Proceeding Paper

Benchmarking Computer-Vision-Based Facial Emotion Classification Algorithms While Wearing Surgical Masks †

Luis Coelho [1,2,*] **, Sara Reis** [1,3]**, Cristina Moreira** [1]**, Helena Cardoso** [1]**, Miguela Sequeira** [1] **and Raquel Coelho** [1]

[1] Instituto Superior de Engenharia Porto, Instituto Politécnico do Porto; 4249-015 Porto, Portugal; ssr@isep.ipp.pt (S.R.); 1210159@isep.ipp.pt (C.M.); 1220163@isep.ipp.pt (H.C.); 1221871@isep.ipp.pt (M.S.); 1180805@isep.ipp.pt (R.C.)

[2] Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC); 1000-029 Lisbon, Portugal

[3] Centro de Inovação em Engenharia e Tecnologia Industrial (CIETI); 4249-015 Porto, Portugal

* Correspondence: lfc@isep.ipp.pt

† Presented at the Advances in Biomedical Sciences, Engineering and Technology (ABSET) Conference, Athens, Greece, 10–11 June 2023.

Abstract: Effective human communication relies heavily on emotions, making them a crucial aspect of interaction. As technology progresses, the desire for machines to exhibit more human-like characteristics, including emotion recognition, grows. DeepFace has emerged as a widely adopted library for facial emotion recognition. However, the widespread use of surgical masks after the COVID-19 pandemic presents a considerable obstacle to its performance. To assess this issue, we conducted a benchmark using the FER2013 dataset. The results revealed a substantial performance decline when individuals wore surgical masks. "Disgust" suffers a 22.6% F1-score reduction, while "Surprise" is least affected with a 48.7% reduction. Addressing these issues improves human–machine interfaces and paves the way for more natural machine communication.

Keywords: emotion perception; facial emotion; emotion classification; surgical mask

1. Introduction

The perception of emotions is a crucial aspect in social interaction, as they play a fundamental role in how individuals communicate, express themselves, and connect with one another. Human beings are social creatures and rely on effective communication to establish and maintain relationships, build trust, and negotiate complex social interactions. Emotions consist of paralinguistic information that is part of the communication process, providing important cues about the mental and physical state of individuals, their intentions, and the context of the interaction. Tone of voice, body language, and facial expressions are all key indicators of an individual's emotional state and can help others understand their feelings and respond appropriately. In fact, individuals who are better at recognizing and interpreting emotions are more successful in their social interactions and have stronger relationships with others [1].

In the rapidly evolving landscape of human–machine interactions, the recognition and understanding of emotions have emerged as pivotal factors in enhancing the quality and effectiveness of these interactions. As machines and artificial intelligence (AI) systems become increasingly integrated into various aspects of daily life, the ability to accurately perceive and respond to human emotions becomes imperative for creating seamless and meaningful interactions in both social and technological contexts [2,3].

One of the main benefits of emotion perception in human–machine interaction is that it can help machines to personalize their responses based on the user's emotional state. Emotional mimicry, the imitation of the emotional context conveyed by the informant, is an unconscious behavior that regulates or mediates social interactions [4]. Machines designed

to interact with humans should be aware of this effect. For example, a virtual assistant that recognizes when a user is frustrated or angry can adjust its responses and the tone to be more soothing or helpful, which can improve the user's experience and satisfaction with the technology and help the user achieve the desired interaction goals [5]. In education, emotion-aware applications can monitor students' state, attention and engagement levels during online or offline learning sessions. If a student seems confused or disinterested, the system can adapt its content delivery in real-time, offering additional explanations, interactive exercises, or even taking short breaks to maintain engagement and optimize the learning experience [6,7]. In autonomous vehicles, facial emotion recognition can assist in ensuring the safety of passengers. If the system detects signs of drowsiness, distraction, or discomfort in the driver's facial expressions, it can trigger alerts or interventions to prevent accidents. Additionally, it could adjust the vehicle's settings to create a more calming environment if stress is detected [8]. In addition, emotion perception can also be useful in several healthcare areas, such as chronic diseases or mental health [2,4]. For example, in the field of mental health, emotion perception can be used to detect changes in a patient's mood or behavior, which can help healthcare providers to identify potential relapses or episodes of anxiety or depression. By detecting these emotions early, healthcare providers can work with patients to develop coping strategies or provide additional support to help manage their condition and prevent further complications. Precise emotion perception is becoming the essential objective in human–machine interaction as it can greatly enhance the effectiveness, efficiency, and personalization of these interactions, integrating a more "human" characteristic into machines.

