EMOTION DETECTION USING MACHINE LEARNING: ASSESSMENT OF EFFECTIVENESS AND APPLICABILITY IN DIFFERENT CONTEXTS

DETECÇÃO DE EMOÇÕES USANDO APRENDIZAGEM DE MÁQUINA: AVALIAÇÃO DE EFICÁCIA E APLICABILIDADE EM DIFERENTES CONTEXTOS

Luciano Ferreira Rodrigues-Filho 1
Alan Silva Martins 2

Abstract*: This research conducts a comprehensive analysis of emotion detection using facial expressions, employing various experiments and methodologies. The study initiates with a robust dataset comprising 35,685 grayscale facial images categorized by distinct emotions. Preliminary tests reveal a significant correlation between language learning settings and data patterns, demon-strating the model's adaptability to diverse contexts. Nevertheless, the re-search transcends mere emotion detection, striving to fathom the intricacies of emotions across diverse scenarios. This investigation entails complex con-siderations, such as the role of neutral expressions as a positive emotion in-dicator and the substitution of emotion capture time with emotion propor-tion in long-term analyses. Various experiments are conducted, altering participant numbers and environmental conditions. The outcomes indicate that, within controlled environments, the system yields favorable results, but real-world simulations, including obstacles and inadequate lighting, present challenges. Depth analysis underscores complexities in representing three-dimensional objects in a two-dimensional space. This study underscores the promise of emotion detection via facial expressions while emphasizing the complexities and challenges across different contexts. The findings offer valuable insights for enhancing model accuracy and applicability in real-world scenarios, pinpointing avenues for future enhancements and invest-ments.*

Keywords: *Emotion Detection. Facial Analysis. Machine Learning. Human-Computer Interaction.*

Resumo: *Nesta pesquisa foi realizado uma análise abrangente da detecção de emoções por meio de expressões faciais, empregando diversos experimentos e metodologias. O estudo começa com um conjunto de dados robusto composto por 35.685 imagens faciais em tons de cinza categorizadas por emoções distintas. Os testes preliminares revelam uma correlação significativa entre as configurações de aprendizagem de línguas e os padrões de dados, demonstrando a adaptabilidade do modelo a diversos contextos. No entanto, a pesquisa transcende a mera detecção de emoções, esforçando-se para compreender os meandros das emoções em diversos cenários. Esta investigação envolve considerações complexas, como o papel das expressões neutras como indicador de emoção positiva e a substituição do tempo de captura da emoção pela proporção da emoção em análises de longo prazo. Vários experimentos foram conduzidos, alterando o número de participantes e as condições ambientais. Os resultados indicam que, em ambientes controlados, o sistema produz resultados favoráveis, mas simulações do mundo real, incluindo obstáculos e iluminação inadequada, apresentam desafios. A análise de profundidade ressalta as complexidades na representação de objetos tridimensionais em um espaço bidimensional. Este estudo ressalta a promessa da detecção de emoções por meio de expressões faciais, ao mesmo tempo que enfatiza as complexidades e desafios em diferentes contextos. As descobertas oferecem informações valiosas para melhorar a precisão e a aplicabilidade do modelo em cenários do mundo real, identificando caminhos para melhorias e investimentos futuros.*

Palavras-chave: *Detecção de Emoções. Análise Facial. Aprendizado de Máquina. Interação Humano-Computador.*

[1] Graduação em Psicologia pelo Centro Universitário de Ourinhos/UNIFIO, graduação em Pedagogia pela Faculdade União Cultural do Estado de São Paulo, mestrado em Psicologia Social pela Pontifícia Universidade Católica de São Paulo - PUC/SP e Doutorando em Pesquisa e Desenvolvimento em Biotecnologia Médica pela Faculdade de Medicina de Botucatu FMB/UNESP. Atualmente é docente no Centro Universitário de Ourinhos/UNIFIO, pesquisador da Universidade Estadual do Norte do Paraná/UENP, pesquisador da Pontifícia Universidade Católica de São Paulo/ NUTAS/PUC/SP. Trabalhou no Hospital del Trabajador de Santiago, Chile, na Clinique La Borde - Cour-Cheverny - França, no Hospital de Saúde Mental de Ourinhos, no Instituto Psicopedagógico Ciudad Joven San Juan de Dios, Sucre, Bolívia. Com estudos e pesquisas na New School for Social Research, New York/EUA, no East Side Institute, NewYork/USA, na Universidad de la Republica, Montevidéu, Uruguai. Na Universidade Eduardo Mondlane, Maputo, Moçambique.. Lattes: http://lattes.cnpq.br/3264658506558579. ORCID: https://orcid.org/0000-0003-1547-9301. E-mail: lu_fr@yahoo.com.br

[2] Graduando em Ciência da Computação pelo Centro Universitário de Ourinhos/UNIFIO. Lattes: http://lattes.cnpq.br/5985730774277021. ORCID: https://orcid.org/0009-0008-8626-4838. E-mail: alanosms711@gmail.com

Introduction

The ability to comprehend and interpret human emotions is paramount for social interaction. Howev-er, this task is not always straightforward, particularly when verbal communication is limited or absent. In such situations, facial analysis has proven to be an effective tool for capturing facial expressions and deducing underlying sentiments.

