

# Sistema de deteção facial de stress

Ana Xu<sup>1</sup><sup>1</sup> Departamento de Física, Instituto Superior de Engenharia do Porto, Porto, Portugal; 1222111@isep.ipp.pt

**Abstract:** O stress corresponde a uma reação natural do organismo a estímulos externos e internos. A exposição prolongada a situações de stress pode agravar ou provocar graves problemas de saúde, tais como: ansiedade, burnout, depressão, entre outros. O estudo da relação entre as expressões faciais e os sinais de stress tem sido aprofundado, com a evolução da interação homem-máquina, permitiu o desenvolvimento de sistemas computacionais capazes da deteção facial de stress. No presente trabalho, desenvolveu-se uma rede neuronal convolucional (CNN) para a classificação e deteção de emoções faciais para as categorias “Stress” e “No Stress, através do conjunto de imagens obtidas das bases de dados FER2013 e CK+. Recorrendo ao modelo desenvolvido, passou-se à classificação das imagens em tempo real por webcam, avaliando a presença ou ausência de stress. O sistema alcançou uma precisão de 80%, mas apresentou algumas limitações, tais como, a leitura incorreta de algumas emoções devido à não distinção entre elas e deste modo não classificando de forma correta e também quando se encontra em diferentes ângulos faciais. Em conclusão, o código criado pode ser melhorado e adaptado a projetos futuros, com o objetivo de ajudar os trabalhadores com cargas horárias exaustivas, como os profissionais de saúde, desde médicos, assistentes e enfermeiros.

**Keywords:** Stress, rede neuronal convolucional, sistemas computacionais, interação homem-máquina, deteção facial de stress.

## 1. Introdução

O stress é um sentimento de sobrecarga de tensão associado a uma variedade de circunstâncias da vida quotidiana [1]. Atualmente, a sua gestão é fundamental para minimizar os riscos para a saúde, pois é uma das principais causas de doenças crónicas, como as doenças cardiovasculares, diabetes, cancro, entre outras [2,3].

Nos últimos anos, devido aos elevados níveis de stress registados na sociedade, tem havido pesquisas baseadas em várias estratégias como a medição de sinais biológicos. No entanto, estes métodos não eram práticos e necessitavam sempre de equipamento especializado. Por isso, houve a necessidade de desenvolver outros métodos mais robustos e não invasivos para a deteção rápida do stress [4].

Como solução promissora nesse cenário, surgiu a machine learning, um ramo da inteligência artificial, que permite aos sistemas aprenderem e melhorarem com base em dados, sem a necessidade de programação explícita. Neste sentido, existe um crescente interesse na área da interação homem-máquina, com o propósito de aproximar os computadores do comportamento humano, promovendo interações mais naturais e procurando criar interfaces mais simples, eficientes e fáceis de utilizar [5].

### 1.1. As emoções faciais e o stress

A sociedade comunica diariamente, tanto verbalmente como através de gestos e expressões faciais. Muitas vezes, estes gestos e expressões faciais ocorrem espontaneamente e o comunicador nem sequer se apercebe da sua ação, sendo a principal forma de conhecer as emoções do comunicador de um modo não invasivo [6].

Citation: Ana Xu

Academic Editor: Firstname Last-name

Received: 12/01/2025

Revised: date

Accepted: date

Published: date



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Na comunidade científica, e baseado na investigação do Paul Ekman, existe um consenso alargado sobre a universalidade de seis emoções: raiva, nojo, alegria, medo, tristeza e surpresa. Salientando que existem expressões faciais que exprimem universalmente a mesma emoção [7]. No âmbito da realização deste projeto optou-se por utilizar as seis emoções universais e a emoção neutra (não tem nenhuma expressão facial característica). No entanto, o stress não tem uma expressão facial universal de emoção, mas existem estudos dos investigadores Dinges et al. [8] e Lerner et al., à qual referem que as emoções negativas como raiva, nojo e medo relacionam-se inequivocamente com o stress.

