
UNIVERSIDADE FEDERAL DE OURO PRETO
DEPARTAMENTO DE CIÊNCIA DE COMPUTAÇÃO
PROCESSAMENTO DIGITAL DE IMAGENS

Projeto de Pesquisa

DETECÇÃO DE EXPRESSÕES EM CÂMERAS DE SEGURANÇA

Alunos:

Carlos Eduardo Gonzaga Romaniello de Souza,

Fabio Henrique Alves Fernandes,

Vinícius Gabriel Angelozzi Verona de Resende,

Resumo

Análise de emoções através da análise facial é um tópico amplamente pesquisado na área da visão computacional, isso devido a sua dificuldade e sua ampla utilização no dia a dia, seja na área de marketing, segurança, ensino, entre outros. Com isto em vista, este trabalho busca estudar abordagens e implementar um programa que utilize uma abordagem para análise de expressões faciais. Para isso, o framework *DeepFace* e a biblioteca *OpenCV* foram utilizados para coleta, treinamento do modelo e predição de expressões. No total, foi utilizado um conjunto de 7 expressões: medo, raiva, surpresa, desgosto, felicidade, neutro e tristeza. Após testes realizados utilizando *webcams* dos integrantes, obtiveram-se resultados satisfatórios quanto a acurácia do modelo e detectores utilizados.

1 Introdução

Reconhecimento de emoções através da análise facial é uma tarefa facilmente executada por seres humanos no dia a dia. No entanto, realizar a mesma atividade computacionalmente com a utilização de algoritmos se mostra uma atividade complexa devido a diversos fatores. Tal problema é de grande importância devido às diversas aplicações da atividade na área da computação, aplicações nas áreas de interação humano-computador, robótica, segurança, marketing entre outros. Tendo isto em vista, esforços são feitos com o intuito de criar novos algoritmos e técnicas que tentam similar a atividade realizada por humanos, técnicas essas que exploram áres de redes neurais, reconhecimento de padrão, processamento de imagem, entre outros. Mesmo após tantos estudos o problema ainda permanece sem solução [Canal et al. \(2022\)](#).

Este trabalho tem como objetivo fazer uma revisão de algumas aplicações descritas na literatura e uma breve análise sobre cada uma. Além disso, iremos estudar e implementar uma aplicação que utilize ou baseia-se nas abordagens que foram analisadas e mostrar os resultados obtidos pelo algoritmo com execução em tempo real.

A organização segue a seguinte ordem, primeiro uma breve análise sobre alguns trabalhos encontrados na literatura a partir de 2016. Em seguida será apresentada a metodologia aplicada. Por fim serão apresentados os resultados e as conclusões tiradas a partir da realização do projeto.

2 Estado da Arte

[Canal et al. \(2022\)](#) apresentam comparações entre diferentes métodos de reconhecimento de expressões faciais classificados em métodos clássicos e baseados em redes neurais artificiais (NNB). Para o apuramento do *survey* realizado, foram estudadas um total de 94 abordagens para o tópico, detalhadas em um total de 51 artigos. Na investigação, observa-se um aumento na presença de Redes Neurais Convolucionais (CNN) e outras abordagens baseadas em redes neurais, como Perceptron Multicamadas (MLP). Além dos métodos mencionados, técnicas clássicas como Rede Bayesiana Dinâmica (DBN) e *Support Vector Machines (SVM)* ainda se mostram muito utilizadas, especialmente pela alta precisão nos ambientes abordados e pela não necessidade de treinamento prévio, como as NNBs. Ambos grupos de técnicas apresentaram prós e contras. Por um lado,

os métodos clássicos são menos custosos que os métodos baseados em redes neurais e podendo atingir cerca de 100% de acurácia durante o reconhecimento. Por outro, os métodos clássicos são limitados ao banco de dados e ambiente utilizado, perdendo significativamente a precisão quando utilizados em situações mais cotidianas. Já as redes neurais apresentaram também casos de 100% de precisão, mas são geralmente extremamente dependentes do banco de dados utilizados para o treinamento das mesmas e nos casos analisados. Apesar de as NNBs apresentarem um comportamento médio menos preciso do que os outros métodos, seu crescimento se da pela possibilidade de utilizar bancos de dados mais realistas, maiores e menos restritos, devido à seu maior nível de abstração quando comparado aos métodos clássicos.

