnicas em prática desenvolvendo um estudo sobre a deteção de *engagement* usando expressões faciais presentes em vídeos e frequência cardíaca. O *dataset* foi construído com dados de 23 estudantes de engenharia de uma universidade na Austrália, constituído por vídeos filmados enquanto cada aluno estava sentado em frente a um computador a escrever uma composição sobre um local que tinha visitado. Após a escrita da redação, os alunos foram separados em quatro grupos: (1) tiveram que entregar a redação em 30 minutos; (2) tiveram que esperar para receber *feedback* mas estavam livres para editar o manuscrito; (3) receberam logo *feedback* humano acerca da qualidade da escrita; (4) tiveram um tempo adicional de 20 minutos para melhorar a sua redação. Após a construção do *dataset* e respetiva extração de *features*, os autores testaram vários classificadores de maneira a obter um modelo com melhor *performance*. Concluíram que o melhor classificador era o *Updateable Native Bayes*, tendo obtido uma *Area Under the Curve* (AUC) de 73,3%.

Com o objetivo de determinar o estado afetivo de um usuário, Gupta, Cunha, Awasthi e Balasubramanian [44] desenvolveram um estudo usando métodos de visão computacional e *Machine Learning*. Primeiramente, estudaram e compararam uma vasta rede de conjuntos de dados diferentes compostos por imagens estáticas, vídeos, áudios, ECG e EDA. No entanto, para

este estudo apenas foi usada a base de dados *DAiSEE* que contém 9.068 sequências de vídeos de 112 sujeitos, com *labels* de *engagement*, frustração, confusão e tédio. Estas *labels* foram classificadas, para cada sequência de vídeo, como “Muito Baixo”, “Baixo”, “Alto” e “Muito Alto”. Os autores testaram cinco modelos diferentes (*InceptionNet Frame Level*, *InceptionNet VideoLevel C3D*, *Training C3D*, *FineTuning*, *LRCN*) sendo que o último foi o que desencadeou melhores resultados de *accuracy* (57,9% para *engagement*).

Segundo Whitehill et al. [45], as melhores medidas para detetar *engagement* passam pelo uso de auto-relatórios, listas de verificação observacionais e escalas de classificação e medições automáticas. No entanto, o tema deste artigo vai ao encontro de técnicas computacionais, reconhecendo *engagement* através do rosto, postura corporal e gestos. Usando um *dataset* composto por 13.584 *frames* de vídeos de 34 estudantes que participaram num treino de capacidades cognitivas, avaliado por 4 estados diferentes (“*Not engaged at all*”, “*Nominally engaged*”, “*Engaged in task*” e “*Very engaged*”), os autores foram capazes de desenvolver um modelo que detetasse o nível de atenção do aluno. Estes usaram 3 classificadores binários: “*Box Filter*”, “*Gabor Energy Filters*” e “*Facial Action Coding System*”, sendo que os dois últimos foram os que melhor desempenho mostraram, tendo conseguido alcançar valores de *accuracy* próximos dos valores humanos (79,3% e 71,4%, respetivamente).

Através da *Expression Recognition Toolbox (CERT)*, Grafsgaard et al. [46] analisaram alguns movimentos faciais, de maneira a detetar *engagement* em alunos de uma universidade dos Estados Unidos. Estes movimentos consistem em levantar e baixar a sobrancelha, apertar as pálpebras e bocejar (abrindo a boca). O *dataset* utilizado é composto por *frames* de vídeo de 67 alunos, captadas durante a realização de uns exercícios de programação na linguagem *Java*. Primeiramente, os autores dividiram os valores de saída de cada estudante em “AU-presente” (tem presente algum dos movimentos faciais a serem estudados) e “AU-não presente” (não tem presente nenhum dos movimentos faciais a serem estudados) e calcularam as suas médias. Estes valores serviram de *input* para o modelo utilizado- *Logistic Regression Model*, produzindo estas categorias previstas de “AU-presente” e “AU-não presente”. Para avaliar a *performance* do modelo, utilizaram a estatística de *Cohen’s Kappa* e a métrica *accuracy*. Os resultados foram os seguintes: levantar a parte interior da sobrancelha, 0,14 para a estatística e 57% para a *accuracy*; levantar a parte exterior da sobrancelha, 0,29 para a estatística e 64% para a *accuracy*; baixar da sobrancelha, 0,05 para a estatística e 54% para a *accuracy*; apertar as pálpebras, 0,29 de estatística e 64% de *accuracy* e, por fim, as típicas “cavinhas” nas bochechas, 0,29 de estatística e 64% de *accuracy*.

As emoções, constituindo uma parte vital da vida humana, desempenham um papel completo no que toca à forma como o ser humano compreende e percebe o mundo à sua volta. Durante muitos anos, foram desenvolvidos diversos métodos manuais com o objetivo de simplificar a análise de emoções, como por exemplo, questionários elaborados por psicólogos. No entanto, com a evolução da tecnologia, surgiram técnicas bastante mais simples para o reconhecimento das emoções.

Segundo Saxena et al. [47], a melhor forma para realizar o reconhecimento de emoções

é através das expressões faciais e do áudio. Neste estudo, foram analisados quatro métodos possíveis para a identificação de emoções humanas, sendo eles as expressões faciais, o áudio, o texto e sinais fisiológicos. Para a primeira análise, foram comparados vinte e três processos de classificação diferentes, sendo que o melhor foi o algoritmo de *Stationary Wavelet Transform*, com um *recognition rate* de 98,83%. No que diz respeito ao reconhecimento a partir do áudio, realizado através da avaliação de *clips* de voz, chamadas telefônicas, etc, também foram testadas várias técnicas, sendo que a que permitiu obter uma melhor *performance* foi um algoritmo de otimização híbrido que combina *Particle Swarm Optimization* (PSO) e *Biogeography-based Optimization* (BBO), com um *rate* de 99,47%. O reconhecimento de emoções através de texto é realizado com base numa seleção de palavras-chave consideradas emocionais, presentes num texto escrito pelo utilizador em questão. À semelhança das técnicas acima descritas, também para este método foram estudados variados procedimentos para o reconhecimento emocional. O mais bem classificado, com um *rate* de 87,02%, foi a combinação entre uma técnica de relacionamentos estruturais- *Rough Set Theory* e *Support Vector Machine* (SVM). Por último, com a ajuda de sinais fisiológicos do tipo ECG, foi possível perceber que a melhor estratégia será a de utilizar sinais *wireless*, tendo-se conseguido um *rate* de 87%. Para a realização desta técnica, é transmitido um sinal de rádio-frequência e, através dos batimentos cardíacos, são analisados os reflexos corporais que, após uma avaliação, serão associados a uma ou várias emoções.

Para adquirir de forma mais eficaz características emocionais em áudios de fala, Jiang et al. [48] fez uma proposta de modelo usando a rede *Parallelized Convolutional Recurrent Neural Network* (PCRN). Inicialmente fez extração de *features*, com o auxílio de *long short-term memory* e a rede escolhida para o treino do modelo foi uma *Convolutional Neural Network* (CNN). O classificador usado para rotular as emoções foi o *SoftMax*. Para verificar a veracidade do algoritmo, este foi testado em quatro *datasets* diferentes. O primeiro, proveniente do *Institute of Automation, Chinese Academy of Sciences*, é composto por 1200 afirmações feitas por quatro atores (dois homens e duas mulheres) que se encontravam a representar seis diferentes emoções-raiva, medo, felicidade, neutro, tristeza e surpresa. O segundo, *EMO-BD* da Universidade de Berlim, consiste em 535 afirmações de dez atores profissionais (5 homens e 5 mulheres). O conjunto de dados *ABC* é uma base de dados alemã composta por um vídeo de 11.5 horas de afirmações de quatro homens e quatro mulheres. Por fim, é utilizado também o *dataset SAVEE* que, à semelhança dos acima descritos, apresenta gravações de quatro atores do sexo masculino, representando sete emoções diferentes. Para comprovar a exatidão do reconhecimento destas emoções, foram apresentadas quatro matrizes de confusão, uma para cada *dataset* testado. A partir delas, é possível concluir o seguinte: para o *dataset CASIA*, as emoções da raiva e da tristeza são as que apresentam melhores resultados de *accuracy*, 75% e 72%, respetivamente; no que diz respeito ao *dataset EMO-DB*, a raiva e a tristeza demonstram valores de *accuracy* superiores a 90% no entanto, o aborrecimento e o neutro também revelam ter uma boa *accuracy*, 87,65% e 88,61%, respetivamente; o maior *rate* de reconhecimento é obtido para a emoção da agressividade no *dataset ABC*, com uma *accuracy* de 86,31%; por último, no *dataset SAVEE*, o melhor *rate* pertence à emoção do neutro, com um valor de 84,17%.

Com o propósito de melhorar a qualidade de vida das pessoas mais idosas que sofrem de

problemas neurológicos, Nahian et al. [49] desenvolveu um projeto que, para além de monitorizar, com precisão, as quedas que cada pessoa dá em tempo real, também faz previsões sobre as condições de saúde da mesma e dá informações sobre o seu estado emocional. Em caso de emergência, será emitido um alerta e uma ambulância inteligente (ou uma clínica móvel) chegarão a casa do idoso, através de coordenadas enviadas no momento, via GPS. Os dados retirados ao longo do dia são armazenados numa *cloud* que poderá ser acedida pelo médico, o que desencadeia uma maior ligação na relação médico/paciente. Existe também uma câmara que regista a expressão facial do idoso, que é imediatamente capturada e mapeada numa emoção. Após o registo de todos os movimentos do idoso, bem como da sua expressão facial no momento da queda, os dados são processados por um modelo de Inteligência Artificial (IA) e, no que toca à expressão facial, o resultado da emoção detetada será enviado para o sistema cognitivo, que servirá para criar um plano para outras medidas que deverão ser adotadas no futuro. Este projeto é composto por três camadas. A primeira constitui uma série de sensores que são colocados na casa do idoso, para a recolha de informação contínua: alguns dispositivos inteligentes, tais como um relógio, uma pulseira e uma cinta com um sensores de movimento (para captar os movimentos no momento da queda), e será também essencial uma câmara fotográfica para captar a expressão facial do idoso. A segunda camada é composta por um servidor que irá filtrar todos os dados e remover os redundantes. Por fim, existirá um *cloud server* onde serão tomadas todas as decisões por parte da IA (deteção dos diferentes estados emocionais da pessoa idosa, deteção de quedas, etc). O *dataset* usado para este estudo é composto por resultados de sensores e câmaras, correspondentes a seis diferentes atividades e cinco diferentes quedas, que constitui um total de 31339 eventos de atividades e 2669 quedas. Para a deteção de quedas, foi usada uma *Long Short Term Memory* (LSTM) e três camadas *dense*. A *accuracy* foi de 93,17% para as atividades e 85,07% para a deteção de quedas. Foram testados variados modelos de classificação de *Machine Learning* tais como *Logistic Regression*, *SVM*, *Naive Bayes*, *Random Forest*, etc, no entanto, o modelo de *Logistic Regression* foi o que obteve uma melhor *performance* (90,28% de *accuracy*), tendo sido, assim, o escolhido. Para o reconhecimento de emoções (divididas em cinco diferentes emoções), foi usado um modelo pré-treinado chamado *MobileNetV1* (uma espécie de CNN no entanto, a convolução é realizada para cada canal de entrada, separadamente), onde se obteve uma *accuracy* de apenas 61,11%, onde as imagens com a emoção da felicidade foram as mais bem classificadas, ao contrário da emoção da tristeza, que obteve a pior *performance*. À semelhança da deteção de quedas, também para as emoções foram testadas as mesmas técnicas de classificação, sendo que a melhor, ainda que baixa, foi a *Random Forest*, com uma *accuracy* de 46%.

As CNN foram introduzidas no final dos anos 90, mas eram limitadas devido à falta de dados de treino e poder computacional. Após a década de 2010, com o aumento do poder de computação e a recolha de conjuntos de dados maiores, estas redes têm sido capazes de alcançar resultados que anteriormente eram considerados como sendo puramente do domínio humano [50].

Apesar disso, existem ainda alguns problemas que envolvem este modelo, como o *overfitting*. Para contextualizar, o *overfitting* ocorre quando um modelo treina os dados demasiado bem.

Isto acontece quando um modelo aprende o detalhe e o ruído nos dados de treino ao ponto de criar um impacto negativo no desempenho do modelo, com novos dados [35]. Por este motivo, foram propostas várias técnicas para melhorar o desempenho. Por exemplo, quando se trata de *overfitting*, o *dropout*, a regularização e o aumento dos dados são utilizados para o evitar, pois melhoram a capacidade de generalização. Além disso, ao utilizar um *Adam Optimizer* em vez de um *SGD*, permite um treino mais rápido, uma vez que este utiliza técnicas como o momentum e taxas de aprendizagem adaptativas [51]. O *Batch Normalization* também foi desenvolvido para ajudar a evitar o desaparecimento e explosão do gradiente. Por outras palavras, esta é uma técnica usada para normalizar as ativações de uma rede neural em cada lote de treino, tornando-as mais estáveis e facilitando este processo [52].

Com os desenvolvimentos e a extensa investigação sobre CNN, estas são agora vistas como uma ferramenta muito mais viável na extração de características, no reconhecimento de padrões e no processamento de imagens. Muitas variantes de CNN obtiveram resultados notáveis com uma precisão entre 65% e 95% [53, 54]. Foram também estudados os desempenhos de diferentes arquiteturas. Na verdade, ao comparar o *VGG*, o *Inception* e o *ResNet*, verificou-se que o *VGG* tinha uma precisão de 72,7%, seguido do *ResNet* com 72,4% e do *Inception* com 71,6% [47].

É notória a preferência pela arquitetura *VGG*. No entanto, existem duas variantes principais: *Visual Geometry Group (16 layers)* (VGG-16) e *Visual Geometry Group de 19 Camadas* (VGG-19). Estas distinguem-se pelo número de camadas convolucionais, totalmente conectadas, uma vez que a primeira apresenta 16 camadas e a segunda 19. Segundo *Gwyn et al.* [55] e *Bodavarapu et al.* [24], que fizeram um estudo comparativo sobre estes dois tipos de *VGG*, a estrutura que apresenta melhores resultados no tema do reconhecimento facial, tanto de *accuracy* como de *loss*, é a VGG-16.

Além disso, foi demonstrado que o desempenho é melhorado quando vários modelos diferentes são agrupados [47, 56]. No entanto, o objetivo deste estudo é otimizar uma única rede, o que permite melhorar ainda mais o desempenho do conjunto, utilizando a CNN VGG-16.