## 1.2. Descrição da problemática

Segundo a Agência Europeia para a Segurança e Saúde no Trabalho o stress relacionado com o trabalho (SRT) é o segundo problema de saúde mais frequente associado com o trabalho, tendo uma incidência de 28% entre os trabalhadores da União Europeia [9]. Atualmente, as perturbações de stress são um problema comum entre os trabalhadores de todas as empresas/instituições, mais nomeadamente em circunstâncias como a exposição frequente ao perigo, exigências elevadas, baixo controlo sobre o trabalho e tarefas repetitivas [10]. Neste sentido, o fenómeno do stress laboral deve ser analisado cuidadosamente, em particular na fase final e crónica do stress laboral (burnout) [11]. Embora muitas empresas e entidades proporcionem programas relacionados com a saúde mental e promovam esforços para melhorar o ambiente de trabalho, o problema persiste [9]. Tal acontece devido à dificuldade de os trabalhadores abordarem abertamente sobre os problemas psicossociais e aceitarem que estão perante uma situação vulnerável. Para além, do stress ser difícil de detetar, uma vez que varia de pessoa para pessoa e ainda, podem ser confundidas à primeira vista com outros estados emocionais ou físicos [12].

## 1.3. Objetivos

Os principais objetivos deste trabalho foram o desenvolvimento de um algoritmo em linguagem Python para a deteção facial de stress em imagens e em tempo real (através de webcam) de modo a oferecer maior rapidez nos diagnósticos, minimizando o erro humano e intervindo em tempo útil nos locais de trabalho.

## 2. Estado de arte

A deteção do stress com recurso à machine learning tem sido investigada ao longo do tempo.

Em 2014, um grupo de investigadores, Gao et al., concebeu um sistema de monitorização em tempo real para detetar o stress emocional nos condutores, analisando as expressões faciais através de uma câmara de infravermelhos próximos (NIR) instalada no painel de controlo. O modelo utiliza descritores Scale-Invariant Feature Transform (SIFT) para extrair características locais e normaliza a postura com um modelo cilíndrico 3D. O treino foi efetuado com Support Vector Machines (SVM), utilizando imagens das bases de dados, FACES e Radboud. O sistema foi validado em dois conjuntos de dados: o conjunto 1, filmado num ambiente de escritório com 21 indivíduos simulando emoções básicas; e o conjunto 2, recolhido num automóvel com 12 participantes diferentes em condições reais de aplicação. A deteção foi avaliada utilizando métricas como a accuracy por vídeo e F1-score por fotograma. O melhor desempenho foi obtido através da adaptação do modelo ao ambiente do escritório, juntamente com a normalização da postura. Obtendo uma precisão de 90,5% no conjunto 1 e 85% no conjunto 2, com valores de F1 respetivos de 0,871 e 0,815. Os investigadores concluíram que este método não intrusivo mostrou alta eficiência e potencial para futuras melhorias, como adaptação online e integração de mais sinais [13].

A investigação desenvolvida por Trigeorgia et al., em 2017, propõe um sistema inovador para o reconhecimento de emoções, com recurso a uma abordagem que integra

informação auditiva e visual. O modelo concebido integra várias arquiteturas de redes neurais profundas, incluindo uma Rede Neural Convolutacional para extrair características do sinal de fala, uma ResNet-50 para a modalidade visual e uma Memória de Longo Prazo para modelar a sequência temporal e o contexto emocional. O dataset escolhido foi o REMote COLlaborative and Affective, que foi utilizado no Audiovisual Emotion Challenge de 2016. Esta base de dados contém gravações de interações humanas que expressam emoções de forma natural e espontânea, permitindo uma avaliação robusta do modelo proposto. O modelo alcançou um accuracy notável, apesar do valor exato não ter sido especificado no trabalho. Os resultados obtidos mostraram que o sistema de treino de ponta a ponta teve um desempenho significativamente superior às abordagens tradicionais que utilizam características auditivas e visuais extraídas manualmente [14].