[Nguyen et al. \(2017\)](#) propõem um sistema capaz de analisar faces e identificar 3 tipos de expressões de forma eficiente, estas são: neutras, positivas e negativas. Para o desenvolvimento foram coletadas 1079 imagens e 68 pontos de referência para cada rosto. Os 68 pontos de referência compõem as principais partes do rosto, contornando o olho, nariz, sobrancelhas, pálpebras e mandíbula. Para cada uma das imagens, foram executados três passos: normalização — que consiste em rotacionar a imagem de forma que dos dois pontos inferiores extremos do nariz formem um ângulo reto com o eixo horizontal da imagem; Detecção de característica do rosto para possibilitar a identificação das emoções e por fim a seleção de técnicas e classificadores como *Support Vector Machine* (SVM) para classificação das emoções. Com a abordagem proposta, observou-se que não há necessidade do sistema analisar quadro a quadro do vídeo e sim 1 a cada 3 quadros visto que não há variação significativa entre três quadros consecutivos. Esta abordagem atingiu uma precisão acertiva de 70.65%.

[Kaur et al. \(2022\)](#) propõe uma abordagem para estudo de emoções de alunos em aulas virtuais. A abordagem é dividida em 5 partes principais: pré-processamento das imagens, aumento de dados, extração de características, treinamento e validação. No primeiro, os dados são preparados para serem lidos pela rede implementada, extraindo e removendo dados. Em seguida, na etapa de aumento de dados, uma variedade de dados são gerados rotacionando e preenchendo a imagem a fim de serem utilizados no treinamento do modelo. No passo da extração de características, os dados são filtrados a fim de combinar valores interessantes para o modelo. Na penúltima parte, os dados coletados são processados pela rede durante o treinamento. Por fim, na validação, os resultados adquiridos pelo treinamento do modelo são validados com um conjunto de valida-

ção pré-estabelecido. Após testes realizados durante aulas com 40 alunos, a abordagem proposta apresentou 67% de acurácia com validação de 53%.

Scardua & Marques (2018) buscam reconhecer emoções básicas do ser humano através do método *Face Detection* (FaD). Essas emoções são: raiva, medo, repulsa, surpresa, alegria e tristeza que são consideradas básicas pois os integrantes de qualquer sociedade conseguem representá-las independente do seu contexto social e cultural. Para conseguir tal feito ele começa detectando os elementos importantes do rosto humano através do método *Feature Detection* (FeD). Esse dados coletados são analisados para classificar as emoções captadas pelas câmeras de segurança através do modelo criado pelos psicólogos Paul Ekman e Wallace V., que é conhecido como Modelo psicológico de Classificação Facial (*Facial Action Coding System – FACS*). Esse Sistema de Reconhecimento de Emoções (SER) pode ser utilizado no âmbito da segurança para a prevenção de tumultos ou pode ser associado a um polígrafo, por exemplo. Porém há o empecilho de o SER não conseguir analisar as expressões caso a face do usuário esteja parcialmente coberta ou caso o usuário tenha aplicado *botox*.

Fernández-Caballero et al. (2016) apresentam um ambiente inteligente que realiza o monitoramento de emoções das pessoas através de cameras de vigilância em locais relacionados a saúde como hospitais, clínicas e centros especializados em cuidados. A arquitetura apresentada opera em 3 estágios: “Detecção Humana”, “Regulação da emoção” e “Feedback do controle das emoções”. Na primeira etapa é realizada a coleta dos dados relacionados a emoção do paciente. No segundo, o paciente é submetido a vários estímulos diferentes para a coleta de novos dados. Por fim é realizado um loop do feedback do controle para avaliar o efeito dos estímulos na regulação das emoções através da detecção delas. Nesse modelo são necessárias fontes de comunicação confiáveis para prevenir a perda ou delay na transmissão dos dados. Além disso, o constante monitoramento do status dos dispositivos inteligentes é necessário para lidar com casos de falha para manter a performance e o funcionamento do sistema.

Cruz-Albarran et al. (2017) buscam um método de reconhecimento de emoções através da temperatura do rosto, além da pressão sanguínea que irradia dos vasos sanguíneos durante uma emoção. Para a detecção, é usado um método de termografias em imagens infravermelhas (IRT) para a detecção de 5 emoções básicas: felicidade, desgosto, raiva, tristeza e medo. Através das IRT's geradas, podem ser feitas análises da temperatura do rosto em partes específicas, como por

exemplo, durante um estado de felicidade, quando uma pessoa sorri, acontece uma diminuição na temperatura do nariz e da testa. Este método, por não precisar de uso de sensores diretamente no corpo do usuário, não é invasivo, e a calibração de cada usuário consegue ser feita automaticamente. O sistema proposto foi testado com 44 pessoas, tendo uma eficácia de 89.9% na detecção das emoções. Como não foram encontrados estudos sobre a mudança de temperatura em relação às emoções de desgosto e tristeza, os autores propuseram uma tabela com possíveis alterações nas temperaturas, para dessa forma conseguirem atingir resultados.