Artificial intelligence (AI) has brought about revolutionary changes across various domains, with faci-al analysis using machine learning (ML) emerging as a promising technique for emotion detection. This approach facilitates the automated processing of facial expressions, which can yield valuable insights across fields such as psychology, marketing, medicine, education, and security.

In psychology, this technology aids in evaluating patients' emotional states and monitoring psycholog-ical disorders. In marketing, it can gauge consumer receptivity to advertising campaigns. In education, it helps identify student engagement in the classroom, while in security, it is employed to detect suspicious emotions at airports, train stations, or other high-traffic locations.

A sound understanding of facial expressions can enhance non-verbal communication in diverse con-texts, whether they involve interpersonal interactions or situations entailing human-machine interaction. By comprehending emotions conveyed through facial analysis, one can respond in a more suitable and empathetic manner, thereby improving interaction quality.

Moreover, ML-based facial analysis can offer more precise and consistent assessments of human emotions than conventional methods. Through machine learning algorithms, systems can be trained to recognize subtle patterns in facial expressions, enhancing sensitivity and accuracy in detecting diverse emotions. This enables continuous, real-time monitoring of facial expressions across different contexts.

Given these justifications, the relevance and necessity of research in implementing AI-driven facial analysis systems for emotion detection are evident. Understanding human emotions through facial ex-pressions can contribute to the development of more efficient solutions across various domains, leading to improvements in communication, well-being, and people's overall quality of life.

However, it is crucial to acknowledge the contradictions and challenges intrinsic to this approach. First and foremost, image quality plays a critical role in the accuracy of facial analysis. Unfavorable lighting conditions or low-resolution cameras in environments can compromise the AI's ability to capture and in-terpret facial expressions. Additionally, variations in facial appearances across individuals from different ethnic groups, age brackets, and genders can impact the generalization of AI models, resulting in varying performance across diverse demographics.

Another fundamental contradiction lies in privacy and ethics. Collecting facial expression data with-out adequate consent or the misuse of such data raises significant ethical and legal concerns. Further-more, facial analysis often deals with 2D images, neglecting the depth dimension of facial expressions. This limitation can lead to inaccurate interpretations since depth is pivotal in non-verbal communication. The ambiguity of facial expressions and the potential for multiple interpretations also pose a challenge, as AI may not always accurately discern emotions.

Other considerations involve the intricate interplay of non-verbal cues, including body language and tone of voice, which are not fully accounted for in facial analysis alone. Additionally, the presence of bi-ases and prejudices in AI models, stemming from training data, can result in biased emotion interpreta-tions, potentially leading to discrimination in practical scenarios.

Therefore, while AI-based facial analysis holds substantial promise in detecting human emotions, it is imperative to recognize that its implementation grapples with a range of complex contradictions and challenges that necessitate comprehensive and ethical addressal. This research seeks to investigate the implementation of advanced facial analysis systems using machine learning techniques for emotion de-tection, with the aim of comprehending their effectiveness, accuracy, and applicability across different contexts. The study endeavors to contribute to scientific and technological progress in this domain, offer-ing pertinent insights for the practical deployment of ML-based facial analysis and its influence across various fields.

Material and Methods

The proposed methodology consists of validating the machine learning (ML) algorithm for faci-al emotion detection, using the categories of disgust, sadness, fear, anger, surprise, happiness and neutral.

The adopted paradigm is firmly anchored in the supervised learning methodology. This ap-proach involves dedicating efforts to exhaustive training of the model through exposure to an extensive and comprehensive data set, which brings together a wide diversity of images representing human faces in different emotional states. At the same time, each image was annotated in order to record the emotion concretely expressed by the face depicted in it.

The genesis of the model presupposes a meticulous and comprehensive analysis of the entire corpus of data, a process that unfolds with detail and extensive dedication. Through this detailed screen-ing, we seek to identify and extract distinctive patterns that unequivocally correlate with each emotional category. As the model is exposed to an increasing amount of data and the training process continues, its intrinsic ability to identify and discriminate emotions based on facial features progressively improves.

This continuous progress is a direct manifestation of the principle underlying machine learning, in which the model iteratively optimizes its capabilities as more data is injected into the process. This phenomenon of iterative improvement is what gives the model the ability to generalize, and is the basis of its power to "learn", thus giving rise to the term machine learning. As a result, the model emerges as a highly competent and versatile tool for accurately identifying emotions based on facial manifestations.

