# Analysis of Gender Differences in Facial Expression Recognition Based on Deep Learning Using Explainable Artificial Intelligence

Cristina Manresa-Yee*, Silvia Ramis, José M. Buades

Universitat de les Illes Balears, Group of Computer Graphics, Computer Vision and IA. Maths and Computer Science Department, Palma (Spain)

Received 29 July 2022 | Accepted 3 March 2023 | Early Access 12 April 2023

## Abstract

Potential uses of automated Facial Expression Recognition (FER) cover a wide range of applications such as customer behavior analysis, healthcare applications or providing personalized services. Data for machine learning play a fundamental role, therefore, understanding the relevancy of the data in the outcomes is of utmost importance. In this work we present a study on how gender influences the learning of a FER system. We analyze with Explainable Artificial intelligence (XAI) techniques how gender contributes to the learning and assess which facial expressions are more similar regarding face regions that impact on the classification. Results show that there exist common regions in some expressions both for females and males with different intensities (e.g. happiness); however, there are other expressions like disgust, where important face regions differ. The insights of this work will help improving FER systems and understand the source of any inequality.

## Keywords

Explainable AI, Facial Expression Recognition, Gender Differences, Explainable Artificial Intelligence, XAI.

DOI: 10.9781/ijimai.2023.04.003

## I. Introduction

THAT men's and women's facial expressions differ is a well-known fact. Gender differences have been examined subjectively but also on physiological measures such as facial electromyography (EMG) or observable Action Units (AUs). Studies show differences in frequency [1]–[3], increased attention [4] or expressiveness specially in the expressions of happiness, fear, anger or disgust [1], [5], [6]. Regarding smiling, Dimberg and Lundquist [3] found that women evoked more facial muscle activity in response to happy faces. This result is consistent with earlier works [7]. Regarding fear, evidence shows that females were more facially expressive when presented with fear-relevant images (e.g. angry faces or snakes) with an increase of the activity on the corrugator supercilia [6]. Women also experience disgust with more intensity than men [4], however, anger is less likely to be displayed by females [8].

Six basic facial expressions -happiness, sadness, anger, surprise, disgust and fear- are recognized across different cultures [9]. Descriptions have been made about the face muscles involved in forming those expressions. The Facial Action Coding System (FACS) [10] describes anatomically all visually discernible facial movement by defining Action units (AUs), which are the actions of individual muscles or groups of muscles. Observing and coding a selection of AUs, Emotion FACS (EMFACS), humans can identify prototypical facial expressions that have been found to suggest certain emotions.

* Corresponding author.
E-mail address: cristina.manresa@uib.es

Fan, Lan and Li [11] analyzed two of these AUS, the AU6 (cheek raiser) and AU12 (lip corner puller) related to the smiling (happiness expression) and found that females were generally more expressive and presented a higher intensity value for AU12 (bigger smile) than males. McDuff et al. [1] also analyzed AUs (AU1, AU2, AU4, AU12 and AU15) to study gender differences. Their results found that women smiled more, and they presented more significantly inner brow raise actions, which are related with fear and sadness.

Houstis and Kiliaridis [12] did not use AUs, but they analyzed a set of facial distances when posing a rest pose, a lip pucker, and a posed smile. Their findings regarding gender, found that males had a vertical upward component more pronounced both in the posed smile and the lip pucker, while females had a more pronounced horizontal component in the posed smile.

There is extensive research in analyzing gender differences both for recognition (perceiver) [13]–[15] and generation (expresser) of facial expressions [16], [17]. However, in the case of current deep learning developments for automatic FER [18], due to their black box nature, it makes it difficult to assess the gender differences in recognition beyond comparisons of for example the accuracy rates or bias [19]–[22]. However, a deeper understanding on how men and women images contribute to the learning of the models could help improving the models or understanding the misclassifications. Further, research in this area is usually based on existing datasets which are not always gender balanced [18] and we even find datasets, such as JAFFE [23] , with only one gender (10 Japanese female subjects).