Said & Barr (2021) propõe identificar emoções humanas analisando imagens de alta resolução. Ao analisar imagens de câmeras de segurança, a distância entre a câmera e o indivíduo varia a cada frame o que dificulta realizar o reconhecimento de emoções. Nesse trabalho, o autor descreve o método utilizado para solucionar esse problema que consiste em três etapas: identificação de faces humanas na imagem, análise das expressões faciais e reconhecimento de emoções. Durante o desenvolvimento do projeto alguns pontos negativos foram levados em consideração como memória utilizada, precisão e as imagens de alta resolução e para equilibrar esses pontos o autor propõe o uso do Face-Sensitive Convolutional Neural Network (FS-CNN).

Turabzadeh et al. (2018) propõe um sistema de reconhecimento facial no MATLAB, usando o MATLAB Simulink, ferramenta capaz de reconhecer expressões faciais em tempo real com uma velocidade de um frame por segundo. Para que o reconhecimento, foi feito um dataset com vídeos de participantes, onde foram gravados dois vídeos de cinco participantes assistindo a algum outro vídeo. O sistema foi montado em torno de um Field-Programable Gate Array (FPGA), com isso, foi possível manter o reconhecimento de emoções em tempo real com um framerate de 30fps. Ao final, o sistema obteve uma acurácia geral de 47.44%.

3 Metodologia

A fim de melhor entender e estudar o reconhecimento de expressões faciais, montaremos uma aplicação que faz o reconhecimento destas através de imagens coletadas de uma *webcam*. Para tal, utilizaremos um framework desenvolvido em *Python* denominado *DeepFace* (Serengil & Ozpinar, 2020) e a biblioteca python *OpenCV* (Bradski, 2000), responsáveis por capturar as imagens e processá-las.

Em suma, o reconhecimento de expressão consiste em três passos básicos: detecção de face, extração de características e atributos (*landmarks*) e casamento com modelos treinados (reconhecimento). Cada fase será brevemente abordada nos subtópicos 3.1–3.3, explicando algoritmos utilizados na aplicação.

3.1 Detecção de faces

A fase de detecção de face é uma das mais cruciais, visto que a má ou falta da detecção impossibilita o resto do processo. De maneira geral para detectar faces em uma imagem, os algoritmos utilizados procuram os olhos para poderem identificar a posição do rosto e assim encontrar outros pontos de interesse como sobrancelhas, boca e nariz. Neste trabalho utilizamos a técnica *Single Shot MultiBox Detector* (SSD) para identificar as faces que por sua vez utiliza a biblioteca *openCV* para a identificação dos olhos e então rotação da imagem.

Tanto o método SSD quanto a biblioteca *openCV* utilizam modelos pré treinados que passam por redes neurais para ajudar nesse reconhecimento, o que não entra no escopo da disciplina

3.2 Extração de características

Para a detecção de pontos de referência (*landmarks*) no rosto de uma pessoa, foi usada uma biblioteca em Python chamada Dlib, que é capaz de detectar 68 pontos de referência, passando pela mandíbula, queixo, sobrancelhas, olhos, nariz e boca.

Para o Deepface, foram usados somente os limites do rosto (maxilar, queixo e sobrancelhas), o que nos deixa somente 27 dos 68 pontos de referência. Usando esses pontos de referência, podemos retirar somente a face do indivíduo a ser analisado, deixando a imagem sem ruídos que possam interferir na detecção.

Dentro do DeepFace, o algoritmo primeiro faz a detecção do rosto. Logo após, é encontrado cada 27 pontos de referência que compõem os limites do rosto. Com isso, é marcada uma linha em cada ponto, e tudo o que estiver do lado de fora é marcado como 0 na escala RGB, deixando todo o background preto e retirando os ruídos.