In this way, the facial expressions of a diverse sample of participants were captured, varying in group size to assess the consistency of results in different social contexts. The groups were composed of 1 individual, 3 people, 5 people, 26 people and 280 people.

Sample by Group Size

- Individual (N = 1): This group included a single participant, aiming to understand how the in-structions and template affected facial expressions in an individual context.

- Group of 3 People (N = 3): This group represented a more intimate environment, where partici-pants had the opportunity to interact and observe each other's expressions during the experiment.

- Group of 5 People (N = 5): This intermediate group configuration was selected to evaluate how the dynamics of a slightly larger group could influence facial expressions.

- Group of 26 People (N = 26): In this case, a larger group size was chosen to investigate whether social dynamics and group pressure could have an impact on participants' facial expressions.

- Group of 280 People (N = 280): This large-scale group was used to gain an overview of how the instructions and knowledge of the answer key affected facial expressions in a larger and more diverse environment.

These groups participated in the research in four distinct stages, each with its own duration and set of instructions:

Stage 1 - Spontaneous Expressions (21 seconds): In this stage, participants were instructed to express randomly dictated emotions without receiving any specific guidance on how they should perform these facial expressions. Each emotion lasted 3 seconds.

Step 2 - Short Spontaneous Expressions (7 seconds): Similar to the previous step, participants were asked to express randomly dictated emotions, but this time each emotion was expressed for just 1 sec-ond. Again, no detailed instructions were provided.

Step 3 - Presentation of the Template (Instructions): After the first two steps, participants were ex-posed to a template that displayed visual examples of what facial expressions should be like for each specific emotion. This step aimed to contaminate the participants' spontaneous process with information from the answer sheet.

Step 4 - Replication with Instructions (21 seconds and 7 seconds): In this step, participants were asked to replicate the same emotions, but this time they received detailed instructions based on the template on how to perform the correct facial expressions. The duration of the expressions was again 3 seconds for each emotion in the first part and 1 second for each emotion in the second part.

Two main analyses were carried out

1. Comparison between Spontaneous Expressions and Replies with Instructions: It was inves-tigated whether exposure to the answer sheet significantly influenced the participants' facial expressions. The facial expressions from the first two stages (without instructions) were compared with the expressions from the last two stages (with instructions), in order to assess whether the instructions and the answer key had an impact on the facial expressions of emotions.
2. Processing Speed: In addition, the research also addressed processing speed, both in the con-text of the computer vision system and the processing speed of the human mind. The analysis compared the results obtained in the two different durations (3 seconds and 1 second) to understand how the time variation affected the quality and accuracy of facial expressions.

The data obtained and statistical analyzes were performed to determine whether there were sig-nificant differences between the different steps and durations, considering both the accuracy of facial expressions and the processing speed. The results provided insights into the influence of instructions and prior knowledge of a template on facial expression of emotions, as well as implications for processing speed in computer vision systems and human mental processes.

In order to obtain a more precise analysis, we constructed an index using a weighted formula. The formula considers the weights assigned to each emotion as follows: disgust (D1), sadness (S1), fear (F1), anger (A1), surprise (S2), happiness (H2), and neutral (N2). The weights vary between 0 and 100 and are associated with different emotions: 0 for negative emotions, 50 for neutral emotions, and 100 for posi-tive emotions.

Participants will indicate on a scale from 0 to 100 how much they feel dissatisfied (for negative emo-tions) or satisfied (for positive emotions), with a weight of 0 assigned to negative emotions, 50 to neutral emotions, and 100 to positive emotions. The index will be calculated as follows:

$$(100*(H2 + S2) + 50*N2 + 0*(D1 + S1 + F1 + A1))/100$$

The results obtained by the facial detection algorithm will be subjected to a comparative analysis be-tween the groups and their stages, aiming to investigate the agreement between the objective analysis provided by the algorithm and the machine/human processing speed. This comparison process allows for a systematic assessment of the validity and reliability of the algorithm in relation to the detection and classification of facial expressions related to the investigated emotions.

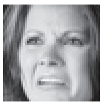
To carry out this comparative analysis, appropriate statistical measures were used, such as correlation coefficients or agreement indices, which made it possible to quantify the degree of relationship between the results obtained by the algorithm and the participants' responses in the questionnaire. Furthermore, in-terjudge agreement analysis techniques were applied to evaluate the consistency between different hu-man raters and the AI algorithm using two video capture machines, each with a researcher. To this end, both the qualitative and quantitative aspects of the responses and analyzes were taken into considera-tion. Possible discrepancies between the participants' subjective perceptions and the algorithm's infer-ences were investigated, in order to identify possible limitations or biases of the facial analysis system in relation to the emotions studied.