Em resumo, a pesquisa na área de reconhecimento de emoções e deteção de *engagement* tem avançado significativamente nas últimas décadas, impulsionada pelo aumento do poder computacional e pela disponibilidade de conjuntos de dados cada vez mais amplos. Estudos recentes exploraram uma variedade de técnicas, incluindo o uso de CNN, como o VGG-16, a aplicação de sensores fisiológicos e análises de expressões faciais, bem como abordagens baseadas em *Machine Learning* (ML). Além disso, alguns investigadores têm procurado aplicar essas técnicas em diversos contextos, desde a deteção de emoções em vídeos, até à avaliação de níveis de *engagement* em ambientes de ensino e saúde. O desafio do *overfitting* tem sido enfrentado com o uso de técnicas como *dropout*, regularização e aumento de dados, contribuindo para melhorar a generalização dos modelos. À medida que a pesquisa avança, acompanhando a evolução dos desafios, é fundamental explorar novas abordagens e aperfeiçoar as técnicas existentes. Esse esforço torna-se ainda mais importante no contexto médico, onde a compreensão das emoções humanas e do *engagement* desempenha um papel crucial na otimização dos resultados do tratamento, promovendo, assim, uma assistência mais eficaz e uma melhoria na qualidade de vida

dos pacientes.



## 4

# *Datasets*

Através do Capítulo 3, pode concluir-se que já existem alguns estudos e pesquisas que ligam o *engagement* à área da saúde. No entanto, nenhum dos investigadores abordou a relação de *engagement* com a doença da demência. O estudo desta associação é necessário, uma vez que permitirá a possível resolução de diversos problemas ligados a este tipo de doença. Por esta razão, o tema desta tese centra-se no problema relacionado com a extrema dificuldade que os profissionais de saúde enfrentam, em conseguir que os seus pacientes mantenham o interesse durante a fase dos tratamentos. Durante esta etapa, o objetivo será detetar *engagement* em pessoas com demência, através da análise das suas expressões faciais.

Numa fase preliminar, o principal objetivo foi detetar a presença de *engagement*, usando um *dataset* composto por imagens faciais de diversas pessoas, aparentemente saudáveis. Posteriormente, utilizando o mesmo modelo de Inteligência Artificial (IA), elaboraram-se testes, fazendo uso de um *dataset* real, constituído também por imagens faciais, sendo que estas pertencem a pacientes que poderão, ou não, conter a doença da demência.

Nesta secção serão apresentados os *datasets* utilizados para a elaboração das experiências realizadas, bem como os métodos, ferramentas e bibliotecas escolhidas para as mesmas.

Para uma abordagem mais aprofundada e conveniência de apresentação, a metodologia experimental será apresentada no próximo capítulo, acompanhada dos resultados obtidos e da discussão correspondente.

A linguagem de programação escolhida como a principal ferramenta para o desenvolvimento do modelo de IA proposto foi *Python*. *Python* é amplamente utilizada na área da IA devido à sua sintaxe intuitiva, vasta coleção de bibliotecas especializadas e a sua popularidade na comunidade de *developers*. A escolha de *Python* como a linguagem de programação central, alinha este trabalho com as práticas e tendências atuais em IA, facilitando a colaboração com outros pesquisadores e a utilização de recursos disponíveis na comunidade.

Foram escolhidos dois tipos de conjuntos de dados: um *dataset* retirado do *Kaggle* [57], composto por imagens faciais de diferentes expressões do rosto humano, e um conjunto de dados real, fornecido pelo *Center for Research in Neuropsychology and Cognitive and Behavioral Intervention* (CINEICC), com base num programa de educação para a saúde cerebral, constituído por *frames* de vídeo das expressões faciais de cada paciente [58].

## 4.1 Dataset 1

O primeiro *dataset* é constituído por 35887 imagens de expressões faciais de diversas pessoas. Este conjunto de dados contém duas *features*: a primeira constitui a informação de qual a expressão do rosto humano a que se refere a imagem e a segunda refere-se aos pixels constituintes da imagem. Estes dados foram adquiridos na competição “*Challenges in Representation Learning: Facial Expression Recognition Challenge*”, que decorreu na plataforma do *Kaggle* [57], no ano de 2013 [59]. Na Tabela 4.2 são apresentadas, em detalhe, as características deste *dataset*.

Emoção	Número de Imagens
0- Raiva	4953
1- Repugnância	547
2- Medo	5121
3- Alegria	8989
4- Tristeza	6077
5- Surpresa	4002
6- Neutro	6198

**Tabela 4.1:** Características do primeiro *dataset* usado.

Na Figura 4.1 está representada uma imagem da expressão do rosto humano relativa a cada emoção presente no primeiro *dataset*.



**Figura 4.1:** Imagens de cada emoção presente no primeiro *dataset*.

## 4.2 Dataset 2

Para a elaboração do segundo *dataset*, procedeu-se à realização de algumas etapas.

Inicialmente, foram recolhidos vídeos das sessões do programa *Reminder*. Estes vídeos são gravações de tela das sessões de prevenção da doença da demência e foram realizados na plataforma *Zoom*. Cada sessão teve a participação de seis pacientes, além da neuropsicóloga Ana Rita Silva. Apesar de serem, no total, 19 sessões, para este estudo apenas foram utilizadas 7, adquirindo, assim, um total de 7 vídeos, cada um com, aproximadamente, 1 hora de duração.

De seguida, procedeu-se à divisão dos vídeos. Cada vídeo foi dividido em 3 partes, representando os 3 momentos principais de cada sessão. Esta divisão permitiu separar as imagens correspondentes a cada momento específico.

Posteriormente, foram extraídas as *frames* de cada sessão. Para isso, foi desenvolvido um algoritmo que percorreu cada terço de cada vídeo e extraiu *frames*, contendo os rostos das pessoas presentes. Estas *frames* representam capturas de tela em momentos específicos dos vídeos.

No algoritmo 1 está representado o pseudocódigo do algoritmo utilizado para converter os vídeos em *frames* de vídeos.

---

**Algorithm 1** Vídeo para *Frames*


---

**Input:** Vídeos **Output:** Frames- imagens faciais

---

```

1: procedure
2:   video_file  $\leftarrow$  path_video_file
3:   facecascade  $\leftarrow$  path_HaarCascade_file
4:   cap  $\leftarrow$  cv2.VideoCapture(video_file)
5:   sensitivity  $\leftarrow$  100
6:   minarea  $\leftarrow$  500
7:   while true do
8:     ret, frame  $\leftarrow$  read
9:     if ret then
10:      gray  $\leftarrow$  converter
11:      framedif  $\leftarrow$  diff(current, previous)
12:      thresh  $\leftarrow$  threshold(framedif, sensitivity, 255, Threshbinary)
13:      dilatation  $\leftarrow$  dilate(thresh, kernel)
14:      contours  $\leftarrow$  findcont(dilatation)
15:      for contour in contours do
16:        if contour_area(contour) > min_area then
17:          faces  $\leftarrow$  facecascade.detectMultiScalee
18:          for x, y, w, h in faces do
19:            face  $\leftarrow$  frame[y : y + h, x : x + w]
20:            imwrite(facefile, face)
21:          end for
22:          prev_gray  $\leftarrow$  gray.copy
23:        end if
24:      end for
25:    end if
26:  end while
27: end procedure

```

---

Para a elaboração deste algoritmo, fez-se uso da biblioteca *OpenCV* e do classificador *Haar Cascade* para processar os vídeos e realizar a detecção facial. Este começa por ler o primeiro *frame* do vídeo e converte-o para a escala de cinza, como se pode observar nas linhas 2 a 4.

No código implementado, optou-se por utilizar o parâmetro *sensitivity* (linha 5) com o valor de 100. Esta escolha foi baseada na necessidade de capturar pequenos movimentos faciais, como mudanças discretas na expressão dos olhos ou da boca, que podem ser relevantes para a análise realizada. Embora seja importante detetar movimentos significativos em relação ao quadro anterior e remover *frames* desnecessárias, também se reconhece a importância de preservar estas variações. Portanto, um valor de *sensitivity* de 100 foi selecionado para permitir a detecção de tais alterações faciais, garantindo que esses pequenos movimentos sejam considerados no contexto do estudo.

O parâmetro de área mínima (linha 6), definido como 500, foi escolhido com base na natureza dos movimentos e no contexto do estudo. Ao definir uma área mínima, procura-se

filtrar regiões de movimento irrelevantes, removendo variações pouco perceptíveis que podem ocorrer devido a ruídos ou flutuações aleatórias. Optou-se por um valor de 500, de maneira a encontrar um equilíbrio entre detetar movimentos relevantes e evitar a captura de ruído ou movimentos insignificantes.

Em seguida, as imagens são passadas para escala de cinza (linha 10) e cada *frame* subsequente é comparada com a *frame* anterior, de maneira a detetar regiões de movimento significativas. Isto é feito calculando a diferença entre as *frames* e aplicando uma limiarização para obter uma imagem binária (linha 12). Após este procedimento, é realizada uma operação de dilatação para unir regiões próximas (linha 13). A partir destas regiões, são encontrados os contornos e, em seguida, verificadas as áreas dos mesmos para determinar se representam movimentos significativos (linha 14).

Se uma região for considerada significativa, o algoritmo utiliza o classificador *Haar Cascade* para detetar faces nessa região e recorta-as como imagens separadas. Esta técnica é utilizada para identificar e capturar apenas as *frames* consideradas significativas para o estudo (linhas 16 a 20).

Apenas quando é identificado movimento significativo, a detecção facial é aplicada e as faces são recortadas como *keyframes*. A utilização deste método evita a extração de *frames* que não apresentam alterações relevantes, concentrando-se apenas nos momentos de movimento no vídeo. Durante todo o processo, a *frame* atual é copiada como *frame* anterior para a próxima iteração do *loop* (linha 22).

Após a extração das *frames*, estas passam por um processo de detecção de desfocagem, demonstrado no algoritmo 2.

Após a extração das *frames*, o código percorre cada imagem individualmente (linhas 3 a 7). Inicialmente, é realizada uma verificação de desfoque na imagem, utilizando o operador de *Laplace* [60]. Se a variação da transformada de *Laplace* for menor que um determinado limiar, a imagem é considerada desfocada e é removida (linhas 8 a 12).

Em seguida, é utilizada, novamente, a técnica de detecção facial com o classificador *Haar Cascade* para garantir que apenas constam imagens faciais. Se nenhuma face for detetada, a imagem é considerada não relevante para o estudo e também é removida (linhas 13 a 16).

Caso sejam encontradas faces na imagem, apenas a primeira face detetada é mantida para análise. Além disso, é feita uma verificação para determinar se a área ocupada pela face representa pelo menos 70% da área total da imagem. Se não atender a esse critério, a imagem é considerada não significativa e é também eliminada (linhas 18 a 22).

Após passar por todas estas verificações, se a imagem não tiver sido removida, é considerada relevante e é mantida no conjunto de dados.

De seguida, procedeu-se à eliminação de imagens indesejadas, ou seja, das imagens relacionadas com a neuropsicóloga Ana Rita Silva, pois não são relevantes para este estudo. A explicação deste algoritmo encontra-se no conjunto de passos 3.

**Algorithm 2** Filtragem de *Frames***Input:** *Frames*- imagens faciais **Output:** *Frames*- imagens faciais (filtradas)

---

```

1: procedure
2:   facecascade  $\leftarrow$  path_HaarCascadefile
3:   for filename in folder_path do
4:     if filename Ends With “.png” or “.jpeg” or “.jpg” then
5:       image_path  $\leftarrow$  JoinPath(folder_path, filename)
6:       image  $\leftarrow$  LoadImage(image_path)
7:       gray  $\leftarrow$  converter(image)
8:       laplacian_var  $\leftarrow$  CalculateLaPlacianVariance(gray)
9:       if laplacian_var < blurthreshold then
10:        RemoveFile(image_path)
11:        Continue
12:      end if
13:      faces  $\leftarrow$  DetectFaces(gray, facecascade)
14:      if Size(faces) == 0 then
15:        RemoveFile(image_path)
16:        Continue
17:      end if
18:      (x, y, w, h)  $\leftarrow$  GetFirtCoordinates(faces)
19:      face_area  $\leftarrow$  CalculateFaceArea(w, h)
20:      image_area  $\leftarrow$  CalculateImageArea(image)
21:      if face_area <  $0.7 \times$  image_area then
22:        RemoveFile(image_path)
23:      end if
24:    end if
25:  end for
26: end procedure

```

---

---

**Algorithm 3** Seleção de imagens mais relevantes

---

**Input:** *Frames-* imagens faciais (filtradas) **Output:** *Frames-* imagens faciais (sem o rosto da neuropsicóloga)

```
1: procedure
2:   function delete_images_of_person(person_image, folder_path)
3:     person_face  $\leftarrow$  load_image(person_image)
4:     fixed_size  $\leftarrow$  (person_face.width, person_face.height)
5:     person_face  $\leftarrow$  resize_image(person_face, fixed_size)
6:     for each file in list_files(folder_path) do
7:       file_path  $\leftarrow$  join_path(folder_path, file)
8:       if is_file(file_path) and file.ends_with(“.png” or “.jpg” or “.jpeg”) then
9:         image  $\leftarrow$  load_image(file_path)
10:        image  $\leftarrow$  resize_image(image, fixed_size)
11:       end if
12:       if compare_images(person_face, image) then
13:         delete_file(file_path)
14:       end if
15:     end for
16:   end function
17:   function compare_images(image1, image2)
18:     gray1 = convert_to_grayscale(image1)
19:     gray2 = convert_to_grayscale(image2)
20:     diff  $\leftarrow$  absolute_difference(gray1, gray2)
21:     mean_diff  $\leftarrow$  calculate_mean(diff)
22:     similarity_threshold  $\leftarrow$  30
23:     if mean_diff < similarity_threshold then
24:       return true
25:     else
26:       return false
27:     end if
28:   end function
29:   Example usage:
30:   person_image  $\leftarrow$  'anarita.png'
31:   folder_path  $\leftarrow$  'frames_2306_p3'
32:   delete_images_of_person(person_image, folder_path)
33: end procedure
```

---

Para isso, utilizou-se uma função (linhas 2 a 16) que é responsável por excluir imagens que sejam semelhantes a uma imagem de referência (neste caso, a imagem da neuropsicóloga).

Primeiro são percorridos todos os arquivos da pasta em questão e é verificado se estes representam uma imagem válida, com extensão *.png*, *.jpg* ou *.jpeg* (linhas 6 a 13).

De seguida, cada imagem válida é redimensionada para que fique com o mesmo tamanho da imagem de referência, e estas são comparadas, calculando a diferença absoluta entre elas e verificando se a média dessas diferenças está abaixo de um limite predefinido, neste caso 30 (linhas 17 a 22). A escolha deste limite específico está relacionada com a natureza das imagens envolvidas. No contexto do estudo realizado, a imagem de referência pertence à neuropsicóloga que passa por mudanças significativas na sua expressão e posição durante os vídeos. Portanto, um valor mais baixo para o limite de média das diferenças (30) foi escolhido, com o objetivo de capturar com mais precisão as variações ligeiras nessas imagens, mesmo quando ocorrem mudanças perceptíveis na expressão facial e posição da mesma ao longo dos vídeos.

Por último, se as imagens forem consideradas suficientemente semelhantes, o arquivo é excluído da pasta em questão (linhas 22 a 27).

Após o processo de filtragem, todas as *frames* selecionadas foram convertidas em conjuntos de pixels, que representam a informação de cor e intensidade dos pontos das imagens. As imagens convertidas em conjuntos de pixels foram armazenadas num arquivo no formato *CSV* e a explicação do procedimento encontra-se no algoritmo 4.