3.3 Reconhecimento da expressão

Inicialmente, para que o reconhecimento de cada emoção seja possível, cria-se um modelo de rede neural convolucional (*CNN*) de 3 camadas utilizando a biblioteca Keras (Chollet et al., 2015) e Tensorflow (Abadi et al., 2016). Para treinar o modelo, é necessário utilizar uma base de dados densa, para melhorar a eficácia da predição. Para tal, utiliza-se a base **FER-2013**, fornecida em um desafio lançado pela Kaggle em 2013, que consiste de 28.709 imagens de treinamento e 3.589 imagens de teste.

Com o modelo treinado, reconhecer a expressão se torna um problema de identificar qual das expressões retornadas pelos modelo tem a maior probabilidade.

4 Resultados

Após a execução do algoritmo obtivemos resultados satisfatórios visto que a precisão das análises foram altas para todas as emoções catalogadas. Vale ressaltar que, como o resultado depende de probabilidades, a predição do algoritmo está sujeito a erros, mas apresenta uma boa acurácia pelo identificado nos testes. Para garantir a boa qualidade dos resultados é necessário que o rosto do usuário esteja nítido e que a iluminação esteja adequada para evitar erros na análise dos frames.

Os testes foram realizados com os próprios integrantes do grupo com exceção de 2 emoções onde foi necessário utilizar as imagens da internet para poder reproduzir as emoções já que não conseguimos reproduzí-las de maneira ideal.

A seguir encontra-se alguns exemplos de resultados obtidos pelo algoritmo utilizado nesse trabalho.