Nevertheless, the use of a certain configuration of facial movements may not be sufficient to fully characterize an individual's true emotional state. The production of facial expressions has similarities across humans (or even some animals) because there is a common underlying muscular system. However, these muscle activation can be faked [9]. Moreover, studies have reported that facial patterns associated with emotions can vary significantly depending on cultural factors, the context of interaction, and even among individuals in similar contexts [10]. For these reasons, an accurate estimation of emotions should not only rely on facial expressions but be multimodal, covering speech, kinesics, or even chemical sources of information.

In this paper, our focus is on decoding (superficial) emotions by examining the correspondence between facial expressions and emotions based on Ekman's definition [1,11] and within the broader context of Western culture. The reported results for facial emotion recognition using computer models show a good maturity of technology [12]. However, after the COVID-19 pandemic, the use of surgical masks in social environments has become a regular behavior. In this context, the use of facial masks significantly reduces the visibility of important facial features, making it more difficult to accurately perceive emotions. In Figure 1a, the facial muscles that are used by humans to manipulate facial expression are depicted. Following Ekman's taxonomy and Facial Action Coding System (FACS), muscle engagement obeys specific Action Unit (AU) codes, as shown in Table 1. Each code represents the transition from a neutral state to the expression of the related emotion. In Figure 1b, we can observe that several muscles are hidden by the use of a mask, but there are many others, related with AUs from the top area of the face, as in Table 1, that remain visible and allow the decoding of emotional expressions.

Table 1. Numerical codes for each facial emotion expression according to Ekman's Facial Action Coding System [11], grouped by facial location/area. Each code represents the activation of a muscle or a group of muscles.

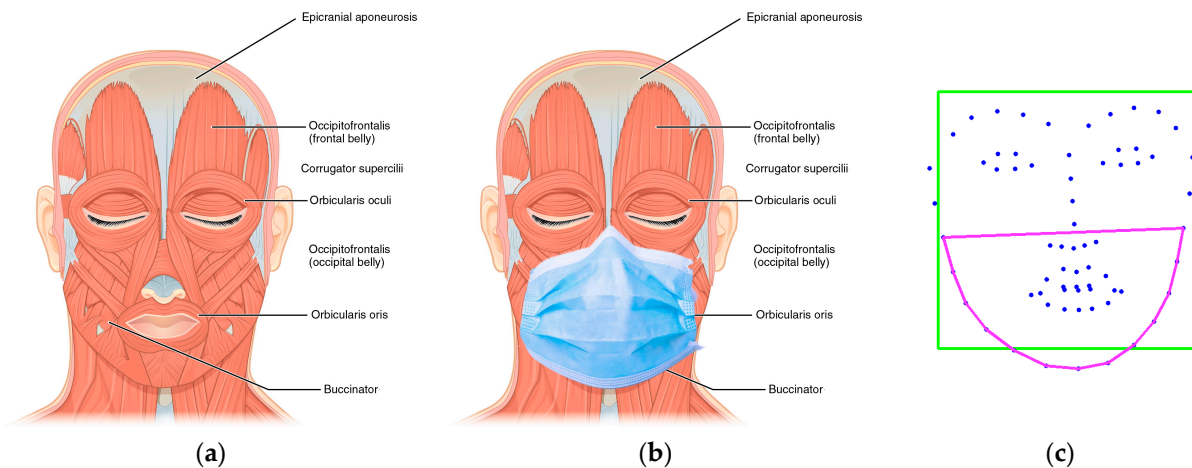| Location	Happiness	Sadness	Surprise	Fear	Anger	Disgust
Top		1 + 4	1 + 2 + 5B	1 + 2 + 4 + 5 + 7	4 + 5 + 7	
Middle	6					9
Low	12	15	26	20 + 26	23	15 + 17

Figure 1. (**a**) Muscles that are involved in human facial expression (with limited labels, for clearer representation), (**b**) superimposed mask and (**c**) example of bounding box after face detection (in green), facial landmark detection (blue dots) and artificially generated superimposed mask (purple line over blue dots) for facial emotion classification task (targeted photographic image was not represented for a clearer visualization) (Facial muscles representations were adapted from an original created by author CNX Anatomy 2013 and distributed under a Creative Commons Attribution 4.0 International License [13]).