Em Portugal, foi realizado um estudo por um grupo de investigadores para desenvolver o sistema EuStress, concebido para monitorizar e prever os níveis de stress dos estudantes de medicina portugueses em tempo real. O modelo utiliza dispositivos vestíveis, como a Microsoft Smartband 2, para recolher sinais fisiológicos, incluindo a frequência cardíaca (FC) e a variabilidade da frequência cardíaca (VFC), processados por técnicas de aprendizagem automática. Os dados foram obtidos durante períodos de baixa pressão académica e períodos de exames, abrangendo 83 estudantes de medicina da Universidade do Minho. As análises compararam várias técnicas de aprendizagem, incluindo redes neurais, regressão logística, Naive Bayes, SVM, Random Forests e k-Nearest Neighbors. As redes neurais apresentaram os melhores resultados, com uma sensibilidade de 75,2% e uma especificidade de 77,9%, utilizando 100 camadas ocultas e a função de ativação tangente hiperbólica. Foram utilizadas métricas de FC e VFC para treinar e validar os modelos, obtendo-se bons níveis de previsão mesmo com amostras limitadas. Concluíram que o sistema fornece informações relevantes para monitorizar o stress e melhorar o desempenho académico dos alunos, propondo aperfeiçoamentos como intervenções personalizadas com base nos resultados recolhidos e a disponibilização de atividades extracurriculares para diminuir o impacto do stress [15].

Num estudo mais recente, em 2024, Fernandes et al., introduziram uma abordagem inovadora para identificar os níveis de stress entre os profissionais de saúde, utilizando técnicas de aprendizagem automática interpretáveis. Os dados recolhidos da Affective-Road continham 49 características de biomarcadores, como a frequência cardíaca, a atividade electrodérmica e a temperatura da pele. O principal modelo utilizado na investigação foi o Random Forest, que obteve uma precisão de 0,747 e uma área sob a curva de 0,994, demonstrando um desempenho superior ao de outros algoritmos. O estudo evidencia a importância de conhecer as características que condicionam os resultados, no caso específico do modelo Random Forest, as particularidades mais significativas foram os dados relativos à temperatura da pele e à frequência cardíaca, sendo que valores elevados da temperatura da pele e da frequência cardíaca estão associados a níveis baixos ou inexistentes de stress, enquanto a atividade eletrodérmica indica stress com valores elevados [16].

### 3. Metodologia

Este trabalho foi desenvolvido em etapas organizadas para lidar com dados de expressões faciais e construir um modelo de classificação binária, visando distinguir entre "Stress" e "No Stress". De modo, a se perceber os desafios associados à identificação de stress no dataset foram realizadas previsões em diferentes cenários, como variações nas idades, alterações de luminosidade, presença de acessórios (óculos) e diferentes ângulos faciais. De seguida, são apresentadas as etapas do trabalho para o desenvolvimento do sistema de deteção facial do stress.

#### A. Pré-processamento manual do dataset

Para iniciar o projeto, foram escolhidas as bases de dados adequadas, utilizando FER2013 e CK+. Em primeiro lugar, foi efetuado um “pré-processamento” manual para remover contradições, ou seja, eliminar imagens classificadas incorretamente nas pastas correspondentes a cada emoção. Este passo assegurou que as imagens em cada pasta estivessem alinhadas com a emoção pretendida, aumentando a sua precisão no treino.

De seguida, os dois conjuntos de dados foram compilados numa única pasta com subpastas para treino e teste, e em cada uma delas com as pastas correspondentes a cada emoção. Assim, o conjunto de dados de treino ficou com 1568 imagens de raiva (angry), 3302 de medo (fear), 372 de nojo (disgust), 5185 de feliz (happy), 4734 de neutro (neutral), 1709 de surpresa (surprise) e 3863 de triste (sad). Por outro lado, as imagens de teste atribuídas foram 392 para raiva, 347 para medo, 127 para nojo, 914 para feliz, 791 para neutro, 283 para surpresa e 358 para triste. No final, obtiveram-se no total 20733 imagens para treino, sendo 5242 imagens correspondentes ao stress e 15491 imagens ao não stress. Enquanto, para teste foram atribuídas 3212 imagens no total, sendo 866 imagens correspondentes ao stress e 2346 imagens a não stress. Após esta “limpeza” inicial, as emoções foram convertidas numa categoria binária:

- “Stress” - Englobava as emoções raiva, medo e nojo;
- “No Stress” - Incluía as emoções felicidade, neutro, surpresa e tristeza.