---

**Algorithm 4** Imagens convertidas para pixels

---

**Input:** *Frames*- imagens faciais (relevantes) **Output:** Cojunto de pixels

---

```

1: procedure
2:   image_folder ← image_path
3:   csv_file ← csv_path
4:   desired_width ← 48
5:   desired_height ← 48
6:   image_names ← os.listdir(image_folder)
7:   With open (csv_file, 'w', newline = '\n') as file:
8:     writer ← csv.writer(file)
9:     for image_name in image_names do
10:      image_path ← os.path.join(image_folder, image_name)
11:      image ← Image.open(image_path)
12:      resized_image ← image.resize(desired_width, desired_height)
13:      grayscale_image ← ImageOps.grayscale(resized_image)
14:      pixels ← list(grayscale_image.getdata())
15:      pixels_str ← ' '.join(str(p)
16:      for p in pixels) do
17:        writer.writerow([pixels_str])
18:      end for
19:    end for
20: end procedure

```

---

Começou-se por definir as dimensões desejadas para as imagens. Escolheu-se 48 x 48 pixels,

uma vez que o modelo tem como *input* imagens com estas dimensões (linhas 4 e 5). De seguida, após converter cada imagem para a escala de cinza (linha 14), os pixels são obtidos e convertidos numa única *string*, separada por vírgulas (linha 15). Seguidamente são colocados num *csv*, linha por linha, ou seja, cada linha contém um conjunto de pixels que, conseqüentemente, representam uma imagem (linha 17).

O *dataset* final contém, então, 62378 imagens. Este é composto por duas *features* principais. A primeira coluna contém os conjuntos de pixels representando as imagens, enquanto a segunda coluna contém informações sobre a sessão e o momento específico de cada imagem.

Sessão	Número de Imagens
13/04	14773
17/04	16791
20/04	7471
24/04	7022
27/04	893
04/05	8039
23/06	4041

**Tabela 4.2:** Características do segundo *dataset* usado.

### 4.3 *Dataset de Engagement para Treino*

Uma vez que era necessário mapear este conjunto de dados com informação relativa à atenção e envolvimento (o conhecido *engagement*) presentes em cada imagem, foram aplicadas algumas técnicas para esse efeito.

#### 4.3.1 *Clustering e Elbow Method*

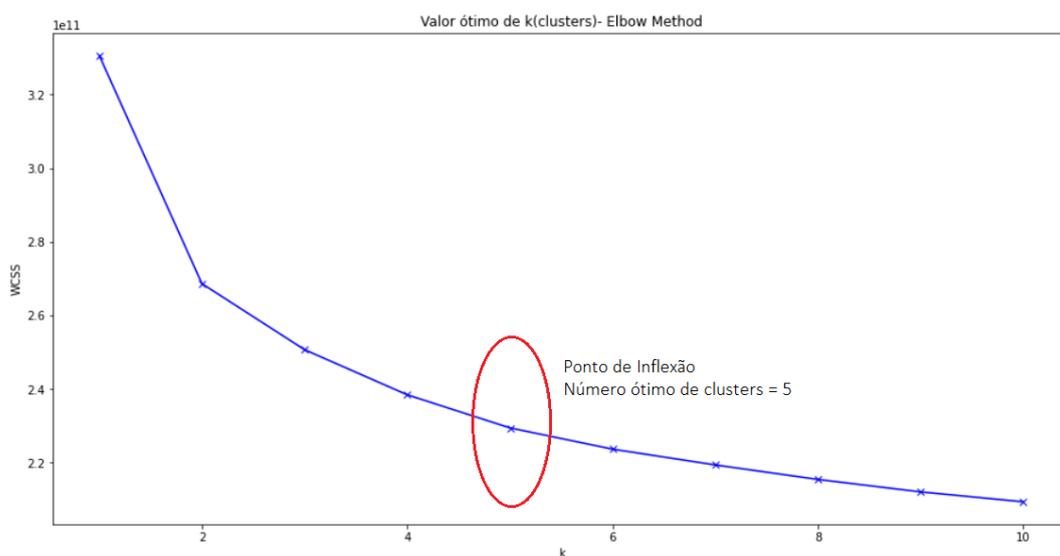
O *clustering* é uma técnica de aprendizagem automática não supervisionada. Isto é, é um tipo de técnica de aprendizagem de máquina em que não há a presença de um conjunto de dados rotulados para orientar o modelo, sendo que o objetivo é encontrar padrões e estruturas que possam ser úteis para entender os dados ou extrair informações. Assim, o *clustering* consiste em separar um conjunto de dados em grupos, cujos membros compartilham características semelhantes [61].

Para determinar o número ideal de *clusters* ou grupos a serem criados, utilizou-se o ***Elbow Method***. O objetivo desta técnica é encontrar um ponto de inflexão (ou “cotovelo”) no gráfico da soma das distâncias quadráticas intra-cluster em relação ao número de *clusters*, ou seja, a soma das distâncias ao quadrado de cada membro do *cluster* em relação ao centróide do *cluster* mais próximo. Esse ponto representa o número ideal de *clusters* em que a adição de um novo não traz grandes melhorias no desempenho do modelo. Por outras palavras, o *Elbow Method* ajuda a encontrar o equilíbrio ideal entre a variação total dos dados (que diminui à medida que se aumenta o número de *clusters*) e a simplicidade do modelo (que aumenta à medida que se reduz o número de *clusters*) [62]. Para isto, usou-se o algoritmo *KMeans*, considerado um dos

algoritmos de *clustering* mais usados devido à sua simplicidade [63].

Assim, para cada valor de  $K$  (que varia de 1 a 10), calculou-se o *Within-Cluster Sum of Square* (WCSS). Este representa a soma da distância ao quadrado entre cada ponto e o centróide num *cluster*. Esta técnica é conhecida como a técnica do “cotovelo”, pois quando é traçado o WCSS com o valor de  $K$ , o gráfico parece-se com um cotovelo. À medida que o número de *clusters* aumenta, o valor de WCSS começa a diminuir e é maior quando  $K = 1$ . A análise do gráfico é feita pela observação da repentina mudança do traço num ponto, criando assim uma forma de cotovelo. A partir deste ponto, o gráfico move-se quase paralelo ao eixo do  $X$ . O valor de  $K$  correspondente a este ponto é o valor ótimo de  $K$ , ou seja, o número ótimo de *clusters* [62, 64].

Na Figura 4.2 está representado o gráfico da soma dos quadrados das distâncias intra-cluster em função do número de *clusters*, gerado para determinar o número de *clusters* necessários para este estudo. Analisando-o é possível reparar que a soma dos quadrados intra-cluster começa a diminuir de forma menos acentuada em torno do número de *clusters* igual a 5. Essa mudança é visível no gráfico, onde o “ponto de cotovelo” é formado. Portanto, escolheram-se 5 *clusters* para a análise de *clustering*, uma vez que essa escolha maximiza a separação entre os grupos, sem adicionar *clusters* desnecessários.



**Figura 4.2:** *Elbow Method* para obtenção do número ótimo de *clusters*.

De seguida, usou-se, novamente, o algoritmo *KMeans* para identificar os centros dos *clusters* e armazená-los numa nova coluna do *dataset*, atribuindo a cada ponto de dados um rótulo de *cluster* com base na sua distância ao centro do *cluster* mais próximo.

Para verificar o rigor e precisão para os *clusters* encontrados pelo modelo *KMeans*, utilizou-se a métrica *Silhouette Score*. Esta é uma métrica utilizada para calcular a qualidade de uma técnica de *clustering*. O seu valor varia de -1 a 1, sendo que se for igual a 1, os grupos estão bem

afastados uns dos outros e distinguem-se claramente, se for igual a 0, a distância entre *clusters* não é muito significativa e se for igual a -1, os *clusters* estão atribuídos de forma incorreta [62].

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4.1)$$

A equação 4.1 representa o cálculo do *Silhouette Score* para um ponto de dados particular, onde:

- $S(i)$  é o coeficiente de *Silhouette* do ponto de dados  $i$ ;
- $a(i)$  é a distância média entre  $i$  e todos os outros pontos do *cluster* ao qual  $i$  pertence;
- $b(i)$  é a distância média de  $i$  a todos os *clusters* aos quais  $i$  não pertence.

O valor desta métrica foi de 0.099, o que indica que a separação entre os *clusters* não é muito clara, ou seja, pode haver alguma sobreposição entre pontos dos diferentes grupos. No entanto, como se trata de um conjunto de dados do mundo real, é admissível que se tenha obtido um valor baixo de *Silhouette Score*, especialmente quando a variabilidade é alta, que é o caso deste *dataset*, uma vez que apresenta imagens faciais de 7 emoções diferentes.

Posteriormente, para se atribuir a *label* de “*engaged*” ou “*not engaged*” a cada imagem pertencente a cada *cluster* diferente, procedeu-se à visualização de 4 imagens de cada *cluster*, recorrendo à biblioteca *Matplotlib*, e à análise das mesmas, percebendo se os rostos presentes em cada uma demonstravam mais a atenção ou a desatenção.

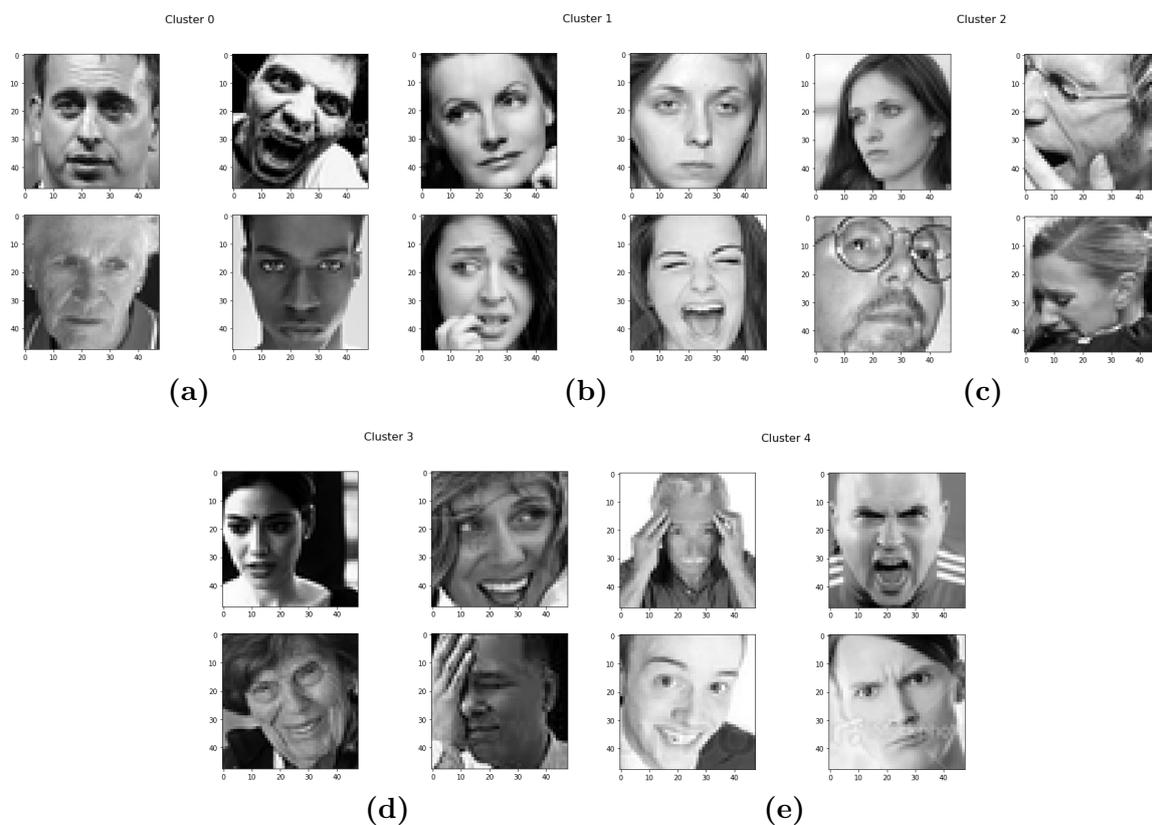
Na Figura 4.3 estão apresentadas as 4 imagens de cada *cluster*, usadas para os classificar acerca da presença de *engagement*.

Observando então as imagens de cada *cluster*, concluiu-se que: os *clusters* 0, 1 e 4 apresentavam imagens com rostos aparentemente atentos e os restantes, *clusters* 2 e 3, faces desatentas. Assim as *labels* atribuídas aos 5 *clusters* foram as seguintes:

- *Cluster* 0: *engaged*
- *Cluster* 1: *engaged*
- *Cluster* 2: *not engaged*
- *Cluster* 3: *not engaged*
- *Cluster* 4: *engaged*

Na Tabela 4.3, estão representadas as duas classes referentes ao *engagement*, bem como o número de dados correspondentes a cada uma.

Após a realização dos procedimentos acima descritos, o *dataset* encontra-se completo e pronto para ser usado para treinar o modelo desenvolvido.



**Figura 4.3:** (a) Imagens pertencentes ao *Cluster 0* (b) Imagens pertencentes ao *Cluster 1* (c) Imagens pertencentes ao *Cluster 2* (d) Imagens pertencentes ao *Cluster 3* (e) Imagens pertencentes ao *Cluster 4*

Classe	Número de Imagens
0	13671
1	22216

**Tabela 4.3:** Número de imagens de cada classe do primeiro *dataset* usado.

#### 4.4 *Dataset de Engagement para Teste*

Em virtude de mapear cada imagem com a informação relativa ao *engagement*, foi solicitada a colaboração especializada da neuropsicóloga Ana Rita Silva. Para assegurar que a colaboração não sobrecarregasse a neuropsicóloga, dadas as complexidades e demandas associadas ao trabalho, apenas lhe foi pedido que esta avaliasse 50 imagens de cada indivíduo, totalizando 300 imagens.

A seleção das imagens foi realizada de forma aleatória, visando garantir a representatividade do conjunto de dados. Entretanto, devido a discrepâncias no número de imagens disponíveis por indivíduo, o total final foi ajustado apenas para 285 imagens. Posto isto, as avaliações fornecidas foram agregadas numa nova classe, referente à classificação. Esta classificação foi considerada como referência, uma vez que a Ana Rita Silva não apenas avaliou a atenção dos pacientes, mas também conduziu as sessões do programa de prevenção, conferindo-lhe uma compreensão profunda e familiaridade com cada indivíduo, o que torna este método uma fonte confiável de informações para a criação deste conjunto de dados único e bem fundamentado.

A Tabela 4.4 apresenta o número de imagens referente a cada classe deste novo *dataset*.