This comparison of results is essential for evaluating the accuracy and validity of the facial detection algorithm, as well as to provide valuable insights into the effectiveness of the technology in interpreting facial expressions associated with the analyzed emotions. This rigorous and scientific methodological approach allows for a robust analysis of the agreement between objective analysis and subjective per-ception, contributing to scientific and technological advancement in this area.

Results and Discussion

For machine learning, the "Emotion Detection" dataset available on Kaggle was used, with a total of 35,685 examples of face images in gray scale of 48x48 pixels (Input). Images are categorized based on the emotion shown in facial expressions (happiness, neutral, sadness, anger, surprise, disgust, fear).

Table 1. Shows random examples of images from the machine learning dataset

Happiness (H2)			
Surprise (S2)			
Neutral (N2)			
Disgust (D1)			
Fear (F1)			
Anger (A1)			
Sadness (S1)			

Source: Created by the authors based on Kaggle (2023).

During the tests carried out to adjust the learning language, we observed that the model achieved positive accuracy in relation to the standards defined by the input data. Even with a reduced number of experiments (only one test, i.e., n=1), we found a statistically significant correlation between the language configurations and the inputs provided. This meant that the model was able to reliably identify and apply the patterns contained in the input data, which was good news.

The statistical calculations used to measure this accuracy indicate that the model is faithful to the inputs it receives. This result is promising, especially considering that the initial experiment had a limited sample size. This suggests that the approach used in testing has the potential to be broadly applicable in different contexts when n=1, making us more confident in the model's ability to adapt to diverse language needs and future tasks.

However, our efforts were not aimed at simply capturing the emotions presented in the sample in a simplistic way (n=1). Instead, we seek to provide results that reflect the emotions aroused in different time periods, such as during the screening of a film, a lecture, a concert, among other events. In this sense, we are faced with the complexity of several factors that could influence our results.

First, we recognize that complete satisfaction or happiness when watching a movie, class, or show is not constant. We do not always experience total happiness, represented by a smile on our face, throughout the event; At some point, we may have moments of reflection or neutrality.

Second, we realize that neutrality plays a significant role in our emotion ratings and therefore cannot be disregarded. We decided to consider it as a positive feeling, giving it a weight of 50.

Third, to address the issue of time during the presentation, we chose to discard the 'emotion capture time' factor, as we believed it would not significantly influence our ratings. Instead, we replace this factor by calculating the proportion of expressions of each emotion to the total expressions of all emotions.

However, if we wanted to analyze emotions at specific moments during expression capture, the 'emotion capture time' factor could not be ignored. For example, in a 120-minute film, how long does the audience experience joy? Or in a 1-hour lecture, at what point in the lecture (or topic) does the peak of happiness occur? This approach may be particularly relevant for performers in stand-up comedy and other similar fields.

Based on these considerations, we aim to improve the interpretation of data obtained through specific measurements. One of them is to highlight the importance of the value of neutrality, as previously presented. Furthermore, we sought to evaluate the system's performance in group settings, taking into account the accuracy of the results, as well as the speed and quality of processing multiple faces.

In the experiment with n=3 participants, conducted in a strictly controlled environment with consistent lighting conditions and two cameras, one with a resolution of 3840px x 2160px and the other with 1920px x 1080px pixels, both placed at a distance of 5 meters from the participants. The environment was strictly controlled, with lighting provided by white light, and all sources of noise in the images were strictly avoided.

In a specific experiment designed to challenge the system's performance, we asked a fourth member to walk behind the three participants, in order to partially hide their faces. It was observed that the system with n=3 presented satisfactory results in a controlled environment. However, when the fourth member walked behind the other participants, there were moments of interruption, especially when the faces were in profile.