#### B. Implementação do código

Inicialmente, as imagens foram reduzidas para 48x48 pixels e convertidas para escala de cinza para reduzir a complexidade computacional, e também normalizadas para valores entre 0 e 1. Em seguida, uma rede neuronal convolucional simples foi utilizada para implementar o modelo, contendo três camadas convolucionais seguidas de operações de pooling para extração de características, terminando em uma camada densa com saída sigmoideal para classificação binária. Para validar o desempenho do modelo, foram analisadas as métricas: matriz de confusão, accuracy, precisão, recall e F1-score, utilizando um limiar de 0,41 para classificar os resultados. Além disso, foi realizada uma avaliação em diferentes cenários, previsões individuais com imagens do diretório para verificar a robustez e a generalização do algoritmo. Finalmente, o código foi adaptado para detetar o stress nos rostos em tempo real recorrendo à webcam.

#### C. Ferramentas utilizadas

Este trabalho foi desenvolvido utilizando o Visual Studio Code, uma IDE muito utilizada para programação devido à sua interface intuitiva e suporte a múltiplas linguagens. A implementação do código foi realizada com Python 3.12.7, utilizando um ambiente virtual configurado com Anaconda Prompt. Este ambiente virtual foi denominado “interacao\_humana”, configurado com o gerenciador de pacotes e ambientes conda, versão 24.11.0. A utilização de um ambiente virtual possibilitou um maior controlo das dependências, garantindo que todas as bibliotecas e versões utilizadas eram compatíveis com o sistema implementado. As bibliotecas python utilizadas no trabalho foram: os, numpy, Tensorflow, Scikit-learn, Matplotlib, Seaborn, OpenCV e Mediapipe.

#### D. Diagrama de Blocos

Previamente à implementação e criação da base de dados utilizada, foi definido um pipeline, no sentido de estipular uma sequência de passos, traduzindo toda a metodologia a desenvolver, que pode ser observada na figura 1.

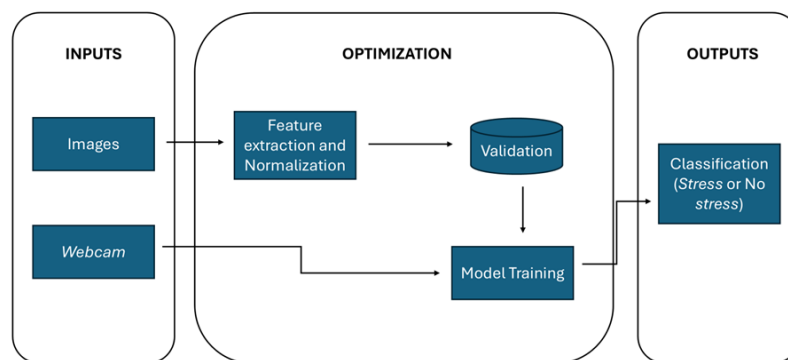


Figura 1. Diagrama de blocos.

#### 4. Resultados e Discussão

##### A. Percepção humana vs Resposta da máquina

Como referido anteriormente realizou-se a previsão de várias imagens do dataset em diferentes cenários para a classificação de “Stress” ou “No Stress”. Os resultados obtidos pela máquina são denominados “Predicted” nas imagens e a resposta “verdadeira” encontra-se colocada na imagem com a designação “Classification”. Na figura 2 estão apresentadas as imagens representativas da classificação “Stress” e “No Stress” para indivíduos de diferentes idades.