Additionally, masks may also create wrinkles or lines on the face, which could be interpreted as conveying a different emotion than the one intended. This poses a problem for humans but also for machines. In this paper, we proposed to benchmark a popular computer vision framework, using a dataset of human faces with superimposed artificially generated facial masks, in a facial emotion classification task.

2. Computer-Vision Based Facial Emotion Recognition

There are many distinct emotions that human beings can interpret; however, besides "neutral", there are six that are considered primary emotions: sadness; fear; anger; disgust; surprise; and happiness. These emotions are more easily interpreted by humans, since they have very different facial movements and expressions between them [6].

Computational classification of facial emotions is an interesting and challenging process since no two facial expressions are the same. The facial expression is a "revelation" made by a human being in a natural way, where the individual expresses what they are feeling at a given moment. However, despite not considering that there are equal facial expressions from people to people, it is possible to find computational mechanisms that recognize certain characteristics/points in each of our facial expressions and assign them a certain value.

In the last few years, Convolutional Neural Networks (CNNs) have been very successful in facial emotion classification because they are specifically designed to recognize patterns and features in images. CNNs can automatically learn and extract high-level features from images that are relevant for classification tasks such as facial emotion recognition. In facial emotion recognition, the convolutional layers are used to detect facial features that are relevant for different emotions, such as the curvature of the mouth for smiling or the wrinkles around the eyes for expressing sadness. Additionally, CNNs that were developed for this purpose, have been trained on large datasets of facial images with known emotions to learn the patterns and features that are most relevant for facial emotion classification. The learned features are then used to classify new images of faces into different emotions. In fact, this is one of the main advantages of CNNs, their ability to handle variations in facial expressions, such as changes in lighting, pose, and occlusion. This makes them robust and effective for real-world applications [14].

DeepFace [15] is a deep learning facial recognition system developed by Facebook's Artificial Intelligence Research team in 2014. It uses a deep neural network with more than 120 million parameters to recognize faces in images and videos with high accuracy. The system works by first detecting and aligning faces in an image, and then using a deep convolutional neural network to extract features from the aligned faces. These features are then used to compare faces and determine if they match. In this process, a set of distinct stages and tools are used.

The use of such systems for detecting emotions in individuals wearing surgical masks as well as their performance in such conditions is not scientifically reported to the authors' knowledge.

3. Development

3.1. Materials

There is a wide availability of photographic image datasets that allow facial emotion classification. The most popular are presented in Table 2, as well as the related decision model and highest benchmarks. For our purposes, we have used the FER-2013 dataset [16] due to its high number of images for each class and because the transmitted emotions are in a natural context (not performed by actors with an expression purpose). It consists of 35,887 grayscale images of size 48 × 48 pixels, each labeled with one of seven facial expressions: angry, disgust, fear, happy, sad, surprise, and neutral. The dataset was introduced in 2013 as part of a facial expression recognition challenge hosted on the Kaggle platform. The dataset's images were obtained from the Google Images search engine and were manually labeled by human annotators. Since its introduction, it has become a widely used benchmark for facial expression recognition and to compare the performance of different models or factors on facial expression recognition, such as image quality, pose, and lighting conditions. For our purposes, the images were upsampled to 144 × 144, suiting the input requirements of the classification system. Human-performed emotion classification accuracy on this dataset is estimated to be 65.5% [17].

Table 2. Popular datasets for facial emotion recognition.

Ref., Year.	Model	Precision	Description
[18], 2015	k-NN, Gaussian SVM, ELM-RBF	99.75%	CK (500 image sequences of 100 people)
[14], 2017	CNN	99.68%	CK+ (147,000 samples)
[15], 2018	VGG, ResNet, GoogleNet, AlexNet,	60.98% 49.16% 63.91% 64.24%	FER2013 (35,887 different facial images divided into 7 categories.)
[17], 2019	VGG-16+, LEMHI	78.40%	MMI (20 people of different ethnicities and ages. Each person recorded 79 series of facial expressions.)
[19], 2005	SVM Multiclass	99.70%	CK (500 image sequences of 100 people.)
[20], 2006	SVM, LP	80.52% 81.82%	CMU-Pittsburg and database without expressions (Only a part of the dataset was used. 253 examples of expressions.)
[21], 2009	DeepFace	97.70% 91.40%	LFW (13,323 internet photos of 5749 celebrities) YTF (3425 YouTube videos from 1595 people)