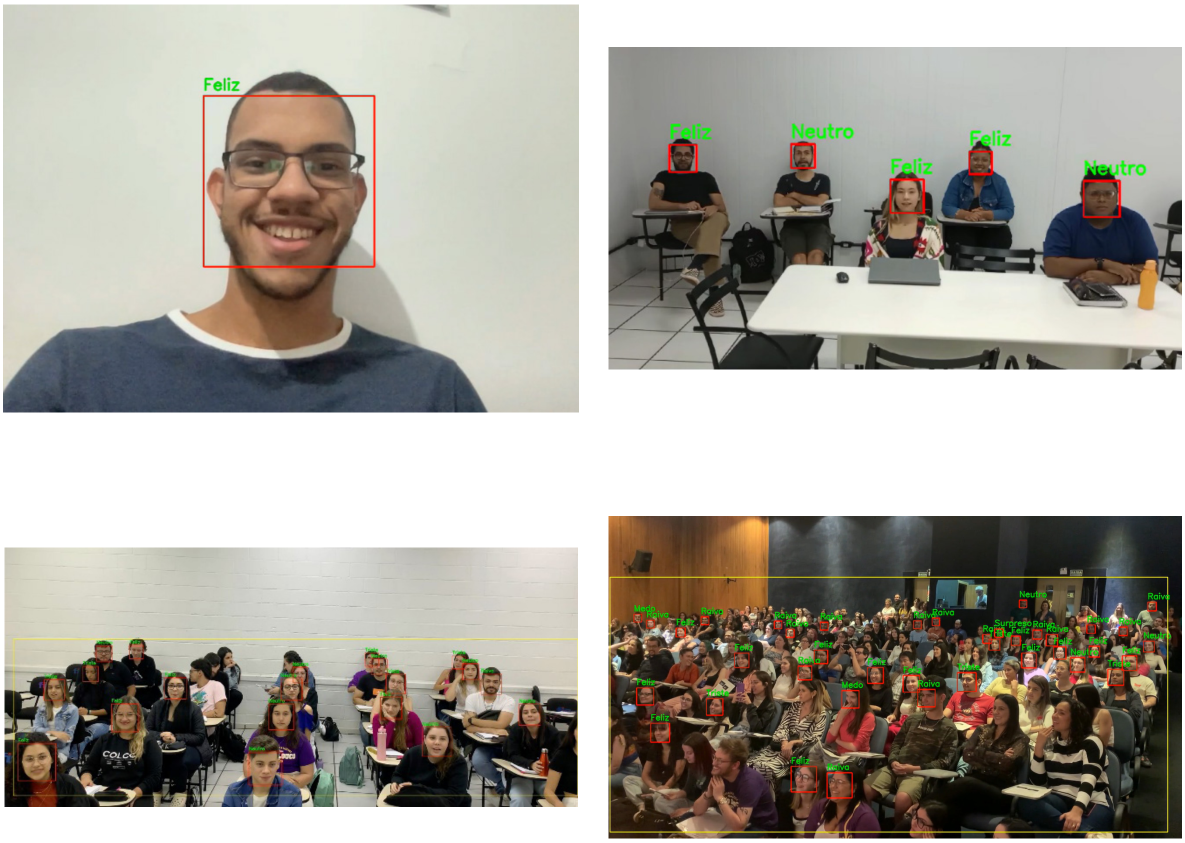
These analyzes seek to improve the understanding of the system's performance, identifying its capabilities and limitations in scenarios that simulate real-world situations, allowing us to optimize its effectiveness and applicability in different contexts.

In the experiment with n=5 participants, conducted in a strictly controlled environment with consistent lighting conditions and using the same cameras and light levels previously mentioned. In the context of this experiment, we applied depth measurements, with some participants positioned at a distance of 5 meters and others at a distance of 7 meters from the cameras.

Additionally, we explored the possibility that the algorithm may introduce racial bias, an issue highlighted by Rhue (2018; 2019), who noted that emotional analysis technology attributes more negative emotions to the faces of Black men than to the faces of White men. However, the experiment did not identify discrepancies based on racial criteria. This can be attributed, in part, to the diverse composition of the dataset that was used to train the algorithm, which encompasses a comprehensive representation of ethnic-racial groups and distinct age groups.

Furthermore, it is important to mention that the algorithm has undergone refinements to improve its language recognition capabilities, given notable flaws in similar algorithms, such as the incident involving Google, when the company's software mistakenly identified black people as 'gorillas'. This case exposed the continuous need for improvement, as highlighted by Yonatan Zunger, chief architect of social at Google, who stated that ""lots of work being done, and lots still to be done. But we're very much on it".

Image 1. Detection of emotions in some experiments (n=1; 5; 26; 280)



Source: authors (2023).

The experiment conducted with n=26 participants, although it provided reliable results, highlighted some substantial limitations. One of the notable limitations is related to the presence of noise in the captured images, particularly when direct lighting on the participants' faces is insufficient. This can occur due to the participants' gaze shifting or head movement, generating shadows that interfere with the quality of the images obtained.

However, the most preponderant challenge encountered is linked to the depth of the images, constituting a dilemma that combines architectural and resolving aspects of the images. In this context, it was noted that the higher resolution camera (3840 pixels x 2160 pixels) presented superior performance in the analysis of emotions, but, on the other hand, the processing speed was substantially reduced.

These complexities and trade-offs are congruent with the observations made by Bayoudh, Knani, and Hamdaoui (2022), which indicate that the deep learning community still seeks to find a more appropriate balance between the complexity in structuring models, the demands on computational power and real-time processing capabilities. The authors add that, in practice, autonomous systems, such as autonomous vehicles and healthcare robots, as well as other real-time embedded systems, consume more hardware, storage and battery resources than other emerging technologies, resulting in a lack of adaptation to future needs.