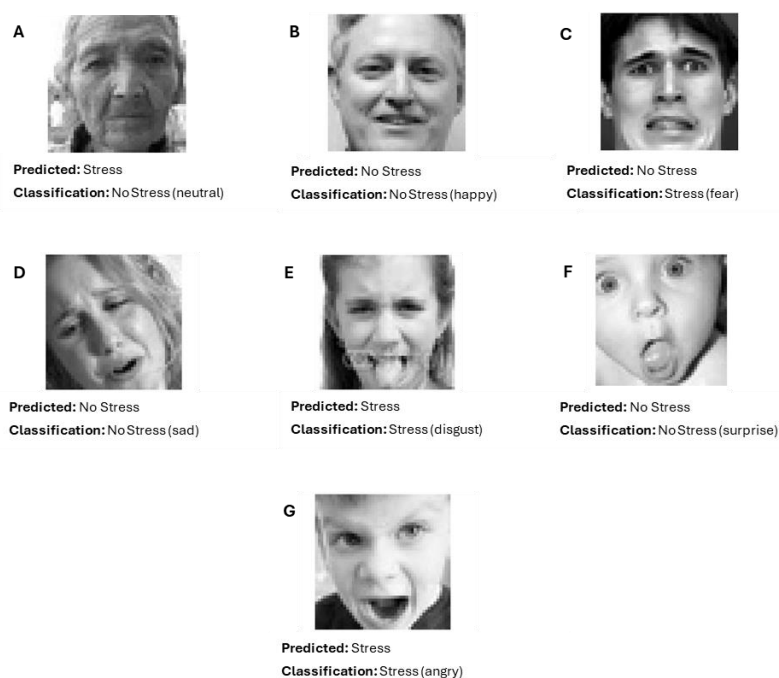
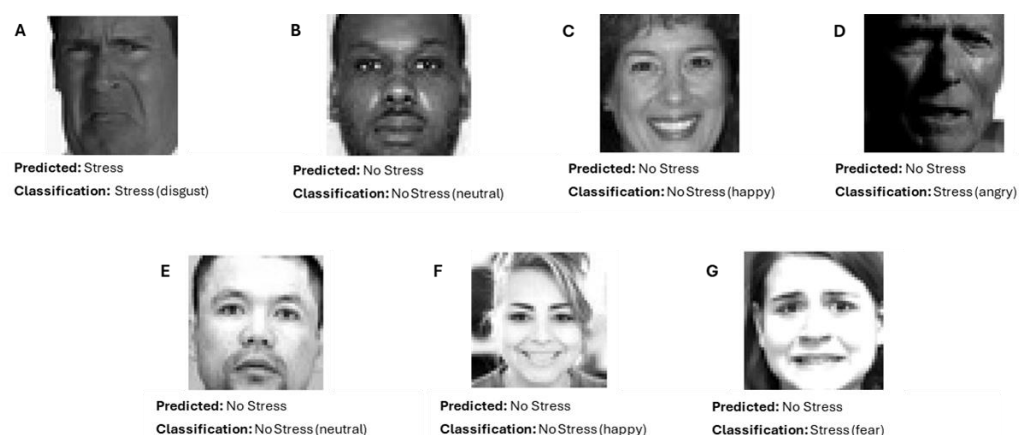


Figura 2. Conjunto de imagens utilizado para comparação de diferentes idades.

Como se pode visualizar na figura 2, a máquina não consegue por vezes distinguir bem as imagens de stress e não stress, tem-se o exemplo da imagem A prevista como “Stress”, mas na realidade tem uma classificação de “No Stress” (correspondente à emoção neutra), e ainda na imagem C que foi dada uma previsão incorreta de “No Stress” e a resposta correta seria “Stress” (correspondente à emoção medo).

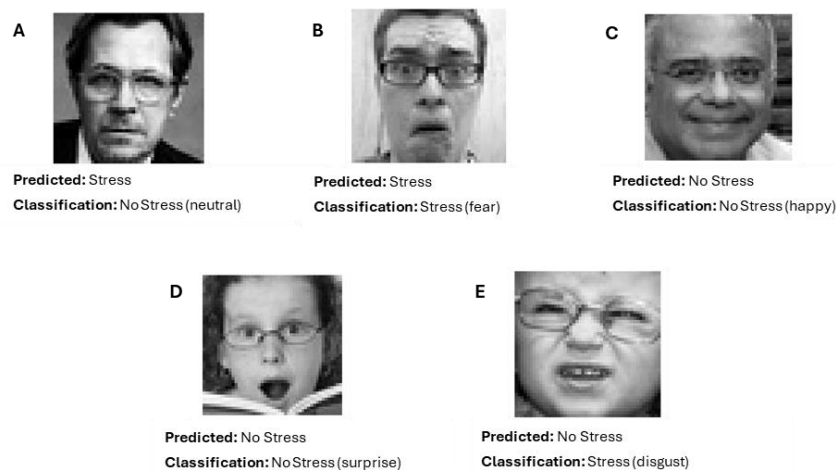
De seguida, prosseguiu-se para a comparação de imagens com muita e pouca luminosidade, como se observa na figura 3.