Explainable Artificial Intelligence (XAI) techniques can provide further information on the internal working on these models and make them more transparent. Although they do not include a gender

Please cite this article in press as:
C. Manresa-Yee, S. Ramis, J. M. Buades. Analysis of Gender Differences in Facial Expression Recognition Based on Deep Learning Using Explainable Artificial Intelligence, International Journal of Interactive Multimedia and Artificial Intelligence, (2023), http://dx.doi.org/10.9781/ijimai.2023.04.003

perspective, we find examples applying these techniques in automated FER to understand automatic emotional annotation [24], to improve transfer learning [25], [26] or to understand the influential face regions in the classification [27], [28]. Heimerl et al. [24] included XAI techniques in their emotional behavior annotation tool addressed to non-expert users. Humans assisted the automatic labeling -only for four out of Ekman's six basic emotions: happiness, sadness, anger and disgust- aided by confidence values of the predicted annotation, as well as visual explanations using XAI (LIME [29] , INNvestigate [30]). Schiller et al. [25] presented saliency maps to identify the most relevant face regions used for the face recognition. The saliency maps were generated by Layer-wise Relevance Propagations (LRP) [31] and by eye-tracking. Then, they evaluated both and transferred that knowledge by hiding the non-relevant information to speed up the training of the neural network in a new domain.

Weitz et al. [27] investigated a Convolutional Neural Network (CNN) trained to distinguish facial expressions of pain, happiness, and disgust. They applied two XAI methods: LRP and LIME. They observed that the CNN did not exclusively look at the face but also to the background of the image. Regarding pain, Prajod et al. [26] also presented a study on the effects of transfer learning for automatic FER for emotions to pain. They applied LRP saliency maps to visually compare and understand the most influent regions for the classification, both for emotion recognition and for pain recognition, and related those regions with AUs. The results showed that specific AUs related to the facial expressions of contempt and surprise were not relevant for pain recognition.

With a gender perspective in mind, in our previous exploratory study, we could sense differences in the learning [32] which motivated us to analyze more thoroughly the impact of gender in automated FER. Further, the interest in this field is due to the multiple and varied domains that can benefit from FER such as diagnosis and treatment of psychiatric illness [33], marketing psychology applications [34] or human computer interfaces [35]. Therefore, by studying the influence of gender differences in FER training, we can improve our understanding of which face regions are important and consider this knowledge to contribute with better models which will impact the applications based on FER.

The work is organized as follows: Section 2 describes the material used and the procedure followed for the study. Section 3 presents and discusses the main results regarding performance, and gender differences and similarities in the important face regions considered by the model. Finally, the main contributions and future line works are presented in the last Section.

## II. Materials and Methods

This section contains detailed description of data, data pre-processing and augmenting, the XAI approach used to understand the internal working of the model and the procedure followed.

### A. Dataset

We train our FER model on the AffectNet dataset [36], a well-known public dataset and widely used in FER. AffectNet comprises more than one million still images of facial expressions in the wild and covers both categorical and dimensional affect models. About half (approximately 440K images) of the images are manually annotated as one of Ekman's basic emotions [37] (anger, fear, disgust, sadness, surprise, happiness, contempt and neutral). Further, the dataset presents issues such as duplicates or non-face images because it was built through web-scraping. In order to study gender differences, we manually labelled images (Female - Male) and selected a similar number of images regarding the gender label and the facial expression (see Fig. 1). Although there are a few datasets with gender labels and facial expressions (e.g. RAFDB with around 30K images [38]), we selected AffectNet due to its size and wide application in FER. It is important to highlight that formally both sex or gender should be informed or self-reported, but on web-scrapped datasets that information is not available. According to the World Health Organization (WHO), sex refers to "the different biological and physiological characteristics of males and females" and gender refers to "the socially constructed characteristics of women and men – such as norms, roles and relationships of and between groups of women and men" [39]. The view of gender used in this paper is binary. When we refer to a particular gender, we are assuming this gender based on the visual characteristics of the individual in the image. Additionally, as Chen and Joo [40] identified, human-generated annotations in FER datasets can include biases (e.g. annotation biases between genders, especially when it comes to the happy and angry expressions). Therefore, this study is limited due to these reasons.