In our experiment, we chose to continue the research with a sample size of n=280 participants in an environment characterized by multiple interfering factors and lacking rigorous control. This environment presented several challenging variables, including background noise, inadequate yellow lighting, and a depth of field of 20 meters. The research was carried out in an auditorium, in the context of an academic event, in which the camera was positioned at a distance of 5 meters from the first row of participants.

The decision to proceed with the research under these conditions was based on the understanding that it was impractical to keep participants in a fixed position and looking directly at the camera, an approach similar to the iconic scene in Stanley Kubrick's film 'A Clockwork Orange', in which participants are kept in a fixed position and looking directly at the camera. The decision to continue the research taking into account the reality of the uncontrolled environment demonstrates the acceptance of inevitable variations and the confidence that the results obtained would still be representative and informative for the objectives of the study. The confidence margin, representing the range of uncertainty associated with an experiment or measurement where natural or unpredictable variations can be tolerated without compromising the interpretation of the results, played a crucial role in this decision.

At this stage of the experiment, we identified some significant limitations in the analysis of emotions, resulting in the invalidation of the results obtained. Several challenges were observed during the experiment, compromising the accuracy of emotional analysis.

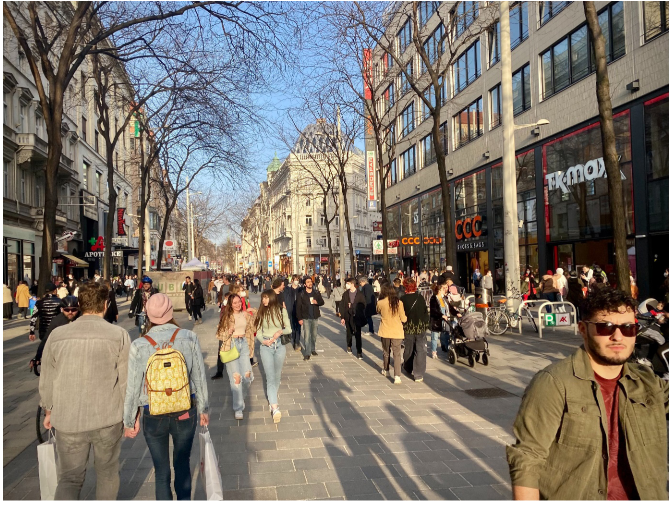
Firstly, the inadequate yellow lighting in the auditorium environment negatively impacted the clarity of facial images in certain contexts, contributing to a dark representation of the participants' faces. This suboptimal lighting condition posed challenges to the accurate analysis of emotions.

Furthermore, limitations related to the equipment used also proved to be relevant. Even when using a high-resolution camera (3840 pixels x 2160 pixels), which had demonstrated satisfactory results in previous experiments, the capture of facial images in the auditorium did not meet the standards necessary for adequate analysis. In this sense, equipment with a resolution equal to or greater than 8K (7680 pixels x 4320 pixels) would be necessary to obtain clearer images. However, such an upgrade would represent a substantial investment in processing hardware.

Finally, the issue of depth was also identified as a challenge. This is not necessarily a limitation of the system in relation to the calculation of emotional data, but rather an intrinsic complexity in the representation of three-dimensionality on a two-dimensional surface. This complexity affected the accuracy of emotion analysis, as depth perception is crucial in interpreting facial expressions. The combination of these technical and environmental challenges negatively impacted the experiment's ability to provide reliable results in analyzing participants' emotions.

Regarding the complexity underlying the representation of three-dimensionality on a two-dimensional surface, it is evident, in images 2 and 3, the persistence of unresolved issues in this context. It is clearly observed that, as the depth in the image increases, the resolution of the faces tends to gradually decrease, reaching a point where their identification becomes impractical. This finding highlights the intricate nature of the two-dimensional representation of three-dimensional objects.

Image 2. Crowd of people in an open environment



Source: Unsplash (2023).

Image 3. Crowd of people on the streets

Source: Unsplash (2023).