Classe	Número de Imagens
0	125
1	160

**Tabela 4.4:** Número de imagens de cada classe do segundo *dataset* usado.

Neste capítulo, é descrita, detalhadamente, a origem e a criação dos dois *datasets* essenciais para a realização deste estudo. O primeiro *dataset*, composto por imagens com *labels* das sete emoções principais, passou por um processo cuidadoso de obtenção dos rótulos de *engagement*, utilizando técnicas de *clustering* e o método do *Elbow*, para identificar cinco *clusters* distintos. Posteriormente, cada quatro imagens de cada *cluster* foram observadas e mapeadas, resultando na formação deste *dataset*. O segundo *dataset* foi adquirido por meio das gravações de consultas do programa Reminder A, compreendendo sessões online de 60 minutos com seis pacientes, das quais sete foram selecionadas para análise. Foi utilizado um processo de extração e filtragem de *frames* das gravações, de maneira a obter apenas imagens relevantes. Após a conversão para conjuntos de pixels e análise criteriosa de 50 imagens de cada paciente pela neuropsicóloga Ana Rita Silva, o segundo *dataset* foi completado com rótulos de *engagement*. Ambos os conjuntos de dados fornecem imagens faciais dos participantes, embora o primeiro seja composto por imagens de várias pessoas, em geral, enquanto que o segundo se concentra especificamente nos pacientes do programa *Reminder*.

# Experimentação: Métodos, Resultados e Discussão

Este capítulo cobre todo o processo de desenvolvimento do modelo, onde serão explicadas as técnicas utilizadas para adquirir as expressões faciais dos pacientes, partindo dos dois *datasets* construídos e explicados no Capítulo 4, bem como os resultados obtidos através do desenvolvimento e aplicação do modelo produzido.

A atenção do paciente durante a fase das consultas e/ou programas de prevenção da doença em questão, é uma medida crucial para avaliar a qualidade de vida e o bem-estar em indivíduos que enfrentam desafios decorrentes de doenças neurodegenerativas.

O presente estudo propôs-se a explorar o potencial das expressões faciais como uma fonte rica de informações sobre o *engagement* dos pacientes, visando oferecer uma abordagem objetiva e não invasiva para avaliar o estado emocional e cognitivo dos mesmos. Através do treino de um modelo de inteligência artificial, foi possível capturar padrões discretos nas expressões faciais dos pacientes e correlacioná-los com o *engagement*.

Nesta secção, serão apresentados os métodos, experimentação, resultados e respetiva discussão, abordando as descobertas mais relevantes e as contribuições significativas, proporcionadas pelo modelo proposto.

A análise dos resultados permitirá a compreensão dos padrões e correlações identificados entre as expressões faciais e a atenção dos pacientes, fornecendo *insights* valiosos para os profissionais da área de saúde e familiares.

## 5.1 Divisão dos Dados

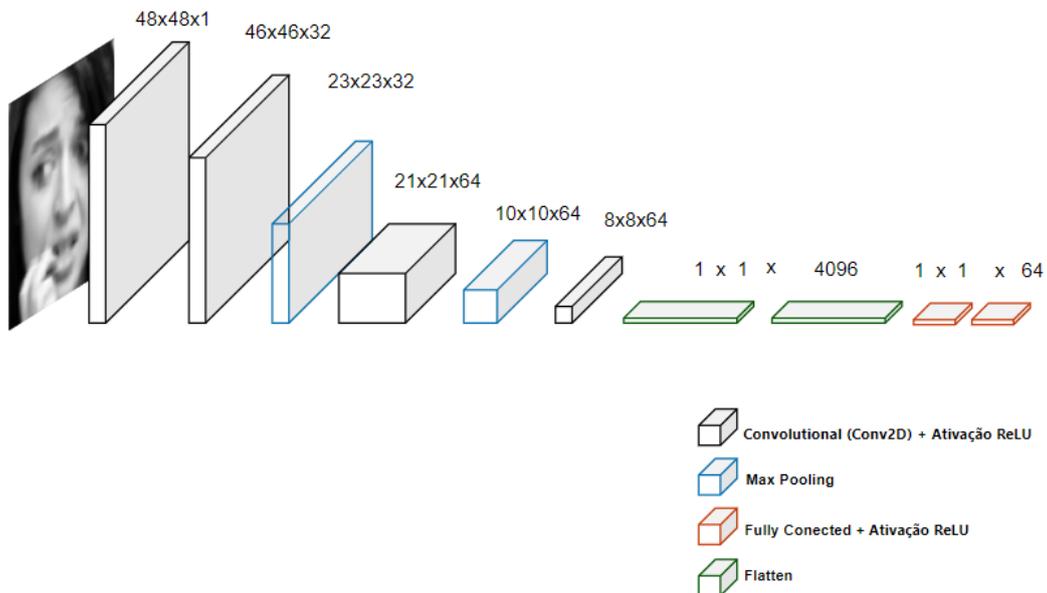
O *dataset* utilizado para esta fase foi o primeiro abordado no Capítulo 4. Os dados foram divididos em três conjuntos: treino, validação e teste, com proporções de 60%, 20% e 20%, respetivamente. O conjunto de treino é usado para ensinar o modelo, permitindo que este aprenda os padrões nos dados. A validação é utilizada para ajustar os hiperparâmetros do modelo e encontrar a melhor configuração. O conjunto de teste é reservado para avaliar o desempenho final do modelo em dados não vistos anteriormente. Esta divisão estratégica dos dados garante um treino adequado, otimização precisa e avaliação confiável do modelo.

## 5.2 Implementação e Configuração do *Visual Geometry Group (16 layers) (VGG-16)*

### 5.2.1 Arquitetura Utilizada

Como mencionado no Capítulo 2, a arquitetura VGG-16[25] foi usada como base para este estudo. No entanto, considerando a eficiência computacional do algoritmo, apenas foram utilizadas 9 camadas. A arquitetura completa da VGG-16 pode ser computacionalmente rigorosa, exigindo recursos de *hardware* consideráveis e um tempo de treino prolongado. Com as 9 camadas selecionadas, foram obtidos resultados promissores em relação à detecção de *engagement*, ao mesmo tempo em que foi otimizada a eficiência do processo de treino [65].

Na Figura 5.1 está representado um diagrama que ilustra a arquitetura da rede VGG-16 utilizada. Este diagrama fornece uma visão geral da estrutura e sequência das camadas neste modelo.



**Figura 5.1:** Diagrama da Arquitetura VGG-16 utilizada neste estudo: *Convolutional Neural Network* (CNN) para Classificação de Imagens.

### 5.2.2 Pré-Processamento dos Dados

Para garantir a consistência e a compatibilidade dos dados com a camada de entrada do modelo, todas as imagens foram redimensionadas para o tamanho de 48 pixels (48x48x1), recorrendo à biblioteca *Numpy*. Esta decisão foi tomada devido ao requisito da camada de *input* do modelo, que espera receber dados nesse formato específico.

De maneira a assegurar a compatibilidade com o modelo, os conjuntos de dados de pixels e as respectivas *labels* foram convertidos em tensores de ponto flutuante de 32 bits (*float32*),

utilizando a biblioteca *TensorFlow*. Esta conversão permite que os dados sejam processados de forma eficiente e garante uma representação numérica precisa e adequada para os cálculos necessários durante o treino.

### 5.2.3 *Early Stopping*

Durante o treino do modelo, foi utilizada a técnica do *Early Stopping* [66], com o objetivo de evitar o *overfitting* e melhorar o desempenho geral. O *Early Stopping* monitora o desempenho do modelo num conjunto de validação, durante o treino, e interrompe o processo quando o desempenho começa a piorar, salvando a melhor performance obtida. Esta abordagem foi implementada neste estudo, onde o *Early Stopping* interrompeu o treino nas 31 *epochs*, garantindo que o modelo alcançasse um bom desempenho sem cair no *overfitting*. O uso desta técnica foi fundamental para obter resultados confiáveis e um modelo que generalize bem para dados não vistos, economizando recursos computacionais e tempo de treino [66].

### 5.2.4 Parâmetros de Treino

No contexto deste estudo, a seleção dos parâmetros de treino desempenha um papel crucial no desenvolvimento de um modelo eficaz para a detecção de *engagement* através do reconhecimento facial. Com o objetivo de otimizar o desempenho e a precisão do modelo, foram utilizados parâmetros específicos durante esta fase.

#### *Adam*

O otimizador *Adam* [67] foi escolhido devido à sua capacidade de ajustar automaticamente a taxa de atualização dos parâmetros durante o treino, combinando as vantagens do algoritmo de *Stochastic Gradient Descent* (SGD) com técnicas adaptativas, resultando numa convergência eficiente na aprendizagem de padrões e características relevantes das expressões faciais. As fórmulas utilizadas para o algoritmo *Adam* foram baseadas no trabalho de *Kingma e Ba* [68].

#### *Binary Crossentropy*

A função de perda *Binary Crossentropy* [69] foi selecionada para este problema de classificação binária, devido à sua eficiência na medição da divergência entre as previsões do modelo e os rótulos verdadeiros, incentivando o modelo a aprender a distribuição correta dos rótulos e melhorar o seu desempenho na classificação. As suas equações podem ser consultadas em [70].

## 5.3 Teste Experimental I

O primeiro teste experimental consiste em avaliar o desempenho do modelo previamente desenvolvido utilizando o primeiro *dataset*, conforme detalhado no Capítulo 4.

Para analisar o comportamento do modelo ao longo das *epochs* e identificar possíveis casos de *overfitting*, foram gerados gráficos de *accuracy* de treino e validação, bem como de *loss* de

treino e validação. Esta escolha gráfica fornece uma visão abrangente do processo de aprendizagem do modelo e permite uma análise aprofundada do seu comportamento ao longo do tempo.

No conjunto de Figuras 5.2 estão representadas, graficamente, as métricas da *accuracy* e *loss* de treino e validação, ao longo das *epochs*, durante o treino do modelo.

Observando o gráfico da Figura 5.2a, é notório que a *loss* de treino, apesar de apresentar valores baixos, mantém-se praticamente constante ao longo das *epochs*, enquanto que a *loss* de validação começa e termina com valores elevados, apresentando oscilações evidentes.

Num modelo ideal, tanto os dados de treino (já conhecidos) como os de validação (novos dados), deveriam apresentar curvas de *loss* descendentes ao longo das *epochs* e manter-se com valores não muito discrepantes entre si. No entanto, o comportamento observado sugere que o modelo pode estar a sofrer de *overfitting*, o que pode impactar negativamente o seu desempenho em dados não vistos anteriormente.

Embora tenha sido identificada a presença de *overfitting*, foram também calculadas métricas adicionais, como *f1-score*, *recall* e *precision*, explicadas no Capítulo ??, para avaliar o desempenho geral do modelo e a capacidade de generalização em relação aos dados de teste.

Modelo	Accuracy	Loss	F1-Score	Precision	Recall
Modelo Inicial	89.87%	53.61%	91.69%	92.92%	90.51%

**Tabela 5.1:** Resultado das métricas do primeiro teste experimental realizado, com o modelo inicial.

Embora as métricas calculadas tenham apresentado valores bastante favoráveis, é importante destacar que a métrica da *loss* de teste também se mantém alta. Deste modo, é necessário realizar alterações no modelo com o objetivo de reduzir a *loss* e eliminar o *overfitting*.

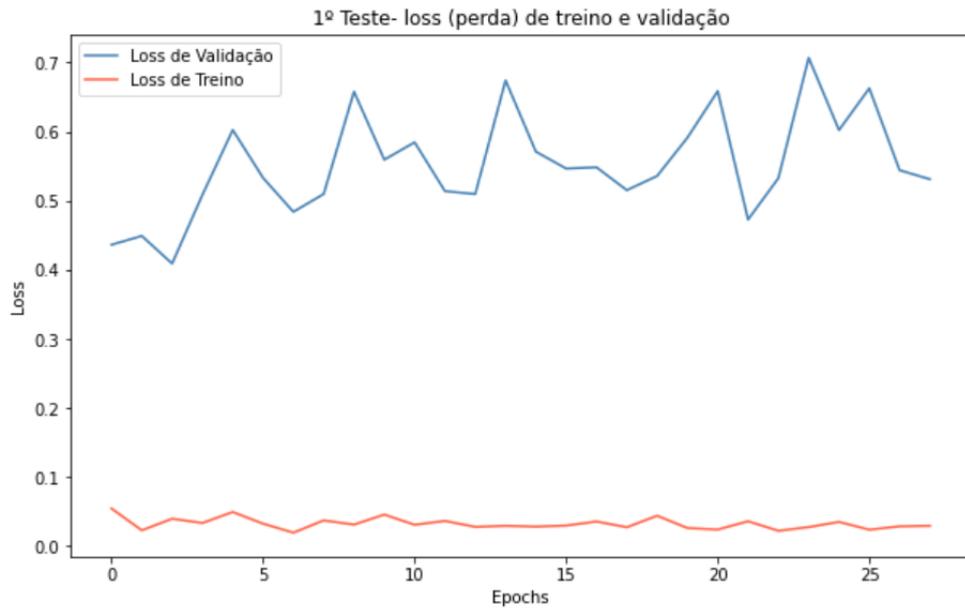
### 5.3.1 Redução/Eliminação de *Overfitting*

Como mencionado no Capítulo 2, têm sido exploradas várias técnicas para mitigar ou reduzir o *overfitting* em modelos de *Machine Learning*. Dentro dessas técnicas, destacam-se o *dropout*, a regularização de pesos *L1* e *L2* e o *k-fold cross-validation*. Nesta subsecção, será apresentada, detalhadamente, a aplicação desses métodos e respetivos resultados obtidos.

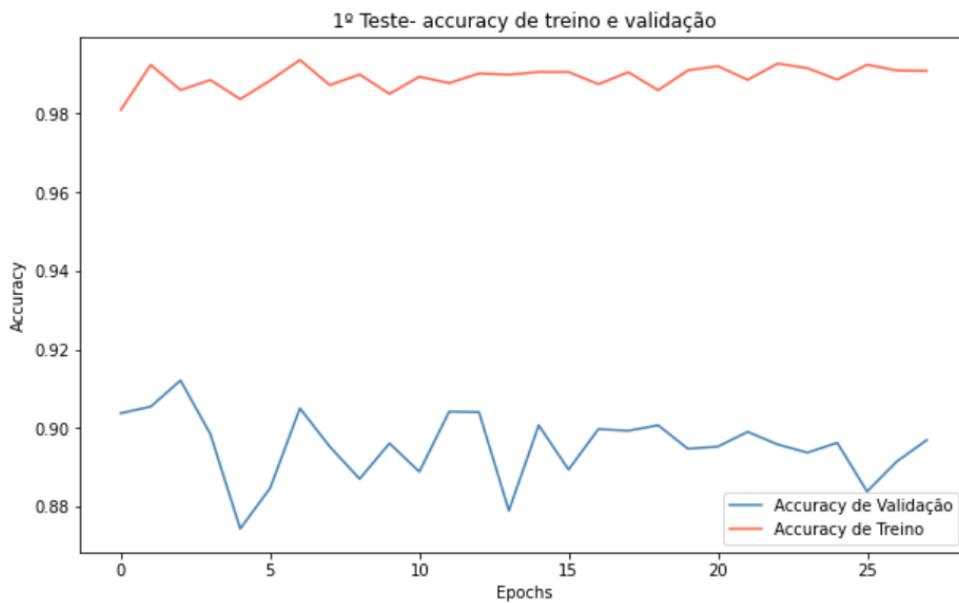
#### *Dropout*

Para se procurar encontrar a configuração mais adequada para o modelo em questão, foram testados diferentes valores de *dropout*.