**Figura 3.** Conjunto de imagens utilizado para comparação de alterações de luminosidade, as imagens A, B, C e D correspondem a baixa luminosidade; as imagens E, F e G correspondem a alta luminosidade.

Na figura 3, verifica-se que em cenários tanto de baixa ou alta luminosidade a máquina não consegue por vezes prever de forma correta, como o caso da imagem D a resposta dada pela máquina foi “No Stress” e a classificação correta seria “Stress” (correspondente à emoção raiva), pode dever-se ao facto da imagem estar com sombra numa parte do rosto, não sendo visível de detetar as características principais da face. E na imagem G na presença de muita luminosidade também não prevê de forma correta a resposta, dando como “No Stress” e a classificação correta era “Stress” (correspondente à emoção medo).

Depois prosseguiu-se para a previsão de imagens pela máquina na presença de um acessório (óculos), como se mostra na figura 4.



**Figura 4.** Conjunto de imagens utilizado para comparação na presença de óculos.

Na figura 4, observa-se que na presença dos óculos no rosto, nem sempre a previsão da máquina é feita de forma correta. As imagens A e E foram dadas respostas incorretas, a classificação certa seria “No Stress” (correspondente à emoção neutra) e “Stress” (correspondente à emoção nojo), respetivamente.

Posteriormente, foi realizada a classificação das imagens no cenário da variação dos ângulos faciais, sendo apresentada na figura 5.

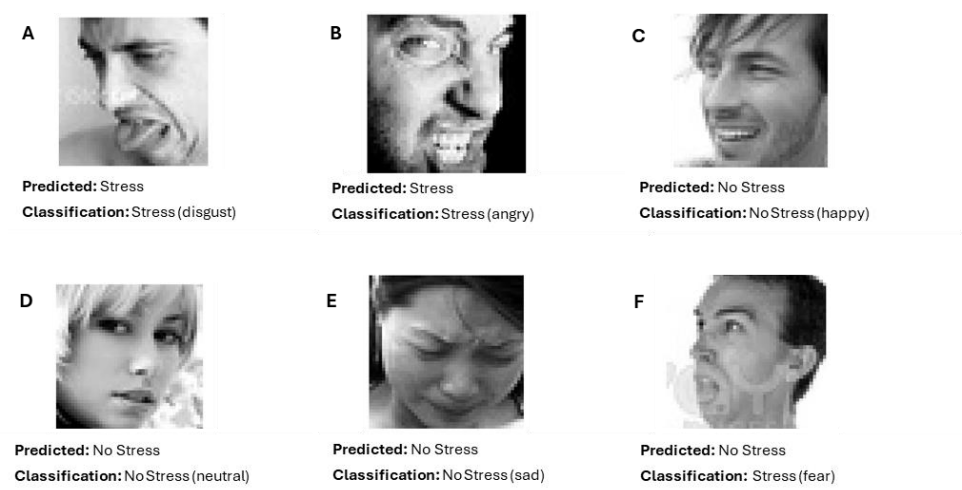


Figura 5. Conjunto de imagens utilizado para comparação em diferentes ângulos faciais.

Pela análise das imagens da figura 5 notou-se que a imagem F não teve uma previsão acertada, o algoritmo assume como “No Stress”, mas devia ser classificada como “Stress” (correspondente à emoção medo).

Para validar o desempenho do modelo, foram analisadas as métricas da matriz de confusão, accuracy, precisão, F1-score e recall (Figura 6 e 7).