3.2. Methods

For our purposes, we followed the pipeline represented in Figure 2. Using the Deep-Face tools, for each image of the dataset, the system first performs face detection. If this is successful, facial landmarks are then created. This is a crucial operation since these points are used for face alignment, a geometric transformation that is applied to the image to

align it according to a predefined template, which usually involves rotating and scaling the image to match the desired orientation and size. This alignment helps to minimize the effects of pose and illumination variations and enables to improve the subsequent classification module. Additionally, these points are also used, for our purposes, to superimpose a simulated surgical mask on the picture. An example of the results obtained after this process is shown in Figure 3, where a set of photos is represented in its original representation, when face detection was not possible, and with the superimposed mask, when face and facial landmarks detection were successful. For the "Emotion Classification with Mask" stage, only the pictures with masks were used.
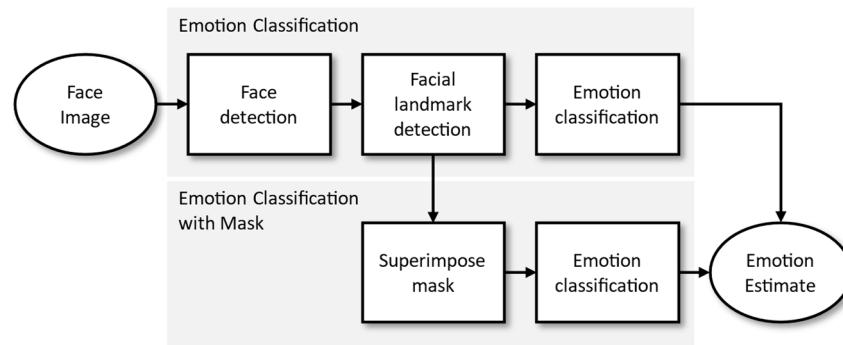

Figure 2. Pipeline for the proposed benchmarking system.

Figure 3. Subset of 100 pictures after using DeepFace for face detection and, when successful, for detection of facial landmarks. Mask representations are based on the resulting face contour points (key-points 2 to 15).

4. Results

For evaluating the face identification algorithm, we have used accuracy, defined as

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Instances} \quad (1)$$

For the evaluation of the classifiers, we have used widely established metrics. Precision and recall are often used together to balance the trade-off between identifying true positive instances and minimizing false positive instances. The F-score is a metric that combines both precision and recall into a single value, providing a more comprehensive evaluation of the model's performance, especially on imbalanced datasets. For our purposes, precision was calculated as

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{2}$$

Recall was calculated as follows:

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \tag{3}$$

Considering FN as the number of false negatives, the F-score was calculated as

$$F_{score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

When considering global metrics, we have calculated a weighted metric to address class imbalance. Using F-score as an example:

$$F_{weighted\_score} = \frac{1}{N}\sum_c N_c F_{c,\ score} \tag{5}$$

where N is the total number of cases and N_c is the number of cases for class c.

4.1. Face Detection

The initial step of facial emotion recognition involves detecting the presence of a face within the given image and establishing a bounding box around it, as the green rectangle depicted in Figure 1c. In the rightmost column of Table 3 (gray background), we can observe the (O)riginal number of images that were evaluated (column Support-O) and the number of images where a face was successfully detected (column Support-FD).

Table 3. Metrics for the results of a facial emotion classification task. The provided values are for the (O)riginal dataset and for a dataset with artificially generated superimposed (M)asks, calculated for each emotion/class. The ratio between these indicators is also provided. The highest value for each column is represented in bold face.