Realizaram-se três testes experimentais. No primeiro, foram aplicadas taxas de *dropout* de 25% e 50% nas camadas ocultas (*hidden*) e densas (*dense*), respetivamente. No segundo, as



(a) *Loss* de treino e validação ao longo das *epochs*.



(b) *Accuracy* de treino e validação ao longo das *epochs*.

**Figura 5.2:** Resultados de treino e validação do 1º teste.

taxas de *dropout* foram reduzidas para 15% e 25% e por fim, no terceiro teste experimental, foram utilizadas taxas mais baixas, de 5% e 15%, respetivamente.

Nas Figuras C.1, C.2 e C.3, presentes no Apêndice C, estão representadas, graficamente, as perdas (*loss*) de treino e validação, após a aplicação das diferentes taxas de *dropout*, respetivamente.

Observando o gráfico a Figura C.1, é notória a grande aproximação (quase sobreposição) das curvas de treino e validação, apesar das oscilações apresentadas na linha de treino. No entanto, os valores de perda são bastante elevados em ambas as curvas, o que denota a possível existência de *underfitting*. O *underfitting* ocorre quando o modelo não é capaz de aprender a partir dos dados de treino e, portanto, não se consegue ajustar bem aos dados, nem captar os seus padrões. Como resultado, tanto a *loss* de treino como a *loss* de validação são altas e próximas uma da outra, indicando que o modelo não está adaptado nem mesmo aos dados de treino.

Analisando o gráfico da Figura C.2, relativo à segunda aplicação da técnica de *dropout*, é evidente que o *overfitting* se mantém, uma vez que, embora a curva de treino continue a descer gradualmente até um valor baixo, a curva de validação possui bastantes picos e termina num valor relativamente alto.

Por fim, examinando a Figura C.3, alusiva ao teste de *dropout*, utilizando 5% e 15% para as camadas *hidden* e *dense*, respetivamente, pode-se concluir que o *overfitting* se mantém. É visível que a curva de treino começa num valor alto, sofre um pico para um valor extremamente elevado e depois retorna ao valor inicial, enquanto que a curva de validação permanece praticamente constante e próxima da curva de treino. Esta observação sugere que o modelo não está a generalizar bem e não tem a capacidade de capturar os padrões relevantes que se aplicariam a novos dados.

### Regularização de Pesos

Foram realizados quatro testes com a regularização de pesos L1 e L2. Primeiramente, foi utilizado o valor de 0.001 tanto para os pesos L1 como para os pesos L2. De seguida, o valor de 0.0001 foi o escolhido, também para ambos os pesos.

#### 5.3.1.2.1 Treino com Regularização L1

No primeiro teste com o peso L1, visualizando o gráfico da Figura D.1 presente no Apêndice D, observa-se que as curvas de *loss* de treino e validação permanecem demasiado próximas uma da outra ao longo das *epochs* e os valores de *loss* finais encontram-se relativamente baixos. Esta avaliação sugere que, aparentemente, existe a presença de *underfitting*.

No entanto, é importante considerar outras métricas como *f1-score*, *recall* e *precision*, que fornecem uma visão mais abrangente do desempenho do modelo em relação a um problema específico. Assim, analisando os valores da primeira linha da Tabela D.1, embora os valores de *accuracy* e *loss* sejam considerados favoráveis, os resultados das restantes métricas encontram-se bastante baixos, o que poderá ser um indicativo de que o modelo não é capaz de aprender com

eficácia as relações complexas presentes nos dados e, como resultado, possui um desempenho fraco, tanto nos dados de treino como nos dados de validação ou teste, confirmando a presença de *underfitting*.

Como foi verificado que o valor do peso de regularização era demasiado alto, uma vez que foi detetado um possível *underfitting*, testou-se o treino do modelo com um valor de penalização mais baixo (0.0001). Observando o gráfico da Figura D.2, este apresenta uma semelhança notável com o obtido após a penalização anterior. As curvas de treino e validação continuam próximas, com algumas oscilações, e os valores de *loss* são relativamente baixos. Porém, os valores das restantes métricas, apresentados na segunda linha da Tabela D.1, continuam baixos revelando que, provavelmente, a regularização de pesos *L1* não será a melhor técnica para eliminar o *overfitting* presente no modelo.

### 5.3.1.2.2 Treino com Regularização *L2*

Seguidamente, realizaram-se dois treinos com regularização do peso *L2*. Após o primeiro treino, com penalização de 0.001, obteve-se o gráfico exibido pela Figura E.1, contida no Apêndice E. Analisando-o, pode-se afirmar que a curva de treino apresenta um bom ajuste, no entanto, é notória a grande oscilação presente na curva de validação. Esta observação pode levar à afirmação de que o *dataset* de validação se classifica como não representativo, ou seja, que não fornece informações suficientes para avaliar a capacidade de generalização do modelo. Os valores das outras métricas consideradas, encontram-se na primeira linha da Tabela E.1. Os resultados são ligeiramente melhores que os anteriores, no entanto, ainda são bastante baixos.

Posteriormente, à semelhança do que foi feito na subsecção anterior, realizou-se um treino com penalização de 0.0001. O gráfico de *loss* de treino e validação resultante encontra-se na Figura E.2. Este é idêntico ao gráfico derivado do treino anterior e o mesmo acontece com os valores das métricas, apresentados na segunda linha da Tabela E.1.

Após a análise dos resultados obtidos através dos testes de regularização *L1* e *L2*, tornou-se evidente que o problema do *overfitting* não foi completamente resolvido. Além disso, os resultados indicam que o conjunto de dados de validação pode não ser totalmente representativo, o que pode impactar a generalização adequada do modelo.

Após estas constatações, uma abordagem promissora para aprimorar o desempenho do modelo e garantir uma avaliação mais robusta é o uso do *K Fold Cross Validation*. Esta técnica oferece a vantagem de avaliar o modelo em diferentes conjuntos de dados de validação, permitindo uma avaliação mais confiável, no que toca à sua capacidade de generalização em diferentes situações. Na subsecção seguinte, será apresentado em detalhe como o *K Fold Cross Validation* foi aplicado [37].

### *k-Fold Cross Validation*

Em seguida, foi aplicada a técnica do *K-Fold Cross Validation*. Utilizaram-se 10 *folds* ( $k=10$ ), pois é um valor amplamente utilizado em modelos de *Machine Learning*, de maneira

a dividir o conjunto de dados em partes iguais. Após o treino, foram gerados 10 conjuntos de resultados em formato *CSV*, contendo informações sobre *loss* e *accuracy* de teste e *loss* e *accuracy* de validação. Entretanto, para a seleção final do modelo, optou-se por considerar apenas a *fold* com os melhores resultados, utilizando a métrica de *loss* de validação como critério para escolha. Desta forma, assegurou-se que o modelo final fosse aquele com melhor capacidade de generalização para novos dados.

Todavia, a *loss* de validação ainda apresenta algumas oscilações e variações. Estas podem dever-se à quantidade desproporcional de exemplos entre as classes, uma vez que o *dataset* é não balanceado, afetando a estabilidade e a capacidade de generalização do modelo.

A fim de mitigar este efeito e melhorar o desempenho do modelo, foi aplicada uma técnica de balanceamento de dados: *oversampling*.

Apesar desta técnica ter sido mencionada no Capítulo 2, optou-se por não testar a *undersampling* neste estudo. Embora o *undersampling* seja uma abordagem válida para equilibrar classes desproporcionais, algumas razões, como as descritas no Capítulo 2, conduziram à decisão de priorizar outras estratégias, como o *oversampling*.

Uma vez que o conjunto de dados utilizado apresenta uma discrepância considerável no número de instâncias entre as classes e, considerando o tamanho limitado da classe minoritária, a aplicação do *undersampling* poderia resultar numa redução ainda maior do conjunto de treino, podendo afetar a capacidade de generalização eficaz do modelo para novos dados [33, 34].

Dado o cenário de desequilíbrio e as limitações inerentes ao *dataset* utilizado, preferiu-se explorar técnicas de *oversampling*, que procuram criar novas instâncias sintéticas da classe minoritária. Assim, confia-se que esta técnica oferece um método mais promissor de melhorar a representação da classe minoritária e fortalecer o modelo no que diz respeito à sua capacidade de lidar com o desafio de classes desproporcionais.

### 5.3.2 Tratamento de Dados Não Balanceados

#### *Oversampling*

Como descrito no Capítulo 2, o *oversampling* é uma técnica que aumenta artificialmente o número de instâncias da classe minoritária num conjunto de dados não balanceados, com a finalidade de melhorar o desempenho de modelos de *Machine Learning*.

Para se conseguir alcançar uma melhor performance do modelo em questão, combinaram-se duas técnicas importantes: *k-Fold Cross Validation* e *oversampling*.

O número de *folds* selecionado para esta avaliação manteve-se o mesmo em relação ao teste anterior e a técnica escolhida para aplicação do *oversampling* foi o *Synthetic Minority Over-sampling Technique* (SMOTE), explicado e detalhe no Capítulo 2.

O algoritmo separa os dados de treino e validação para cada dobra e, assim, o *oversampling* é aplicado apenas ao conjunto de treino, usando a técnica SMOTE.

As Figuras 5.3a e 5.3b representam graficamente a *loss* e a *accuracy* de treino e validação, ao longo das *epochs* e a Tabela 5.2 apresenta os valores das métricas escolhidas, aplicadas ao conjunto de teste.

<b>Modelo</b>	<i>Accuracy</i>	<i>F1-Score</i>	<i>Precision</i>	<i>Recall</i>
<i>K-Fold Cross Validation + Oversampling</i>	97.32%	92.31%	92.80%	91.82%

**Tabela 5.2:** Resultado das métricas do primeiro teste experimental realizado, com o modelo após a aplicação de *K-Fold Cross Validation* e *Oversampling*.

Observando o gráfico da Figura 5.3a, é notório que o valor da *loss* de validação se encontra ligeiramente mais baixo do que o da *loss* de treino. Geralmente, é esperado ver a *loss* de treino diminuir mais rapidamente, indicando que o modelo está a aprender com os dados de treino. No entanto, uma *loss* de validação ligeiramente mais baixa sugere que o modelo está a generalizar bem para dados não vistos, o que constitui uma característica positiva.

O decréscimo das duas curvas de *loss* ao longo das *epochs* constitui um padrão satisfatório, indicando que o modelo está a ajustar os seus parâmetros, de maneira a se alinhar melhor com os padrões dos dados. O pico visualizado cerca da *epoch* 56 poderia constituir uma preocupação, no entanto a *loss* estabiliza rapidamente, o que sugere que o modelo consegue superar o obstáculo e continuar a aprender.

O mesmo raciocínio pode ser aplicado ao analisar o gráfico da Figura 5.3b, mas com um ponto de vista inverso. Enquanto que a *loss* decresce ao longo das *epochs*, a *accuracy* tende a aumentar, ilustrando a capacidade do modelo de fazer previsões mais precisas à medida que se torna mais experiente com os dados de treino.

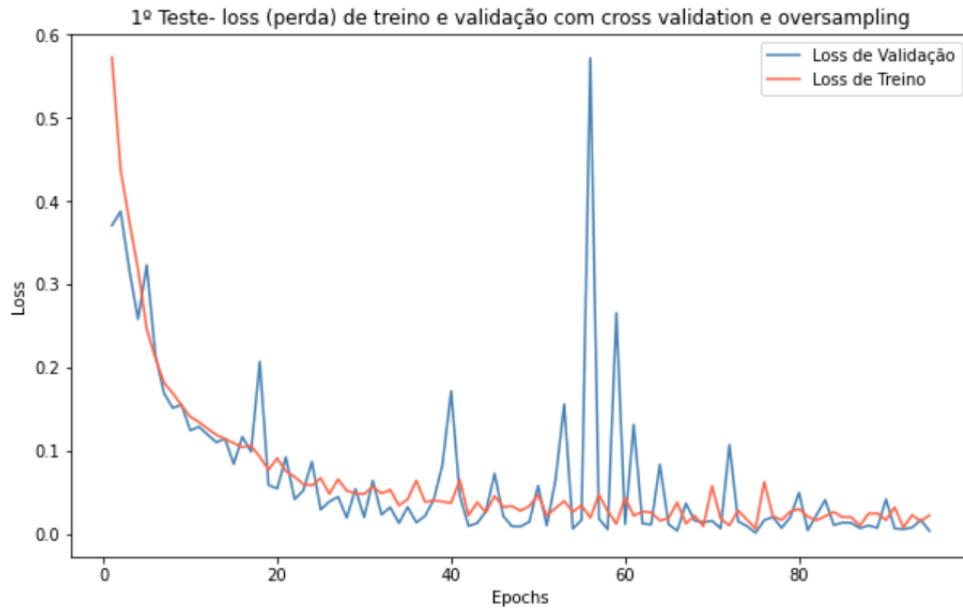
Examinando as métricas de avaliação indicadas na Tabela 5.2, é possível concluir que estas indicam que o modelo realiza classificações precisas e confiáveis, e apresenta bons resultados para dados não vistos durante o treino.

Combinando a análise dos gráficos da Figura 5.3 com os valores das métricas de avaliação, apresentadas na Tabela 5.2, no geral, apesar da estrutura das curvas de *loss* ser considerada atípica, a combinação das métricas de avaliação positivas e do comportamento geral das curvas, sugere que o seu modelo está bem treinado e é capaz de reconhecer, eficazmente, o *engagement*, através do reconhecimento facial.

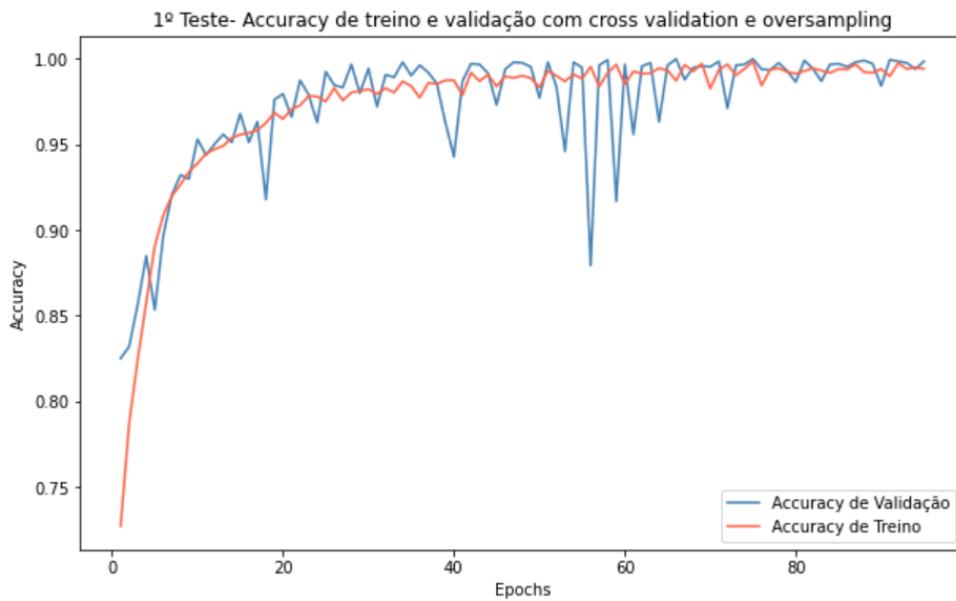
## 5.4 Teste Experimental II

O segundo teste experimental consiste em analisar e avaliar a performance do modelo treinado anteriormente, utilizando o segundo *dataset*, conforme especificado no Capítulo 4.