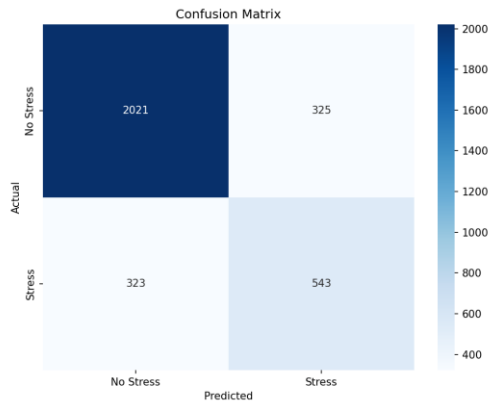


Figura 6. Matriz de confusão.

Classification Report:				
	precision	recall	f1-score	support
No Stress	0.86	0.86	0.86	2346
Stress	0.63	0.63	0.63	866
accuracy			0.80	3212
macro avg	0.74	0.74	0.74	3212
weighted avg	0.80	0.80	0.80	3212

Figura 7. Valores de precision, recall, f1-score e accuracy impressas pela máquina.

Observa-se pela figura 6 que as imagens classificadas como “Stress” e “No Stress” identificadas de forma correta foram superior às incorretas, porém as imagens do “Stress” previstas de modo errado (falsos) ainda foram consideravelmente altos. Para além, dos valores de precisão, recall e f1-score (Figura 7) serem inferiores nas imagens classificadas como “Stress” do que “No Stress”. E apesar do accuracy ser aceitável, de valor 0,80, o

modelo ainda tem algumas limitações, uma vez que o algoritmo não tem uma performance a 100% correta na previsão da classificação das imagens.

#### B. Detecção em tempo real por webcam

Depois de efetuados os testes de detecção de stress na face, a partir das imagens do dataset, procedeu-se à recolha dos resultados da webcam normais e com adição dos óculos, como se observa na figura 8 e 9, respetivamente.



Figura 8. Imagens classificadas em tempo real (webcam).

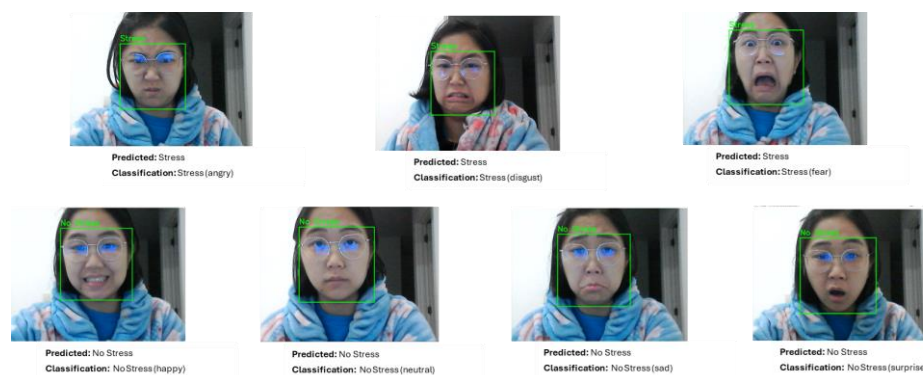


Figura 9. Imagens classificadas em tempo real (webcam) com óculos.

Baseado no teste através de webcam, permitiu constatar que as emoções “medo” e “nojo” foram muitas vezes confundidas com “surpresa”, classificadas como “No Stress” pela máquina. Assim sendo, pode-se afirmar que o algoritmo desenvolvido não é muito exato para a detecção do stress na face.

### 5. Conclusões e Perspetivas futuras

Conclui-se que o sistema desenvolvido é simples, mas apresenta limitações nos diferentes cenários, dado que não deteta as características principais do rosto e ou devido às diferentes emoções faciais terem traços semelhantes nas suas expressões dificultando a detecção precisa e exata pela máquina. O presente trabalho pode ser aplicado em trabalhos futuros, melhorando a otimização do dataset com um número maior de imagens e quantidades mais uniformes entre as classes de emoções para a detecção facial de stress.