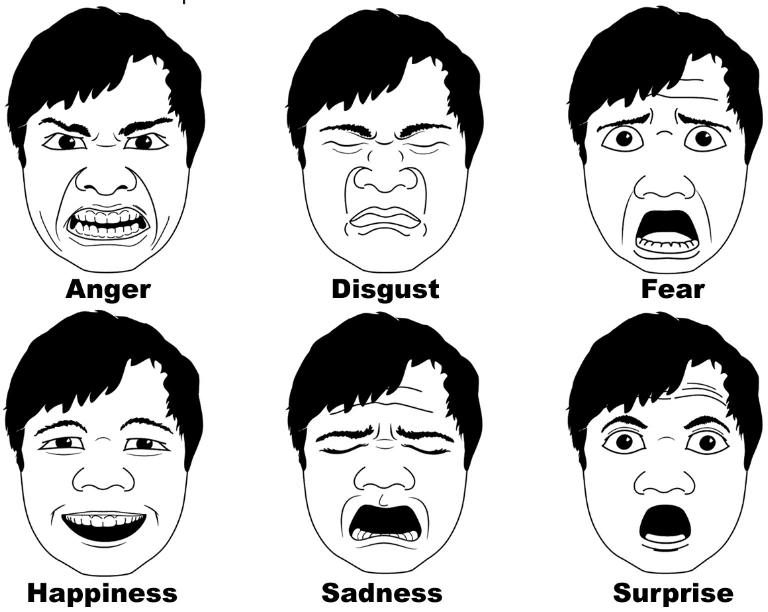
It is important to emphasize that the satisfactory results obtained in the experiment with n=1 participant should not be interpreted as an indication that the number of participants does not influence the results. Instead, the effectiveness of this experiment is attributed to the participant's proximity to the camera. The distance between the participant and the camera plays a crucial role as it directly affects the ability to capture facial details.

It is worth noting that the validity of the results in an experiment with n=280 participants is theoretically viable, as long as all participants can be positioned at a similar focal distance in relation to the camera. However, this condition becomes impractical due to the Principle of Impenetrability, which establishes the impossibility of physical superposition of all participants in the same depth of field. Therefore, the feasibility of large-scale experiment is restricted by limitations imposed by the laws of optics and the physical nature of objects.

Based on the results derived from the experiments conducted, a series of challenges of considerable magnitude emerge, highlighting, in particular, the question about the feasibility of a computer vision tool in discerning the individual's authentic emotional experience at the time of image acquisition. When examining Image 4, it is noticeable that the template used in phases 3 and 4 of the experiment predominantly reflects the Western paradigm of facial manifestations related to emoticons (Huang, 2023).

However, the question of the inherent complexity of emotions as phenomena rooted in the psychological system is raised, questioning whether they can be adequately captured through facial expressions alone. Could, perhaps, a face display traces of sadness without the individual actually experiencing such an emotion, due to characteristics intrinsic to their personality or influences from external factors, such as environmental context and cultural influences?

Image 4. Emotions template



Source: authors (2023).

Although the visualized emotion features reveal that the mouth and nose region contain the predominant information, while the eyes and ears contain secondary information when the neural network learns to perform facial emotion recognition (FER), this paradigm resembles the process of observation of emoticons by human beings (Huang, 2023). This view is supported by the opinion of Michio Kaku, who opined that programming emotions into a computer is not an extremely complex task (Mayekar, 2018). However, it is important to highlight that Theories of Emotions cannot be simplified in this way, since detecting affect represents a substantial challenge due to the conceptual nature of emotions, which are abstract quantities that cannot be measured directly, presenting imprecise boundaries. and considerable variations in individual expression and experience (Calvo; D'Mello, 2010).

It is crucial to recognize the need for multidisciplinary approaches in designing effective tools for emotion recognition, promoting collaboration between programmers and scientists in human areas, such as psychology. It is essential to understand that, despite the vast research literature on emotions and affective science, the literature on automatic emotion recognition has been predominantly influenced by computer scientists and artificial intelligence researchers, often ignoring the controversies inherent in the underlying psychological theories (Calvo; D'Mello, 2010).

This implies, in addition to analyzing the characteristics of emotional expression in facial images, the urgent need to consider the contribution of intrinsic emotional characteristics, whose manifestation is not readily visible. Although artificial neural networks demonstrate the ability to effectively identify visible cues present in the perioral and nasal regions of facial images for the purpose of recognizing emotional states, the intricate complexity that permeates the domain of human emotions should not be underestimated. Emotion, as a psychophysiological phenomenon, extends beyond mere facial expression, incorporating physiological factors, such as variation in heart rhythms, skin conductance and hormonal responses.

However, we also face the challenge of developing a comprehensive approach to emotion recognition, covering both the visible manifestations of facial expressions and the underlying physiological reactions. The perspective expressed by Michio Kaku regarding the programming of emotions in computational systems, although intriguing, should not be interpreted as an underestimation of the inherent challenges associated with this endeavor. Human emotions emerge from a complex interaction between biological factors, cognitive processes and sociocultural influences. They cannot be encapsulated by simplified algorithms, given the richness of inherent nuances and subtleties, which intertwine and manifest themselves in a highly individualized way.