À semelhança do teste feito no *Teste Experimental I*, também foram consideradas as métricas da *accuracy*, *f1-score*, *precision* e *recall*.



(a) *Loss* de treino e validação ao longo das *epochs*, após a aplicação de *K-Fold Cross Validation* e *Oversampling*.



(b) *Accuracy* de treino e validação ao longo das *epochs*, após a aplicação de *K-Fold Cross Validation* e *Oversampling*.

**Figura 5.3:** Resultados de treino e validação, após a aplicação de *K-Fold Cross Validation* e *Oversampling*.

Modelo	Accuracy	F1-Score	Precision	Recall
<i>K-Fold Cross Validation + Oversampling</i>	45.96%	45.39%	52.46%	40.25%

**Tabela 5.3:** Resultado das métricas do segundo teste experimental realizado, com o modelo após a aplicação de *K-Fold Cross Validation* e *Oversampling*.

As métricas de avaliação estão apresentadas na Tabela 5.3. Estas fornecem *insights* sobre o desempenho do seu modelo em diferentes aspetos.

Observando os valores das métricas, é possível afirmar que o modelo teve um desempenho desafiador no novo conjunto de dados reais. A combinação do *F1-score*, *Precision*, *Recall* e *Accuracy* oferece uma visão abrangente das capacidades e limitações do modelo em relação a diferentes categorias de previsões.

O resultado destas métricas indica que o modelo enfrenta dificuldades em alcançar um equilíbrio ideal entre a taxa de previsões corretas e a capacidade de identificar exemplos verdadeiros positivos. Além disso, este algoritmo não consegue capturar cerca de 40% dos exemplos positivos reais, demonstrando, mais uma vez, a sua limitação em identificar verdadeiros positivos.

Com o objetivo de avaliar, de uma maneira mais aprofundada, o desempenho deste modelo de classificação binária, para além das métricas discutidas anteriormente, a curva *Receiver Operating Characteristic* (ROC) e a *Area Under the Curve* (AUC) também foram consideradas, uma vez que constituem ferramentas importantes para avaliar o desempenho deste tipo de modelo.

Na Figura 5.4 está representada, graficamente, a curva ROC.

Analisando atentamente o gráfico da Figura 5.4, é possível observar que a curva associada ao modelo de classificação encontra-se notavelmente próxima da linha diagonal, que representa a abordagem aleatória. Além disso, é perceptível que, em certos pontos, a curva representativa do modelo se posiciona abaixo dessa linha.

O valor da AUC, calculado para a Curva ROC, o qual regista 0.46, fornece, evidentemente, que o modelo desenvolvido está a enfrentar desafios ao diferenciar as classes positivas das classes negativas. Este valor indica que a capacidade do modelo em fazer previsões corretas está abaixo do nível desejado.

Estas interpretações reforçam as conclusões anteriores, que indicam um desempenho aquém do esperado para este modelo. A Curva ROC e o AUC oferecem uma visão tangível dessa dificuldade em separar as classes de maneira eficiente.

Na Figura 5.5 está representada a matriz de confusão que, como mencionado no Capítulo x, constitui uma ferramenta essencial na avaliação do desempenho de modelos de classificação. Ao analisar os valores presentes na matriz, é possível tirar conclusões valiosas acerca da capacidade do modelo em classificar as instâncias de forma precisa.

Observando a diagonal principal da matriz, que reflete os verdadeiros positivos e verdadeiros negativos, é possível identificar que o modelo acertou 67 instâncias classificadas como positivas

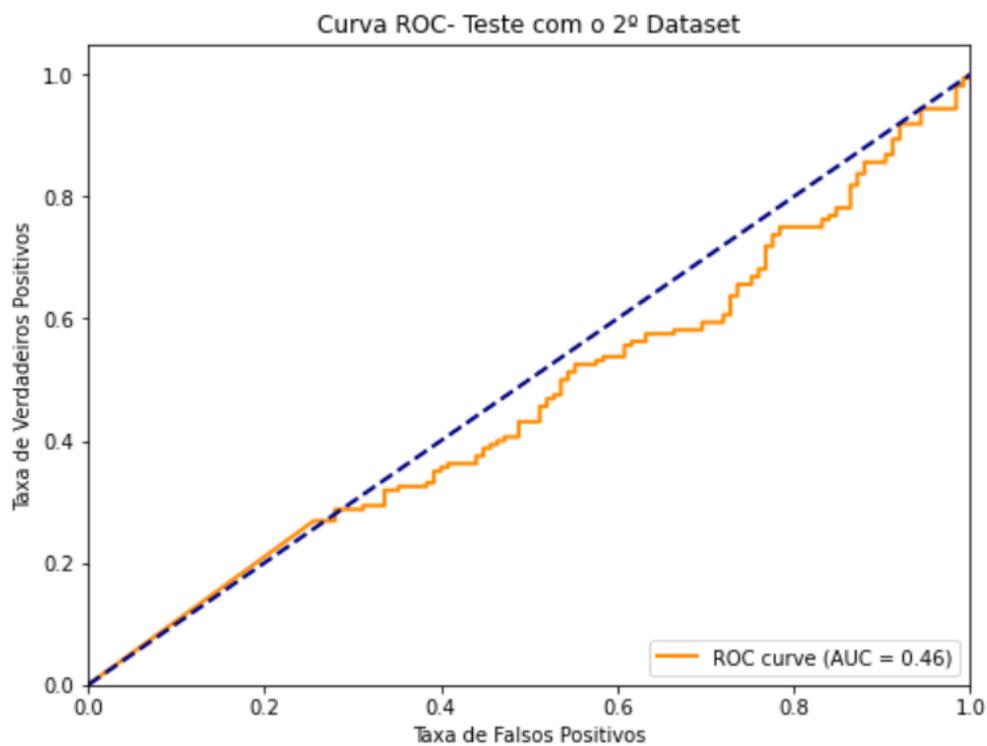


Figura 5.4: Curva ROC após o segundo teste experimental.

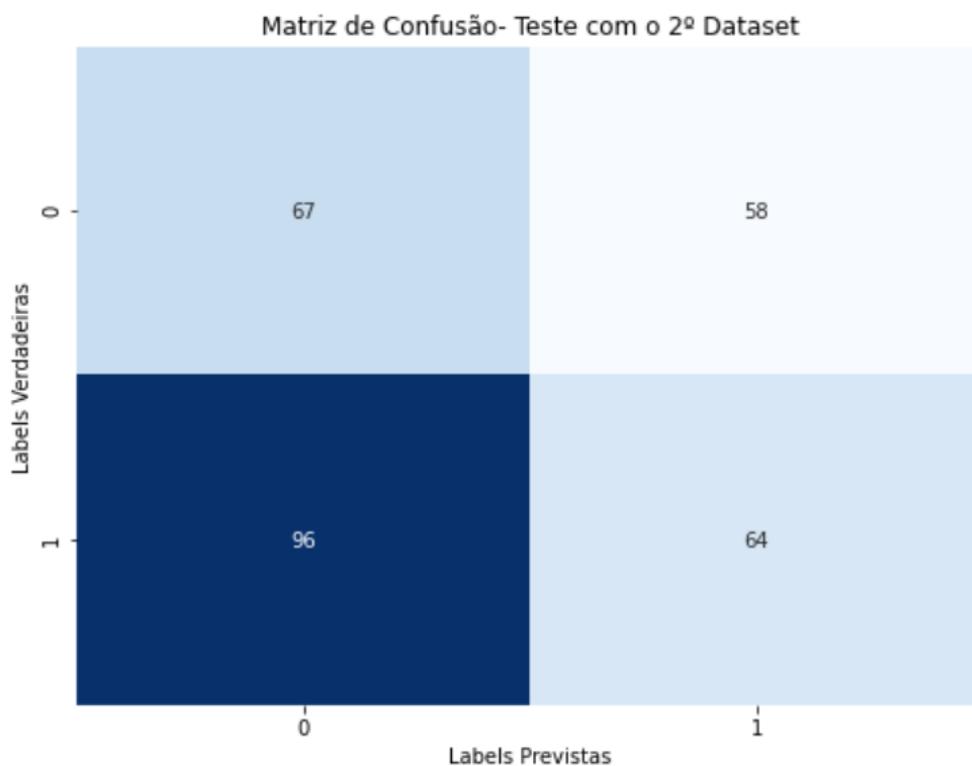


Figura 5.5: Matriz de Confusão após o segundo teste experimental.

e 64 instâncias classificadas como negativas. No entanto, quando são examinadas as outras células da matriz, que representam os falsos positivos e falsos negativos, identifica-se que houve 96 casos em que o modelo classificou erradamente como positivas as instâncias que eram, na verdade, negativas, e 58 casos em que classificou, também de maneira errada, como negativas as instâncias que eram, na verdade, positivas.

Esses resultados ilustram, mais uma vez, que este modelo está a ter dificuldades ao distinguir entre as classes positivas e negativas. A assimetria nas taxas de falsos positivos e falsos negativos é indicativa de um desequilíbrio no desempenho do modelo, o que é coerente com as constatações obtidas a partir das métricas anteriores.

Uma análise detalhada das discrepâncias nas métricas de avaliação entre os dois testes experimentais levanta importantes considerações sobre os fatores subjacentes a esses resultados divergentes. É fundamental observar que a principal distinção entre os testes, reside na utilização exclusiva do segundo *dataset* - o conjunto de dados real - para fins de teste, sem qualquer integração no processo de treino do modelo.

A decisão de empregar o segundo *dataset* apenas para fins de teste implica que o modelo foi treinado somente com um conjunto de dados retirado da *internet*, ou seja, nenhuma imagem do segundo *dataset* foi incorporada no processo de treino. Esta abordagem visa avaliar a capacidade do modelo de generalizar para dados não vistos durante o treino, num cenário que simula as condições do mundo real.

Outro aspeto crucial a ser considerado é o tamanho do conjunto de dados empregado no segundo teste. O conjunto de teste abrange cerca de 300 imagens, o que revela uma quantidade relativamente pequena. É importante destacar que a natureza restrita do conjunto de teste pode ter tido um impacto nas métricas observadas. O motivo pode dever-se ao facto de o modelo não ter tido a oportunidade de ser exposto a uma diversidade representativa de exemplos, o que é essencial para que este possa generalizar os padrões complexos presentes nas imagens. Desta forma, é fundamental compreender que as métricas derivadas desse conjunto de teste podem não refletir integralmente o desempenho potencial do modelo em situações mais diversificadas e realistas. O mesmo se deduz pelo facto de as classes não apresentarem o mesmo número de exemplos. Quando existe um desequilíbrio significativo entre as classes, as métricas podem ser influenciadas, uma vez que o modelo pode ter tendência para prever a classe maioritária com mais frequência, o que pode afetar a *precision* e o *recall*.

Importa ressaltar que a construção do segundo *dataset* foi feita em colaboração com a neuropsicóloga Ana Rita Silva, conforme abordado no Capítulo 4. Dada a meticulosidade e o esforço necessários para avaliar e atribuir rótulos às imagens, foi possível considerar e avaliar somente 283 imagens desta nova coleção.

Neste capítulo, é abordada a divisão dos dados em treino, validação e teste (60%, 20% e 20%, respetivamente), utilizando uma versão simplificada da arquitetura VGG-16 com nove camadas, considerando restrições computacionais. É implementada a técnica do *Early Stopping*, para mitigar o *overfitting* (onde o treino é interrompido após 31 *epochs*) e são utilizados, como

parâmetros de treino, o otimizador *Adam* e a *Binary Crossentropy*, como função de perda. São realizados dois testes experimentais: o primeiro com o primeiro conjunto de dados, que, embora tenha gerado métricas promissoras, revelou *overfitting*. Para o combater, foram aplicadas soluções como *dropout*, regularização de pesos L1 e L2 e *K-fold Cross Validation*, obtendo melhores resultados com esta última técnica. Devido ao desequilíbrio nos dados, recorreu-se ao uso de *oversampling* (SMOTE), conseguindo gerar métricas em torno de 90%.

No entanto, o segundo teste experimental, realizado com o segundo conjunto de dados, apresentou resultados menos satisfatórios. Isto deveu-se ao desafio de lidar com dados não balanceados, onde havia um número significativamente maior de pacientes atentos do que de pacientes não atentos. Além disso, o tamanho do conjunto de dados de teste era muito pequeno, o que influenciou negativamente as métricas de desempenho. Esses fatores - a exclusividade do segundo *dataset* para testes e a limitação do conjunto de teste - fornecem um contexto crucial para a interpretação das métricas de avaliação discrepantes. A sua compreensão enriquece a análise dos resultados experimentais, destacando a importância da composição e dimensão dos conjuntos de dados ao avaliar o desempenho do modelo em cenários reais.

## Conclusão e Trabalhos Futuros

O presente estudo tem como objetivo primordial detetar os níveis de *engagement*, em indivíduos com demência, por meio da análise das suas expressões faciais, durante a fase de tratamento. Esta etapa é desafiadora para os profissionais de saúde visto que, manter o interesse dos pacientes ao longo desse processo, frequentemente, revela-se complexo.

Para atingir este propósito, foi desenvolvido um modelo de Inteligência Artificial (IA), baseado na arquitetura *Visual Geometry Group (16 layers)* (VGG-16), capaz de identificar níveis de *engagement*, a partir de expressões faciais. Dois testes experimentais foram conduzidos para validar a eficiência deste modelo. No primeiro, o modelo foi treinado e testado com um conjunto de dados emocionais, retirados da *internet*. Já o segundo teste experimental, avaliou o modelo previamente treinado com um conjunto de dados reais obtidos a partir do programa *Reminder*, um projeto de prevenção para a saúde mental. Este último conjunto de dados foi composto pela gravação de sessões realizadas na plataforma *Zoom*, com os *frames* de cada paciente, extraídos e analisados. As avaliações das imagens foram realizadas por uma neuropsicóloga responsável pelo projeto, assegurando a confiabilidade dos rótulos associados ao *engagement*.

Os resultados do primeiro teste experimental foram altamente positivos, com métricas de desempenho, como *F1-score*, *precision* e *recall*, excedendo a marca de 90%. No entanto, os resultados do segundo teste não foram tão favoráveis, atingindo cerca de 40%/50%, nestas mesmas métricas.

Os resultados menos satisfatórios do segundo teste experimental, podem ser atribuídos a diversos fatores que influenciaram as métricas. A limitação de dados e o desequilíbrio entre as classes afetaram a precisão dos resultados. Além disso, a ausência de treino com as imagens reais, também contribuiu para um desempenho inferior.