## Referências

1. Giannakakis, G.; Pediaditis, M.; Manousos, D.; Kazantzaki, E.; Chiarugi, F.; Simos, P.G.; Marias, K.; Tsiknakis, M. Stress and Anxiety Detection Using Facial Cues from Videos. *Biomed Signal Process Control* **2017**, *31*, 89–101, doi:10.1016/j.bspc.2016.06.020.
2. Li, R.; Liu, Z. Stress Detection Using Deep Neural Networks. *BMC Med Inform Decis Mak* **2020**, *20*, doi:10.1186/s12911-020-01299-4.
3. Zhang, J.; Yin, H.; Zhang, J.; Yang, G.; Qin, J.; He, L. Real-Time Mental Stress Detection Using Multimodality Expressions with a Deep Learning Framework. *Front Neurosci* **2022**, *16*, doi:10.3389/fnins.2022.947168.
4. Naidu, R.P.; Sagar, P.S.; Praveen, K.; Kiran, K.; Khalandar, K. Stress Recognition Using Facial Landmarks and Cnn (Alexnet). In Proceedings of the Journal of Physics: Conference Series; IOP Publishing Ltd, November 30 2021; Vol. 2089.
5. Vural, Ş.F.; Yurdusever, B.; Oktay, A.B.; Uzun, I. Stress Recognition from Facial Images in Children during Physiotherapy with Serious Games[Formula Presented]. *Expert Syst Appl* **2024**, *238*, doi:10.1016/j.eswa.2023.121837.
6. Ke, Q.; Liu, J.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F. Computer Vision for Human–Machine Interaction. In *Computer Vision for Assistive Healthcare*; Elsevier, 2018; pp. 127–145.
7. Ekman, P. What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science* **2016**, *11*, 31–34, doi:10.1177/1745691615596992.
8. Dinges, D.F.; Rider, R.L.; Dorrian, J.; McGlinchey, E.L.; Rogers, N.L.; Cizman, Z.; Goldenstein, S.K.; Vogler, C.; Venkataraman, S.; Metaxas, D.N. Optical Computer Recognition of Facial Expressions Associated with Stress Induced by Performance Demands. *Aviat Space Environ Med* **2005**, *76*, B172–82.
9. EU-OSHA *Work Related\_stress*;
10. Bakker, A.B.; Hakanen, J.J.; Demerouti, E.; Xanthopoulou, D. Job Resources Boost Work Engagement, Particularly When Job Demands Are High. *J Educ Psychol* **2007**, *99*, 274–284, doi:10.1037/0022-0663.99.2.274.
11. Bakker, A.B.; Demerouti, E.; Sanz-Vergel, A.I. Burnout and Work Engagement: The JDR Approach. *Annual Review of Organizational Psychology and Organizational Behavior* **2014**, *1*, 389–411.
12. EU-OSHA *Third European Survey of Enterprises on New and Emerging Risks (ESENER 2019): Overview Report*;
13. Gao, H.; Yuce, A.; Thiran, J.P. Detecting Emotional Stress from Facial Expressions for Driving Safety. In Proceedings of the 2014 IEEE International Conference on Image Processing, ICIP 2014; Institute of Electrical and Electronics Engineers Inc., January 28 2014; pp. 5961–5965.
14. Tzirakis, P.; Trigeorgis, G.; Nicolaou, M.A.; Schuller, B.W.; Zafeiriou, S. End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE Journal on Selected Topics in Signal Processing* **2017**, *11*, 1301–1309, doi:10.1109/JSTSP.2017.2764438.
15. Silva, E.; Aguiar, J.; Reis, L.P.; Sá, J.O. e.; Gonçalves, J.; Carvalho, V. Stress among Portuguese Medical Students: The EuStress Solution. *J Med Syst* **2020**, *44*, doi:10.1007/s10916-019-1520-1.
16. Fernandes, M.S.; Filho, R.R.; Sene-Junior, I.; Sarkadi, S.; Panisson, A.R.; Morales, A.S. An Interpretable Machine Learning Approach for Identifying Occupational Stress in Healthcare Professionals. In Proceedings of the International Conference on Agents and Artificial Intelligence; Science and Technology Publications, Lda, 2024; Vol. 1, pp. 499–506.