Therefore, the implementation of computational emotion recognition demands a multidisciplinary approach that transcends the traditional limits of computing, embracing significant contributions from the areas of psychology, neuroscience and other related disciplines. Only through this interdisciplinary collaboration will it be possible to develop truly effective tools capable of capturing the full complexity of human emotions, recognizing them in their entirety, both in the visible dimensions and in the underlying physiological layers

Conclusion

In conclusion, when considering the results of the experiment involving the 'Emotion Detection' da-taset available on Kaggle, which consists of 35,685 examples of 48x48 pixel grayscale face images cate-gorized based on demonstrated emotions in facial expressions, several conclusions can be drawn.

During the tests conducted to adjust the language learning, the model demonstrated positive accuracy in relation to the standards defined by the input data, despite the limited number of experiments (only one test, i.e., n=1). It was possible to establish a statistically significant correlation between the language configurations and the inputs provided. These results indicate that the model is able to reliably identify and apply the patterns contained in the input data. Importantly, this conclusion is promising, especially given the limited size of the initial sample in the experiment, suggesting that the testing approach has the potential for broad applicability in different contexts when n=1.

However, it is essential to emphasize that the objective of the experiment was not limited to a simplis-tic analysis of the emotions presented in the sample (n=1). Instead, we sought to provide results that re-flected the emotions aroused at different time intervals. In this sense, the experiment faced the complexi-ty of several factors that could influence the results.

Regarding experiments with different sample sizes (n=3, n=5, n=26, n=280), each presented unique challenges and limitations. For instance, in the experiment with n=3 participants, despite satisfactory re-sults in a controlled environment, interruptions occurred when a fourth member walked behind the partic-ipants, particularly when their faces were in profile.

In the experiment with n=5 participants, carried out in a controlled environment, depth analysis was introduced, with participants positioned at different distances from the cameras. Additionally, the possi-bility of racial bias in the algorithm was explored, which was not identified due to the diversity of the training data set. The experiment with n=26 participants, although it provided reliable results, faced chal-lenges related to noise in the images and the complexity of depth representation. The choice of camera resolution also impacted processing speed.

Finally, the experiment with n=280 participants was conducted in an uncontrolled environment, simu-lating real-world situations. Despite significant challenges such as inadequate lighting and an extended depth of field, the decision to continue the research under these conditions was made based on an under-standing of the experiment's confidence margin. The results of these experiments provide valuable in-sights into the applicability and limitations of the emotion analysis model. They highlight the importance of considering the complexity of emotions in different contexts and the continued need for refinement and adaptation of the model to address these nuances. Furthermore, the technical and environmental challenges faced indicate areas where future improvements and investments can be directed to improve the accuracy and effectiveness of emotion analysis in real-world scenarios.

Referências

BAYOUDH, K.; KNANI, R.; HAMDAOUI, F.; *et al.* A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. **Vis Comput,** v. 38, p. 2939–2970, 2022. DOI: https://doi.org/10.1007/s00371-021-02166-7.

CALVO, R. A.; D'MELLO, S., Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applica-tions. **IEEE Transactions on Affective Computing**, v. 1, n. 1, p. 18-37, Jan. 2010. DOI: 10.1109/T-AFFC.2010.1.

HE, L.; WANG, G.; HU, Z., 2018. Learning depth from single images with deep neural network embedding focal length. **IEEE Transactions on Image Processing**, v. 27, n. 9, p. 4676-4689, 2018.

HUANG, Z. Y.; CHIANG, C. C.; CHEN, J. H.; *et al*. A study on computer vision for facial emotion recognition. **Sci Rep,** v. 13, 2023. DOI: https://doi.org/10.1038/s41598-023-35446-4.

LI, Z.; SNAVELY, N. Megadepth: Learning single-view depth prediction from internet photos. *In*: **Proceedings of the IEEE conference on computer vision and pattern recognition** (p. 2041-2050), 2018.

MAYEKAR, S. **Emotion AI Has Come To Light**. Analytics Insigth. 2018. Available at: https://www.analyticsinsight.net/emotion-ai-has-come-to-light/. Access in: 3 jan. 2023.

RHUE, L. **Emotion-reading tech fails the racial bias test**. The Conversation, 2019. Available at: https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404. Access in: 3 jan. 2023.

RHUE, L. Racial Influence on Automated Perceptions of Emotions. **Ssrn.** Nov. 2018. DOI: http://dx.doi.org/10.2139/ssrn.3281765.

VASILJEVIC, I.; *et al*. Diode: A dense indoor and outdoor depth dataset. **arXiv preprint** Ago. 2019. arXiv:1908.00463.

WANG, J.; ZHANG, S.; MARTIN, R. R. New advances in visual computing for intelligent processing of visual media and augmented reality. **Sci. China Technol. Sci.** v. 58, p. 2210–2211, 2015. DOI: https://doi.org/10.1007/s11431-015-5991-0.

Recebido em 03 de janeiro de 2023.
Aceito em 07 de novembro de 2023.