A investigação, inicialmente, idealizou incluir um terceiro teste experimental, que consistiria no treino do modelo com a fusão dos dois conjuntos de dados. Contudo, a complexidade e a demora na obtenção e processamento das imagens, inviabilizaram a realização deste teste adicional. Como sugestão para futuros aprimoramentos deste estudo, aconselha-se a condução desse terceiro teste experimental, que poderia ser implementado ao combinar os dois conjuntos de dados e treinar o modelo com essa união. Também é recomendável o uso de um conjunto de imagens reais mais abrangente para testes, o que possivelmente contribuiria para a melhoria do desempenho do modelo.

Adicionalmente, uma sugestão para futuros estudos, seria a análise do *engagement* individual de cada paciente, ao longo das sessões, considerando o *dataset* inicialmente dividido por sujeitos. Isso poderia proporcionar *insights* relevantes sobre a evolução de cada paciente ao longo das sessões e comprovar o sucesso das intervenções propostas.

Em síntese, esta pesquisa avançou na procura por soluções inovadoras para a detecção de *engagement* em pacientes com demência. Embora haja espaço para aprimoramentos, os resultados alcançados até ao momento indicam que a análise de expressões faciais com modelos de IA pode ser um recurso valioso no contexto do tratamento de pacientes com demência.

Da mesma forma que as máquinas codificam aspetos complexos, como a saúde mental, diabetes e problemas cardiovasculares, através de números e algoritmos, a conquista da IA no domínio dos veículos autónomos, ressoa com o potencial sucesso nas áreas de saúde. Ao observar os triunfos alcançados na condução autónoma, uma das aplicações mais exigentes da IA, é possível antecipar um resultado semelhante em relação aos desafios de saúde. Afinal, se a IA pode dominar o mundo das estradas sem motoristas, é plausível esperar que possa também oferecer avanços significativos para a saúde. Esta sinergia entre a tecnologia e os cuidados de saúde sinaliza um futuro promissor e aberto a inovações, que podem impactar positivamente a vida do ser humano.

# Referências

- [1] A. Silva, C. Baptista, S. Pinho, M. Lima, A. Faria, and R. Afonso, “The reminder program – protocol for a randomized controlled trial of a neuropsychological group intervention for dementia risk reduction,” under review.
- [2] A. Silva, S. Pinho, M. Lima, A. Faria, R. Afonso, C. Baptista, J. Cerejeira, I. Baldeiras, A. Silva-Spínola, and L. Macedo, “Validation trial of a multicomponent neuropsychological interventions for individuals at risk of dementia: the reminder program.” Poster presented at the MIA Ageing Spring School (4th to 8th April 2022), 2022.
- [3] A. Silva, M. Lima, R. Afonso, and M. Pinho, *Programa REMINDER – Manual para Dinamizadores*. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, 2023.
- [4] P. R. Norvig and S. A. Intelligence, “A modern approach,” *Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems*, vol. 90, pp. 33–48, 2002.
- [5] J. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, 1980.
- [6] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, “A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities,” *Journal of Innovation & Knowledge*, vol. 8, no. 1, p. 100333, 2023.
- [7] N. Jahan, I. B. Rashid, O. Al Numan, A. T. Hasan, and N. Begum, “Collaborative ai in smart healthcare system,” in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, pp. 1–5, IEEE, 2021.
- [8] A. Toffler, *A Terceira Onda*. Record, 1981.
- [9] C. Pramerdorfer and M. Kampel, “Facial expression recognition using convolutional neural networks: State of the art,” 2016.
- [10] N. Abdulsattar and M. Hussain, “Facial expression recognition using transfer learning and fine-tuning strategies: A comparative study,” in *2022 International Conference on Compu-*

- ter Science and Software Engineering (CSASE)*, (Los Alamitos, CA, USA), pp. 101–106, IEEE Computer Society, mar 2022.
- [11] A. F. Magalhães, *O código de Ekman: O cérebro, a face e a emoção*. FEELab Science Books, 2013.
- [12] S. L. Koole, “The psychology of emotion regulation: An integrative review,” *Cognition and emotion*, vol. 23, no. 1, pp. 4–41, 2009.
- [13] R. W. Picard, *Affective computing*. MIT press, 2000.
- [14] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, jul 1988.
- [15] P. Ekman and W. V. Friesen, “Facial action coding system,” *Environmental Psychology & Nonverbal Behavior*, 1978.
- [16] P. Ekman and C. Szlak, *A Linguagem Das Emoções*. Leya Casa da Palavra, 2011.
- [17] C. Darwin, *The expression of the emotions in man and animals*. John Murray, 1872.
- [18] L. Macedo, A. Cardoso, R. Reizenzein, and E. Lorini, “Artificial surprise,” *Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing and artificial intelligence*, pp. 267–291, 2009.
- [19] A. Portugal, “O que é demência,” *Lisboa: Associação Alzheimer Portugal*, 2019.
- [20] K. A. Johnston, M. Taylor, and EBSCOhost, *The handbook of communication engagement / edited by Kim A. Johnston, Maureen Taylor*. John Wiley Sons, Inc Hoboken, NJ, 2018 2018.
- [21] B. Settles, “Active learning literature survey,” *University of Wisconsin, Madison*, vol. 52, 07 2010.
- [22] R. M. Monarch, *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. Simon and Schuster, 2021.
- [23] S. Agarwal, H. Arora, S. Anand, and C. Arora, “Contextual diversity for active learning,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, pp. 137–153, Springer, 2020.
- [24] P. Bodavarapu and P. Srinivas, “Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques,” *Indian J. Sci. Technol*, vol. 14, pp. 971–983, 2021.
- [25] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual-vgg16 cnn model for big data places image recognition,” in *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*, pp. 169–175, IEEE, 2018.
- [26] S. Saha, “A comprehensive guide to convolutional neural networks — the eli5 way.” Accessed on 29.05.2023.

- [27] K. Leung, *VGG16 architecture*. Towards Data Science, Aug 2021.
- [28] D. A. Da Rocha, F. M. F. Ferreira, and Z. M. A. Peixoto, “Diabetic retinopathy classification using vgg16 neural network,” *Research on Biomedical Engineering*, vol. 38, no. 2, pp. 761–772, 2022.
- [29] G. James, D. Witten, T. Hastie, R. Tibshirani, *et al.*, *An introduction to statistical learning*, vol. 112. Springer, 2013.
- [30] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” in *Proceedings of the 23rd international conference on Machine learning*, pp. 233–240, 2006.
- [31] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, “Learning from class-imbalanced data: Review of methods and applications,” *Expert systems with applications*, vol. 73, pp. 220–239, 2017.
- [32] X. Lai, Y. Lu, L. Zhang, Y. Feng, and G. Zhang, “Imbalanced-type incomplete data fuzzy modeling and missing value imputations,” in *2021 The 5th International Conference on Machine Learning and Soft Computing, ICMLSC’21*, (New York, NY, USA), pp. 33–37, Association for Computing Machinery, 2021.
- [33] H. He and E. Garcia, “Learning from imbalanced data,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, pp. 1263 – 1284, 10 2009.
- [34] M. M. Rahman and D. N. Davis, “Addressing the class imbalance problem in medical datasets,” *International Journal of Machine Learning and Computing*, vol. 3, no. 2, p. 224, 2013.
- [35] J. Brownlee, “Overfitting and underfitting with machine learning algorithms.” Accessed on 22.06.2023.
- [36] J. Ashfaque Minsp Maat and A. Iqbal, “Introduction to support vector machines and kernel methods,” 04 2019.
- [37] T. Fontanari, T. C. Fróes, and M. Recamonde-Mendoza, “Cross-validation strategies for balanced and imbalanced datasets,” in *Brazilian Conference on Intelligent Systems*, pp. 626–640, Springer, 2022.
- [38] S. Yadav and S. Shukla, “Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification,” in *2016 IEEE 6th International conference on advanced computing (IACC)*, pp. 78–83, IEEE, 2016.
- [39] S. Li, “Measuring cognitive engagement: An overview of measurement instruments and techniques,” *International Journal of Psychology and Educational Studies*, vol. 8, pp. 63–76, 07 2021.
- [40] R. Azevedo, “Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues,” *Educational Psychologist*, vol. 50, no. 1, pp. 84–94, 2015.

- [41] S. D’Mello, E. Dieterle, and A. Duckworth, “Advanced, analytic, automated (aaa) measurement of engagement during learning,” *Educational psychologist*, vol. 52, no. 2, pp. 104–123, 2017.
- [42] H. Al Osman and T. Falk, *Multimodal Affect Recognition: Current Approaches and Challenges*. 02 2017.
- [43] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D’Mello, “Automated detection of engagement using video-based estimation of facial expressions and heart rate,” *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 15–28, 2017.
- [44] A. Gupta, R. Jaiswal, S. Adhikari, and V. Balasubramanian, “DAISSEE: dataset for affective states in e-learning environments,” *CoRR*, vol. abs/1609.01885, 2016.
- [45] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The faces of engagement: Automatic recognition of student engagement from facial expressions,” *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.
- [46] J. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. Lester, “Automatically recognizing facial expression: Predicting engagement and frustration,” in *Educational data mining 2013*, 2013.
- [47] A. Saxena, A. Khanna, and D. Gupta, “Emotion recognition and detection methods: A comprehensive survey,” *Journal of Artificial Intelligence and Systems*, vol. 2, pp. 53–79, 01 2020.
- [48] P. Jiang, H. Fu, H. Tao, P. Lei, and L. Zhao, “Parallelized convolutional recurrent neural network with spectral features for speech emotion recognition,” *IEEE Access*, vol. 7, pp. 90368–90377, 2019.
- [49] M. J. Nahian, T. Ghosh, M. Uddin, M. M. Islam, M. Mahmud, and M. S. Kaiser, “Towards artificial intelligence driven emotion aware fall monitoring framework suitable for elderly people with neurological disorder,” pp. 275–286, 09 2020.
- [50] J. Teuwen and N. Moriakov, “Chapter 20 - convolutional neural networks,” in *Handbook of Medical Image Computing and Computer Assisted Intervention* (S. K. Zhou, D. Rueckert, and G. Fichtinger, eds.), The Elsevier and MICCAI Society Book Series, pp. 481–501, Academic Press, 2020.
- [51] Hargurjeet, “7 best techniques to improve the accuracy of cnn w/o overfitting.” Accessed on 19.06.2023.
- [52] B. Han, J. Sim, and H. Adam, “Branchout: Regularization for online ensemble tracking with convolutional neural networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 521–530, 2017.
- [53] N. A. S. Badrulhisham and N. N. A. Mangshor, “Emotion recognition using convolutional neural network (CNN),” *Journal of Physics: Conference Series*, vol. 1962, p. 012040, jul 2021.

- 
- [54] Y. Khairuddin and Z. Chen, “Facial emotion recognition: State of the art performance on fer2013,” 2021.
- [55] T. Gwyn, K. Roy, and M. Atay, “Face recognition using popular deep net architectures: A brief comparative study,” *Future Internet*, vol. 13, p. 164, Jun 2021.
- [56] K. Liu, M. Zhang, and Z. Pan, “Facial expression recognition with cnn ensemble,” pp. 163–166, 09 2016.
- [57] “Kaggle Datasets.” <https://www.kaggle.com/datasets>, 2018. Accessed: 2023.
- [58] CINEICC, “Center for research in neuropsychology and cognitive and behavioral intervention.” Accessed: 2023-04-20.
- [59] P. L. Carrier and A. Courville, “Challenges in representation learning: Facial expression recognition challenge.”
- [60] N. Peinecke, F.-E. Wolter, and M. Reuter, “Laplace spectra as fingerprints for image recognition,” *Computer-Aided Design*, vol. 39, no. 6, pp. 460–476, 2007.
- [61] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern Recognition*, vol. 36, no. 2, p. 451–461, 2003.
- [62] M. A. Syakur, B. K. Khotimah, E. M. Rochman, and B. D. Satoto, “Integration k-means clustering method and elbow method for identification of the best customer profile cluster,” *IOP Conference Series: Materials Science and Engineering*, vol. 336, 2018.
- [63] T. Cao and M. Li, “Facial expression recognition algorithm based on the combination of cnn and k-means,” in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, pp. 400–404, 2019.
- [64] P. Bholowalia and A. Kumar, “Ebk-means: A clustering technique based on elbow method and k-means in wsn,” *International Journal of Computer Applications*, vol. 105, no. 9, 2014.
- [65] F. Siddique, S. Sakib, and M. A. B. Siddique, “Recognition of handwritten digit using convolutional neural network in python with tensorflow and comparison of performance for various hidden layers,” in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, IEEE, sep 2019.
- [66] B. Mohamed, M. Daoud, B. Mohamed, and A. taleb ahmed, “Improvement of emotion recognition from facial images using deep learning and early stopping cross validation,” *Multimedia Tools and Applications*, vol. 81, 09 2022.
- [67] O. Melinte and L. Vladareanu, “Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified adam optimizer,” *Sensors*, vol. 20, 05 2020.
- [68] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.

- [69] U. Ruby and V. Yendapalli, “Binary cross entropy with deep learning technique for image classification,” *Int. J. Adv. Trends Comput. Sci. Eng*, vol. 9, no. 10, 2020.
- [70] Y. Ho and S. Wookey, “The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling,” *IEEE access*, vol. 8, pp. 4806–4813, 2019.





# Apêndices



A

## Descrição do Programa Reminder

## Programa REMINDER – intervenção neuropsicológica de redução do risco de demência

O Programa REMINDER (Silva et al., 2023) consiste numa intervenção não farmacológica baseada nos princípios da reabilitação neuropsicológica que visa reduzir o risco de demência em pessoas com 60 ou mais anos. Este programa é inspirado na iniciativa FINGER () e pretende incluir um conjunto de conteúdos e atividades que visam reduzir o impacto dos vários fatores de risco modificáveis associados à demência.

O programa REMINDER é uma intervenção grupal, para pessoas com 60 ou mais anos, tendo os grupos de ser constituídos entre 5 e 7 pessoas.

O programa inclui um total de 20 sessões, que ocorrem ao longo de 10 semanas, com periodicidade bissemanal, sendo dinamizado por um neuropsicólogo clínico.

As sessões do programa REMINDER incorporam quatro componentes principais: um exercício de relaxamento *mindful*; um momento psicoeducativo; uma atividade experiencial ou de partilha (e.g. tarefa de memória auditiva); uma tarefa para realizar entre sessões (e.g. trabalho para casa). Cada sessão tem a duração aproximada de 60 minutos.

Os objetivos das sessões do programa REMINDER são descritos abaixo, e visam facilitar a aquisição ou otimização de competências cognitivas, emocionais, comportamentais e sociais para prevenir o declínio cognitivo e motivar para a aquisição ou reforço de estilos de vida saudáveis e protetores do cérebro.