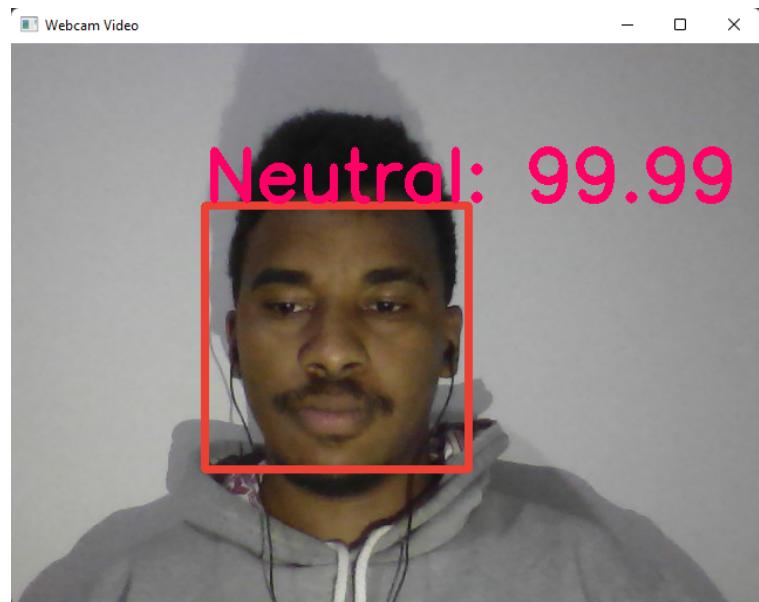


Figura 1: Neutral expression

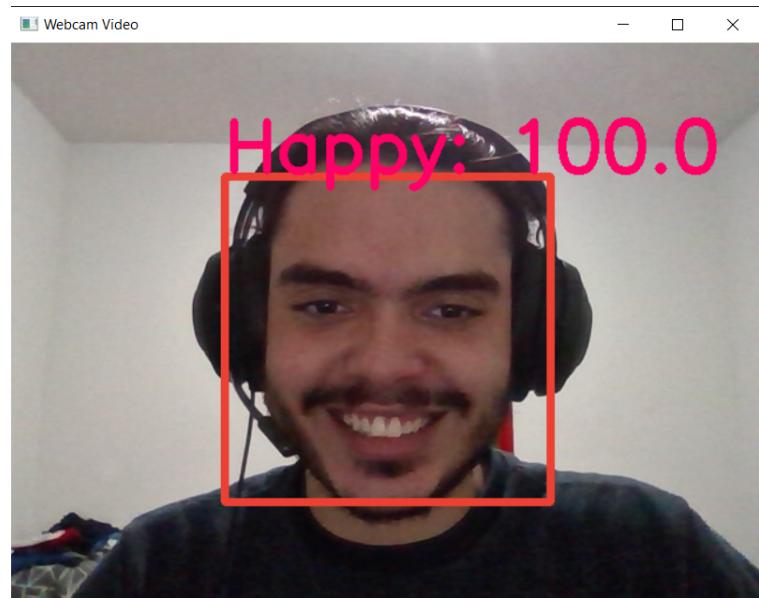


Figura 2: Happy expression

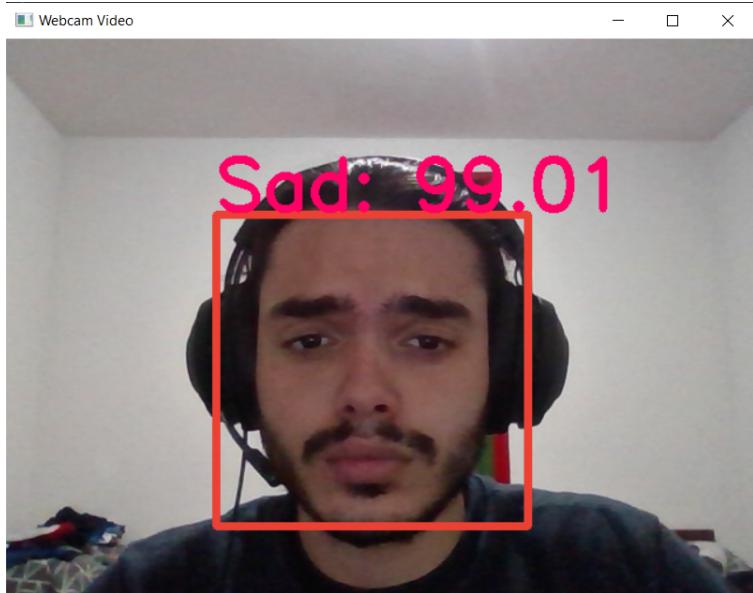


Figura 3: Sad expression

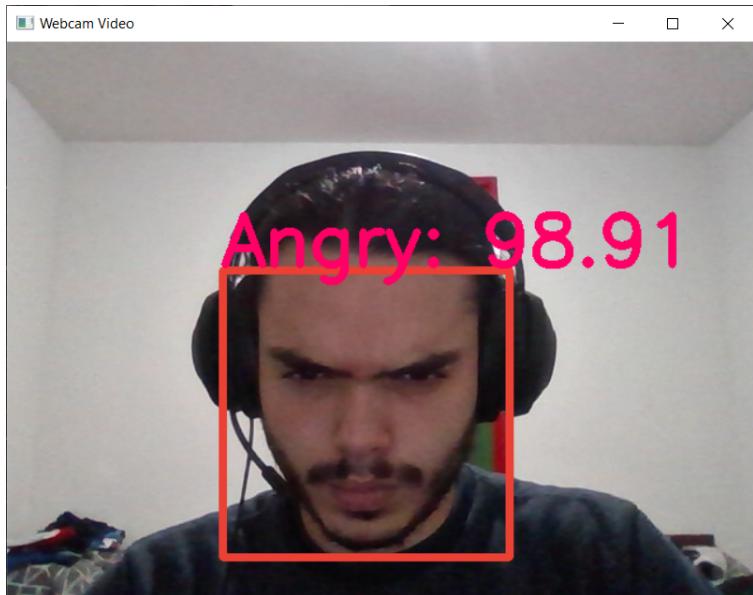


Figura 4: Angry expression



Figura 5: Surprise expression

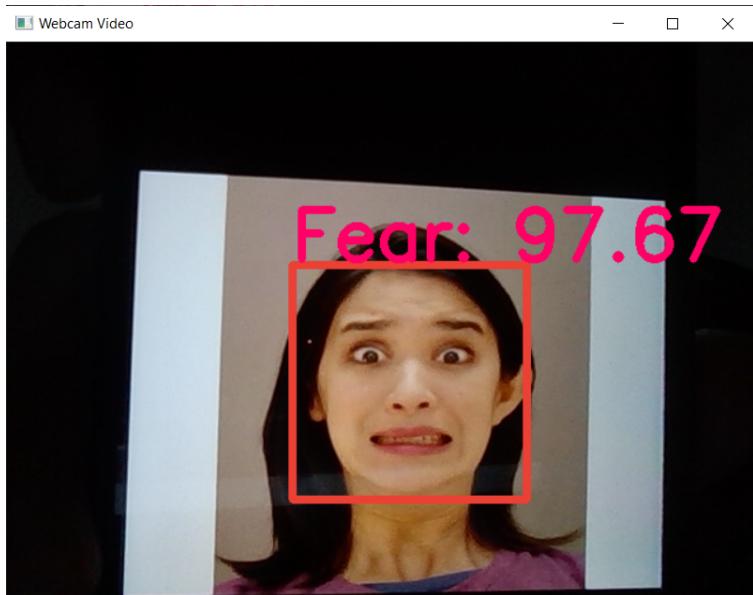


Figura 6: Fear expression

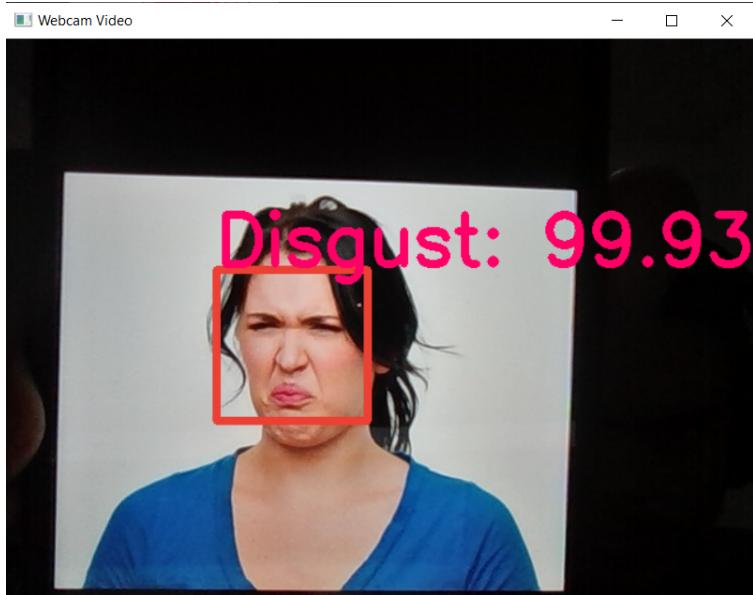


Figura 7: Disgust expression

Como podemos perceber, a precisão do algoritmo é elevada mostrando que ele funciona conforme o esperado identificando as emoções nas imagens. Vale ressaltar que o algoritmo funciona para vídeos e que as imagens presentes neste artigo são apenas frames tirados durante a execução do algoritmo.

Além disso, dado a simplicidade de utilização do modelo e das bibliotecas, para fins educacionais da disciplina, a abordagem tem grandes benefícios. No entanto, para aqueles que buscam maiores estudos, a utilização do *framework* tende a ser mais benéfico para a fins de testes e comparação.

Referências

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265–283).
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., & Sobieranski, A. C. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582, 593-617. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0020025521010136> doi: <https://doi.org/10.1016/j.ins.2021.10.005>
- Chollet, F., et al. (2015). *Keras*. GitHub. Retrieved from <https://github.com/fchollet/keras>
- Cruz-Albaran, I. A., Benitez-Rangel, J. P., Osornio-Rios, R. A., & Morales-Hernandez, L. A. (2017). Human emotions detection based on a smart-thermal system of thermographic images. *Infrared Physics Technology*, 81, 250-261. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1350449516304182> doi: <https://doi.org/10.1016/j.infrared.2017.01.002>
- Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, M. T., ... Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics*, 64, 55-73. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1532046416301289> doi: <https://doi.org/10.1016/j.jbi.2016.09.015>
- Kaur, P., Kapoor, A., Solanki, Y., Singh, P., & Sehgal, D. (2022, 05). Deep learning based emotion detection in an online class.. doi: 10.1109/DELCON54057.2022.9752940
- Nguyen, B. T., Trinh, M. H., Phan, T. V., & Nguyen, H. D. (2017). An efficient real-time emotion detection using camera and facial landmarks. In *2017 seventh international conference on information science and technology (icist)* (p. 251-255). doi: 10.1109/ICIST.2017.7926765

- Said, Y., & Barr, M. (2021). Human emotion recognition based on facial expressions via deep learning on high-resolution images. *Multimedia Tools and Applications*, 80(16), 25241–25253.
- Scardua, D. A., & Marques, K. (2018). Estudo da identificação de emoções através da inteligência artificial. Retrieved from <https://multivix.edu.br/wp-content/uploads/2018/08/estudo-da-identificacao-de-emocoes-atraves-da-inteligencia-artificial.pdf>
- Serengil, S. I., & Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. In *2020 innovations in intelligent systems and applications conference (asyu)* (pp. 1–5).
- Turabzadeh, S., Meng, H., Swash, R. M., Pleva, M., & Juhar, J. (2018). Facial expression emotion detection for real-time embedded systems. *Technologies*, 6(1). Retrieved from <https://www.mdpi.com/2227-7080/6/1/17> doi: 10.3390/technologies6010017