Emotion	Precision			Recall			F1-Score			Face Detection		
Class	O	M	M/O	O	M	M/O	O	M	M/O	O	FD	FD/O
Angry	0.72	0.26	36.1%	0.65	0.25	38.5%	0.69	0.26	37.7%	3994	3002	75.2%
Disgust	0.58	0.10	17.2%	0.66	0.27	40.9%	0.62	0.14	22.6%	436	357	81.9%
Fear	0.54	0.19	35.2%	0.78	0.28	35.9%	0.64	0.22	34.4%	4097	2772	67.7%
Happy	**0.84**	**0.44**	**52.4%**	**0.89**	0.32	36.0%	**0.86**	0.37	43.0%	7215	**5974**	**82.8%**
Neutral	0.83	0.37	44.6%	0.59	0.30	50.8%	0.69	0.33	47.8%	**9930**	5334	53.7%
Sad	0.66	0.27	40.9%	0.76	0.27	35.5%	0.71	0.27	38.0%	4830	2982	61.7%
Surprise	0.77	0.34	44.2%	0.79	**0.43**	**54.4%**	0.78	**0.38**	**48.7%**	3171	2484	78.3%
Accuracy	-	-	-	-	-	-	0.73	0.31	-	33673	22905	-
Macro Avg.	0.70	0.28	-	0.73	0.30	-	0.71	0.28	-	33673	22905	-
Weigth. Avg.	0.75	0.33	-	0.73	0.31	-	0.73	0.31	-	33673	22905	-

Faces from the "Happy" and "Disgust" classes were the most easily detected, with accuracy values of 82.8% and 81.9%, respectively. On the other hand, the classes "Neutral" and "Sad" exhibited identification rates of only 53.7 and 61.7%. The values obtained for the "neutral" class are surprising, since this is the most prevalent class/emotion and it

represents the reference expression to estimate the others, according with FACS system. For this reason, the correct classification of this expression should not be poorly performed, nor should it constitute a challenging problem for a computer-vision system. This may be the result of factors such as pose, lighting conditions or partial representation, conditions that should be correctly handled by the system but, since there is no reference meta-information, this is an unexplored dimension of the challenge.

4.2. Facial Emotion Classification

After face detection, the next stage consists of classifying the respective emotion. To perform the facial emotion classification benchmark, we have calculated the confusion matrices using the pictures without masks and with mask, based on the pictures where a face was detected (numbers in Table 3, column FD). The obtained results are represented in Figure 4, left and right, respectively. For the original dataset, happy faces were the most successfully classified, reaching an 88.8% success rate. As for face detection, neutral was shown to be the most difficult facial emotion to classify, reaching only 59.3%, and with high confusion rates with "fear" (11.0%), "sad" (10.0%) and "happy" (8.3%). This is a deceptive result since, not only because, in most cases, an individual is not exhibiting a marked expression, but also due to the fact that neutral can be used as the reference state from which other emotions can be estimated by observing facial dynamics. Facially expressed "fear", the emotion that involves the highest number of muscles activated, is ranked third in terms of success. For the dataset with superimposed masks (Figure 4, right), we can observe a high performance decline for all emotions. The classifier performed best for "surprise" (42.8%) and "happy" (32.4%) and worst for "angry" (25.2%) and "disgust" (26.6). "Neutral" was the least affected in absolute terms, with a 29.2% decline, and "surprise" was the least affected in relative terms. This was an expected result, since the masks hide important expression landmarks; however, eyes and eye brows, which are important cues, remain visible. "Angry" and "surprise", facial expressions that have limited landmarks on the areas covered by the mask have been highly impacted. This seems to point out that the algorithm's classification strategy heavily relies on this information, or that the higher displacement of the lips and chin are more informative. By considering both confusion matrices, we can observe that "surprise" and "happy" are the facial expressions that are better classified overall.