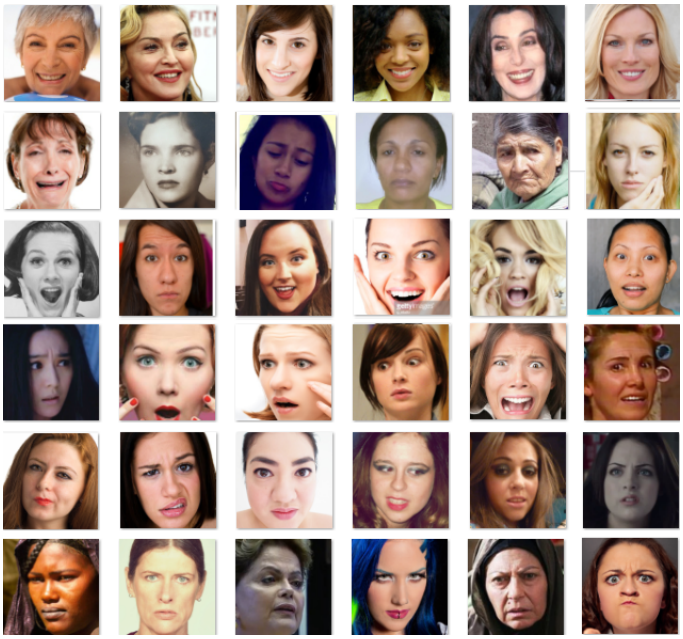


Fig. 1. Excerpt from the dataset of females' images for each expression.

The dataset used in the experimentation is comprised by an initial subset of 19044 images (1587 images x six expression x two gender) randomly selected where duplicated images, non-face images and images of individuals difficult to identify their gender by observation (e.g. babies or androgynous faces) were not considered. To meet the balanced dataset requirement, the subset was chosen considering the maximum number of manually labelled images per gender and facial expression, which was limited by the expression of Disgust. The Disgust expression counts with 4303 images in AffectNet, but 553 are duplicates, 290 images were undetermined and 1873 correspond to males and 1587 correspond to females. Therefore, the maximum number of images per gender and facial expression is 1587. Table I describes quantitatively the dataset in terms of expression and gender.

### B. Pre-Processing and Data Augmentation

Images are pre-processed and augmented before the training. The pre-processing steps carried out are face detection, face alignment and cropping. To detect the face, we apply the *a contrario* framework proposed by Lisani et al. [41]. For its alignment, we initially detect the eyes using the 68 facial landmarks proposed by Sagonas et al. [42]. We find the geometric centroid of each eye from these landmarks and compute the distance between them to draw a straight line and

TABLE I. Number of Images in Terms of Facial Expression and Gender From the Original Dataset to the Training and Testing Datasets. In Gray the Expressions Used in This Study. Dupl.: Duplicates. Und: Undetermined. F: Female. M: Male.

| | | | Gender label | | | Selected | | Pre-processing Face detected | | | Female datasets | | Male datasets | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | # | Dupl. | F | M | Und. | F | M | F | M | \|F-M\| | Test | Train | Test | Train |
| Neutral | 75374 | 5093 | | | | | | | | | | | | |
| Happiness | 134915 | 8430 | 2458 | 1590 | 581 | 1587 | 1587 | 1254 | 1182 | 72 | 253 | 1001 | 238 | 944 |
| Sadness | 25959 | 3425 | 1588 | 1691 | 613 | 1587 | 1587 | 1130 | 1038 | 92 | 227 | 903 | 209 | 829 |
| Surprise | 14590 | 1183 | 1588 | 1670 | 681 | 1587 | 1587 | 1183 | 1058 | 125 | 237 | 946 | 214 | 844 |
| Fear | 6878 | 1082 | 1608 | 1588 | 553 | 1587 | 1587 | 1150 | 1023 | 127 | 235 | 915 | 206 | 817 |
| Disgust | 4303 | 553 | 1587 | 1873 | 290 | 1587 | 1587 | 1230 | 1093 | 137 | 243 | 987 | 218 | 875 |
| Anger | 25382 | 3169 | 1588 | 3944 | 475 | 1587 | 1587 | 1233 | 1123 | 110 | 246 | 987 | 225 | 898 |
| Contempt | 4250 | 315 | | | | | | | | | | | | |
| None | 33588 | 2342 | | | | | | | | | | | | |
| Uncertain | 12145 | 943 | | | | | | | | | | | | |
| Non-face | 82915 | | | | | | | | | | | | | |
| Non-labeled | 6999 | | | | | | | | | | | | | |
| Total | 427298 | 26535 | 10417 | 12356 | 3193 | 9522 | 9522 | 7180 | 6517 | | 1441 | 5739 | 1310 | 5207 |

calculate the rotation angle. This angle is then used to align the eyes horizontally. Finally, we crop the face and resize it to 224x224 pixels to feed it into the network as input.

The pre-processing steps discarded images (see Table I) when the algorithm did not detect the face (e.g. the face was not completely visible or it was a side face).