### Overview das Sessões do PROGRAMA REMINDER (Manual Técnico, Silva, A.R., 2023)

Sessão	Objetivos
1 Introdução ao programa	Apresentação da estrutura do programa, Estabelecimento das regras do funcionamento das sessões e do grupo Sensibilização para a prática de treino <i>mindful</i>
2 REMIND o cérebro	Sensibilização para o conhecimento de fatores de risco e protetores de declínio cognitivo (DC) Reconhecimento da importância de aproximar o Eu real do Eu ideal
3 REMIND o cérebro 2	Consolidação de conhecimento sobre fatores de risco e fatores protetores do DC Potenciar o conhecimento geral sobre o cérebro como via para a motivação para a saúde cerebral Aprender a valorizar os comportamentos positivos
4 REMIND os objetivos	Reconhecer a importância de ter objetivos pessoais concretos para estimular a mudança Identificar barreiras e facilitadores na concretização de objetivos Aprender a criar objetivos SMART

5 REMIND a atenção 1	Reconhecer a importância do foco atencional para o bem-estar cognitivo e emocional Promover a prática de exercícios de atenção
6 REMIND a atenção 2	Reconhecer a importância de estarmos focados para atingirmos os nossos objetivos O problema do piloto automático Estimular as auto-instruções de atenção STOP-FOCUS
7 REMIND a atenção 3	Valorizar a importância da mente presente Aprender a utilizar todos os sentidos para promover uma atenção plena
8 REMIND a memória 1	Conhecer o funcionamento da memória, os seus limites e potencialidades Identificar dificuldades de memória Identificar hábitos protetores da memória
9 REMIND a memória 2	Conhecer a importância da memória autobiográfica no funcionamento neuropsicológico Aprender sobre o conceito de reminiscência e revisão de vida como prática essencial para uma boa adaptação ao envelhecimento
10 REMIND a memória 3	Conhecer as funções da reminiscência da revisão de vida Identificar acontecimentos chave na linha da vida como estruturadores da identidade Aprender estratégias de praticar reminiscência de forma autónoma
11 REMIND a memória 4	Aprender a utilizar estratégias de memorização no dia a dia Identificar de forma colaborativa, estratégias mnemónicas eficazes
12 REMIND a memória 5	Exercitar estratégias internas de memorização Perceber a aplicabilidade das estratégias internas e externas conforme as necessidades Reconhecer a existência de ajudas externas à memória
13 REMIND a memória 6	Reconhecer o papel das ajudas externas no planeamento e capacidade de recordar eventos futuros Treinar a utilização de ajudas externas à memória
14 REMIND as funções executivas 1	Reconhecer a importância das funções executivas para o dia a dia Perceber a relação entre funções executivas e objetivos

	Praticar a divisão de objetivos em tarefas
15 as funções executivas 2	Identificar o papel das funções executivas na monitorização do comportamento e flexibilidade Reconhecer as aplicações das técnicas de controlo comportamental no dia a dia Treinar a flexibilidade cognitiva na resolução de problemas
16 REMIND a comunicação	Reconhecer a importância da comunicação para manter um cérebro ativo Identificar estratégias facilitadoras da comunicação Exercitar a comunicação bem-sucedida
17 REMIND a socialização	Reconhecer o papel da comunicação na cooperação entre as pessoas Reconhecer a importância da socialização na promoção do bem-estar pessoal e no desenvolvimento pessoal
18 REMIND as emoções e eu	Reconhecer o papel das emoções para a saúde do cérebro Identificar dificuldades de regulação emocional e fatores de risco associados ao envelhecimento Treinar para um autoconceito positivo e realista como forma de melhorar a regulação emocional
19 REMIND as emoções e os outros	Reconhecer a importância da auto-compaixão e do autoconceito em momentos de crise Identificar modelos de resiliência pessoal Reconhecer a importância da rede de suporte que contribui para a autonomia pessoal e qualidade de vida
20 REMIND as aprendizagens	Saber identificar os elementos-chave de competências cognitivas, psicológicas e sociais que devo estimular diariamente para proteger o cérebro. Reconhecer a importância do grupo para realizar mudanças no estilo de vida.

Referências:

- Silva, A.R., Pinho, S., Lima, M.P., Faria, A.L., Afonso, R.M., Baptista, C., Cerejeira, J., Baldeiras, I., Silva-Spínola, A.M., Macedo, L., (2022). Validation trial of a multicomponent neuropsychological interventions for individuals at risk of dementia: the REMINDER program. Poster presented at the MIA Ageing Spring School (4th to 8th April 2022)
- Silva, A.R.; Lima, M.P., Afonso, R.M., & Pinho, M.S.(2023). Programa REMINDER – Manual para Dinamizadores. Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra

- Silva, A.R., Baptista, C., Pinho, S., Lima, M.P., Faria, A.L., & Afonso, R.M., (under review).  
The REMINDER program – protocol for a randomized controlled trial of a neuropsychological group intervention for dementia risk reduction.



B

Póster apresentado no *IV Congresso  
Internacional do Center for  
Research in Neuropsychology and  
Cognitive and Behavioral  
Intervention (CINEICC)*

# Towards an Artificial Intelligence model to automatically infer patient's mental states associated with engagement

Beatriz Carvalho uc2021201040@student.uc.pt

Luís Macedo macedo@dei.uc.pt -orientador

Ana Rita Silva anaritaess@fpce.uc.pt -colaboradora

## Main Goal

This study aims to detect engagement in people with dementia, through the analysis of their facial expressions, during the treatment phase, since there is a great difficulty on the part of health professionals in getting their patients to maintain interest during this stage.

## Introduction

There are many health problems in the world that require long-term treatment and sometimes these are difficult for patients to cope with. Generally, it is recognised that patient engagement plays a critical role in the success of these therapeutic procedures and frequently patients lose motivation and interest during the treatment period. Therefore, it is important to detect fluctuations in patient engagement throughout the duration of treatment and to produce alerts when it deviates from normal levels, so that professionals can act in time to try to correct the situation. Patients usually do not report such situations because they are unable to detect them or simply do not want to give such feedback. The solution may lie in incorporating Artificial Intelligence (AI) techniques so that these states can be detected automatically and quickly [1].

## Methods

In this study a dataset, taken from Kaggle [2], consisting of 35887 facial images of different people, was initially used. To divide the images into "engaged" and "not engaged", it was used a the clustering technique and then the dataset was divided into a training, test and validation data. To build the Machine Learning model, the VGG16 architecture was applied, a type of CNN (Convolutional Neural Network), which consists of an input layer, an output layer and several hidden layers that altogether make VGG16 the algorithm with the best results in this type of classification (image classification). In Fig.1 it is possible to visualise the network structure used, in detail. To bring this study as close as possible to reality, a real dataset will be used, composed of video frames of adults in their 55s, participating in a brain health education programme. This new dataset will be used to test the model already developed and trained.

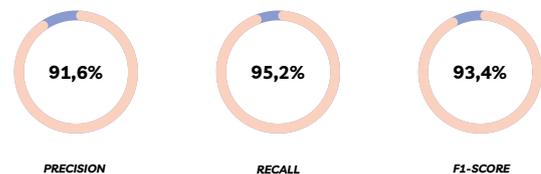
```

Model: "sequential_1"
Layer (type) Output Shape Param #
-----
conv2d_3 (Conv2D) (None, 46, 46, 32) 328
max_pooling2d_2 (MaxPooling (None, 23, 23, 32) 0
2D)
conv2d_4 (Conv2D) (None, 21, 21, 64) 18496
max_pooling2d_3 (MaxPooling (None, 10, 10, 64) 0
2D)
conv2d_5 (Conv2D) (None, 8, 8, 64) 36928
flatten_1 (Flatten) (None, 4096) 0
dense_2 (Dense) (None, 64) 262288
dense_3 (Dense) (None, 2) 138
-----
Total params: 318,082
Trainable params: 318,082
Non-trainable params: 0
    
```

Figure 1: Structure of the VGG16 network used to build the model.

## Results

With the model already trained, some metrics were calculated to prove its performance, such as Precision, Recall and F1-Score. The results are presented below.



Analyzing these results, it is possible to confirm that the model works well during the training and validation phases, and is ready to be tested with the real dataset. After this test, the model is expected to obtain good results, indicating that this algorithm can identify engagement in individuals at risk of dementia at the moment of intervention, detecting early lack of engagement and loss of interest and motivation.

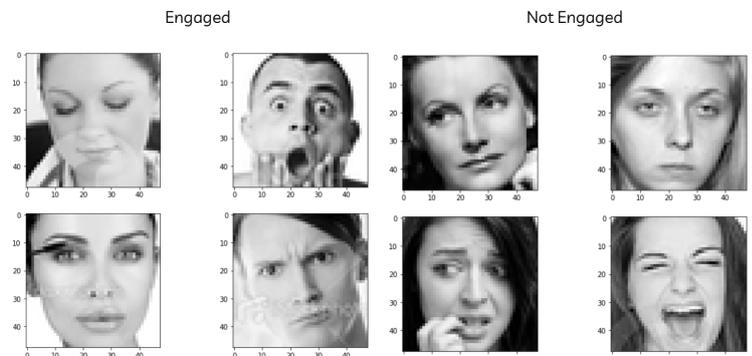


Figure 2: Set of images belonging to the class "Engaged" (left side) and "Not Engaged" (right side) at the time of the test.

## Future work

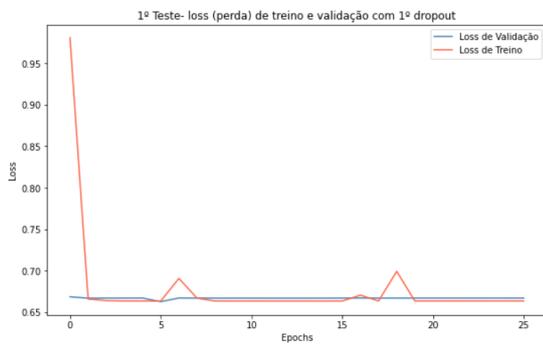
The next phase aims to test this model with the new dataset (the real dataset) that will be provided by CINEICC (Center for Research in Neuropsychology and Cognitive and Behavioral Intervention), based on a brain health education program. These data will initially be in video and then it will have to be converted into static images, with the desired sizes, which indicates that they will have to be previously processed.

## References

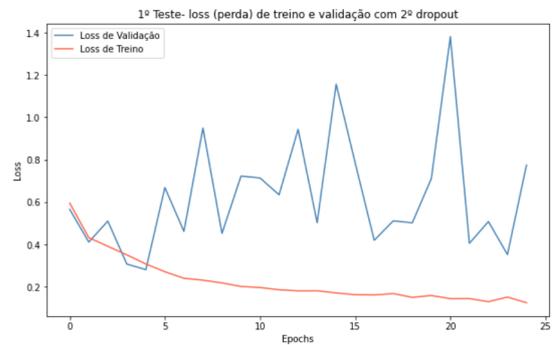
- [1] D'Mello, S.; Dieterle, E. & Duckworth, A. Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning Educational Psychologist, Routledge, 2017, 52, 104-123. doi:10.1080/00461520.2017.1281747
- [2] P.-L. Carrier and A. Courville, "Challenges in representation learning: Facial expression recognition challenge," accessed on 02.02.2023. [Online]. Available: <https://www.kaggle.com/drcapa/facial-expressionedacnn/notebook>

# C

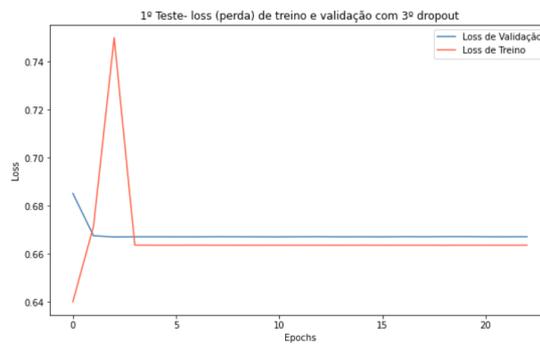
## Dropout para Redução/Eliminação de *Overfitting*



**Figura C.1:** *Loss* de treino e validação ao longo das *epochs*, com *dropout* de 25% e 50%.



**Figura C.2:** *Loss* de treino e validação ao longo das *epochs*, com *dropout* de 15% e 25%.

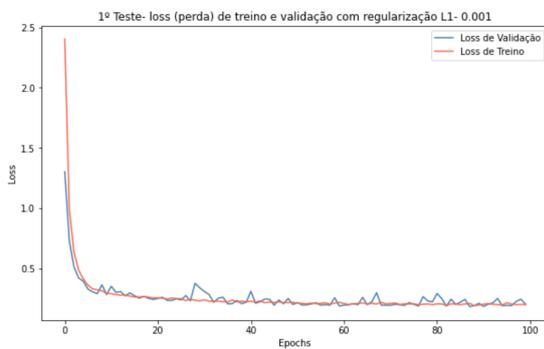


**Figura C.3:** *Loss* de treino e validação ao longo das *epochs*, com *dropout* de 5% e 15%.

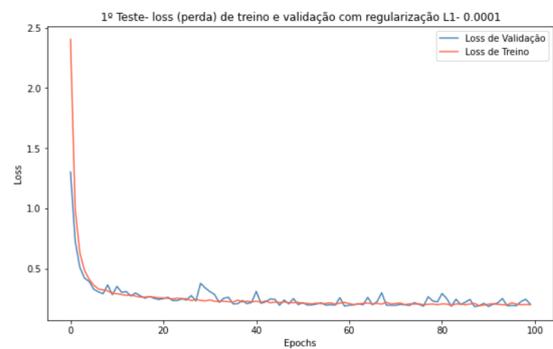


## D

# Regularização de Pesos $L1$ para Redução/Eliminação de *Overfitting*



**Figura D.1:** *Loss* de treino e validação ao longo das *epochs*, com regularização do peso  $L1$ , com 0.001.



**Figura D.2:** *Loss* de treino e validação ao longo das *epochs*, com regularização do peso  $L1$ , com 0.0001.

Modelo	Accuracy	Loss	F1-Score	Precision	Recall
Modelo Reg. L1 (0.001)	94.11%	20.73%	2.72%	2.12%	3.75%
Modelo Reg. L1 (0.0001)	94.69%	17.27%	6.21%	5.07%	8.00%

**Tabela D.1:** Resultado das métricas do primeiro teste experimental realizado, com o modelo após as duas aplicações da regularização do peso  $L1$ .



# E

## Regularização de Pesos $L2$ para Redução/Eliminação de *Overfitting*



**Figura E.1:** *Loss* de treino e validação ao longo das *epochs*, com regularização do peso  $L2$ , com 0.001.



**Figura E.2:** *Loss* de treino e validação ao longo das *epochs*, com regularização do peso  $L2$ , com 0.0001.

Modelo	Accuracy	Loss	F1-Score	Precision	Recall
Modelo Reg. $L2$ (0.001)	94.22%	17.14%	7.85%	9.92%	6.49%
Modelo Reg. $L2$ (0.0001)	92.24%	27.18%	7.72%	6.24%	10.11%

**Tabela E.1:** Resultado das métricas do primeiro teste experimental realizado, com o modelo após as duas aplicações da regularização do peso  $L2$ .