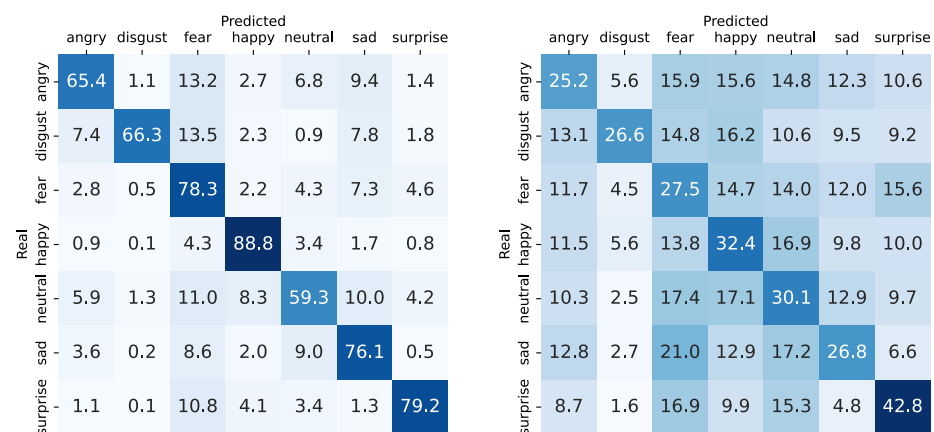|	Predicted								Predicted						
	angry	disgust	fear	happy	neutral	sad	surprise		angry	disgust	fear	happy	neutral	sad	surprise
angry	65.4	1.1	13.2	2.7	6.8	9.4	1.4	**angry**	25.2	5.6	15.9	15.6	14.8	12.3	10.6
disgust	7.4	66.3	13.5	2.3	0.9	7.8	1.8	**disgust**	13.1	26.6	14.8	16.2	10.6	9.5	9.2
fear	2.8	0.5	78.3	2.2	4.3	7.3	4.6	**fear**	11.7	4.5	27.5	14.7	14.0	12.0	15.6
happy	0.9	0.1	4.3	88.8	3.4	1.7	0.8	**happy**	11.5	5.6	13.8	32.4	16.9	9.8	10.0
neutral	5.9	1.3	11.0	8.3	59.3	10.0	4.2	**neutral**	10.3	2.5	17.4	17.1	30.1	12.9	9.7
sad	3.6	0.2	8.6	2.0	9.0	76.1	0.5	**sad**	12.8	2.7	21.0	12.9	17.2	26.8	6.6
surprise	1.1	0.1	10.8	4.1	3.4	1.3	79.2	**surprise**	8.7	1.6	16.9	9.9	15.3	4.8	42.8

Figure 4. Confusion matrices after a facial emotion classification task. On the left, using photographic images of natural faces and, on the right, using photographic images of faces wearing a surgical mask (artificially generated and superimposed). Color range, from white to dark blue, correspond to values from 0 to 100.

Detailed information, comprising precision, recall and F1-metrics, are shown in absolute values in Table 3. The precision and F1-score, were most affected in the "disgust" class, with 17.2% and 22.6% performance after the addition of the facial mask. On the other

hand, "surprise" was the easiest emotion to classify in the simulated mask conditions, with a 47.8% reduction in F1-score, when compared with an uncovered face.

5. Conclusions

Emotions are crucial for a natural and successful interaction between humans and between humans and machines. They convey important paralinguistic and pragmatic information that allows interlocutors to adapt their dialog and communication intentions. In recent years, the use of facial masks in social contexts has hindered the ability to observe important visual clues that allow us to identify facial emotions, both for humans and machines. In this paper, we have presented a benchmark for DeepFace, a freely available CNN-based facial emotion classification system. The results show that both stages, face detection and emotion classification, can still be highly improved (in the context of the tested dataset). Only half or less of the expressed emotions were correctly identified when a mask was used, exposing the limitation of such systems. The importance of facial landmarks used in the representation of emotions has explainable models [22] and the obtained results can help to understand the relative contribution of each aspect to achieve an effective classification of facial emotions.

It is also important to mention that the use of automated facial emotion recognition technology raises several ethical concerns across various dimensions. These include privacy considerations, as emotions are monitored in public spaces without explicit consent, potential inaccuracies and biases that might disproportionately affect certain groups, the risk of promoting discriminatory practices based on perceived emotional states, and the creation of a surveillance culture that infringes upon personal freedom. Furthermore, issues surrounding data security, consent for vulnerable populations, emotional manipulation, and unintended consequences necessitate thorough assessment. Striking a balance between technological assistance and genuine human interaction remains a challenge. Addressing these ethical questions requires collaboration among technologists, ethicists, policymakers, and society, ensuring that facial emotion recognition adheres to ethical standards, respects individual rights, and avoids harmful societal impacts.

References

1. Zadra, J.R.; Clore, G.L. Emotion and Perception: The Role of Affective Information. *Wiley Interdiscip. Rev. Cogn. Sci.* **2011**, *2*, 676–685. [CrossRef] [PubMed]
2. Mandal, M.K.; Habel, U.; Gur, R.C. Facial Expression-Based Indicators of Schizophrenia: Evidence from Recent Research. *Schizophr. Res.* **2023**, *252*, 335–344. [CrossRef] [PubMed]
3. Djordjevic, M.; Glumbić, N.; Brojčin, B.; Banković, S.; Žunić Pavlović, V. Differences in Pragmatic Communication Skills of Adults with Intellectual Disabilities and Dual Diagnoses. *Front. Psychiatry* **2023**, *14*, 1072736. [CrossRef] [PubMed]
4. Gerłowska, J.; Dmitruk, K.; Rejdak, K.; Gerłowska, J.; Dmitruk, K.; Rejdak, K. Facial Emotion Mimicry in Older Adults with and without Cognitive Impairments Due to Alzheimer's Disease. *AIMS Neurosci.* **2021**, *8*, 226–238. [CrossRef] [PubMed]
5. Coelho, L.; Braga, D.; Dias, M.; Garcia-Mateo, C. An Automatic Voice Pleasantness Classification System Based on Prosodic and Acoustic Patterns of Voice Preference. In Proceedings of the Interspeech 2011-12th Annual Conference of the International Speech Communication Association, Florence, Italy, 27–31 August 2011.

6. Zhang, S.; Meng, Z.; Chen, B.; Yang, X.; Zhao, X. Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects. *Front. Psychol.* **2021**, *12*, 728495. [CrossRef] [PubMed]
7. Coelho, L.; Reis, S. Enhancing Learning Experiences Through Artificial Intelligence: Classroom 5.0. In *Fostering Pedagogy Through Micro and Adaptive Learning in Higher Education: Trends, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2023; ISBN 978-1-66848-656-6.
8. Xiao, H.; Li, W.; Zeng, G.; Wu, Y.; Xue, J.; Zhang, J.; Li, C.; Guo, G. On-Road Driver Emotion Recognition Using Facial Expression. *Appl. Sci.* **2022**, *12*, 807. [CrossRef]
9. Jack, R.E.; Garrod, O.G.B.; Yu, H.; Caldara, R.; Schyns, P.G. Facial Expressions of Emotion Are Not Culturally Universal. *Proc. Natl. Acad. Sci.* **2012**, *109*, 7241–7244. [CrossRef] [PubMed]
10. Barrett, L.F.; Adolphs, R.; Marsella, S.; Martinez, A.M.; Pollak, S.D. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychol. Sci. Public. Interest.* **2019**, *20*, 1–68. [CrossRef] [PubMed]
11. Ekman, P.; Friesen, W.V.; Hager, J.C. *Facial Action Coding System. Manual and Investigator's Guide*; Research Nexus: Salt Lake City, UT, USA, 2002.
12. Canedo, D.; Neves, A.J.R. Facial Expression Recognition Using Computer Vision: A Systematic Review. *Appl. Sci.* **2019**, *9*, 4678. [CrossRef]
13. Creative Commons—Attribution 4.0 International—CC BY 4.0. Available online: https://creativecommons.org/licenses/by/4.0/deed.en (accessed on 10 August 2023).
14. Lopes, A.T.; de Aguiar, E.; De Souza, A.F.; Oliveira-Santos, T. Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order. *Pattern Recognit.* **2017**, *61*, 610–628. [CrossRef]
15. Gan, Y. Facial Expression Recognition Using Convolutional Neural Network. In Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, Las Vegas, NV, USA, 27–29 August 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–5.
16. FER-2013. Available online: https://www.kaggle.com/datasets/msambare/fer2013 (accessed on 11 August 2023).
17. Hu, M.; Wang, H.; Wang, X.; Yang, J.; Wang, R. Video Facial Emotion Recognition Based on Local Enhanced Motion History Image and CNN-CTSLSTM Networks. *J. Vis. Commun. Image Represent.* **2019**, *59*, 176–185. [CrossRef]
18. Ali, H.; Hariharan, M.; Yaacob, S.; Adom, A.H. Facial Emotion Recognition Using Empirical Mode Decomposition. *Expert Syst. Appl.* **2015**, *42*, 1261–1277. [CrossRef]
19. Kotsia, I.; Pitas, I. Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. *IEEE Trans. Image Process.* **2007**, *16*, 172–187. [CrossRef] [PubMed]
20. Anderson, K.; McOwan, P.W. A Real-Time Automated System for the Recognition of Human Facial Expressions. *IEEE Trans. Syst. Man Cybern. Part B* **2006**, *36*, 96–105. [CrossRef] [PubMed]
21. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
22. Hussain, N.; Ujir, H.; Hipiny, I.; Minoi, J.-L. 3D Facial Action Units Recognition for Emotional Expression 2017. *arXiv* **2017**, arXiv:1712.00195.