Finally, to increase the number of images to train and add diversity, we augmented data by modifying lighting and appearance [43]. We use the gamma correction technique (see Eq. 1) to modify the lighting conditions, with four gamma values ($\gamma = 0.5$, $\gamma = 1.0$, $\gamma = 1.5$ and $\gamma = 2.0$).

$$y = \left(\frac{x}{255}\right)^{\frac{1}{\gamma}} \cdot 255 \tag{1}$$

where $x$ is the original image, $y$ is the new image and $\gamma$ is the value modified to change the illumination. To modify the appearance, we apply four-pixel translations of the image in both axes.

### C. Explainability Approach

There exist a varied number of XAI techniques [44] for explaining models and data both at global and local levels. In this work, to analyze visually the outcome of the system, we use LIME, Local Interpretable Model-agnostic Explanation [29]. LIME can be applied on any classifier and offers locally faithful explanations of the instance being explained. When LIME is applied to images, explanations are the parts of the image that are most positive towards a certain class. We present a novel strategy to use LIME to acquire global knowledge based on instance-level information in this context. To study the differences of the model for female and male globally, we merge all LIME masks obtained from the same training set, test set and expression in an average heatmap.

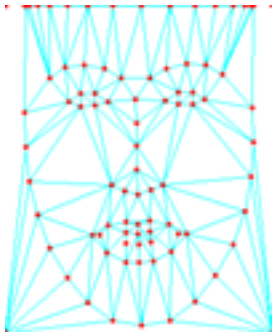


Fig. 2 Normalized location of the landmarks

Instead of computing the heatmap on the input space of the network (a 224 x 224 matrix), it is more relevant to compute it on the face representation space, that is, the parts of the face that are more relevant to identify one or another expression, regardless the orientation, translation or scaling of the image.

To compute the heatmap in the face space, we normalize all images with LIME applied to make the points of interest coincide. Faces are transformed so that the landmarks coincide with the normal form (see Fig. 2).

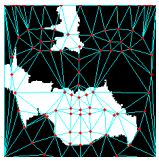
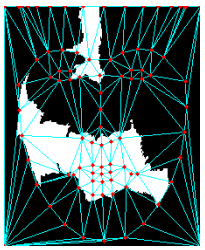
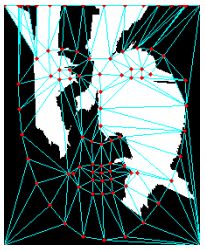
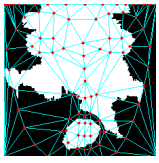
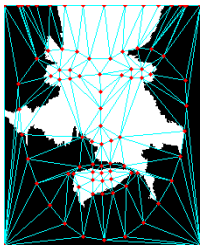
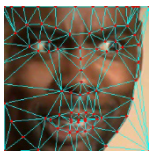
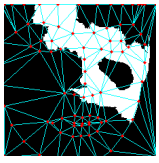
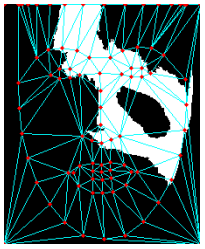
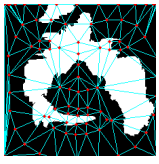
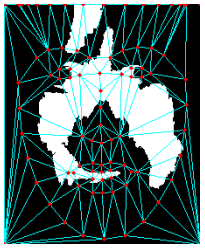
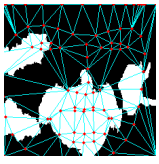
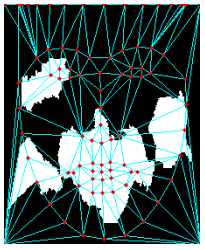
In Table II, we show the result of the process with six sample images, one for each expression starting from the original image: we identify the landmarks and compute the triangularization, then we apply LIME and superimpose the landmarks and triangularization, and finally we normalize this last image to the normal form. The detail of the process is described in the Algorithm 1.

**Algorithm 1**: Computing the normalized LIME image

**procedure** GetNormalizeLIME (*img*) ▷ original image
  *black_image* ⟵ create black image with *img* size
  $L'$ ⟵ landmarks(*black_image*) ▷ 68 normalized landmarks
  $L'$ ⟵$L'$ ∪ 17 top points ∪ four corners ▷ 89 landmarks
  $L$ ⟵ landmarks (*img*)
  $L$ ⟵$L$ ∪ 17 top points ∪ four corners ▷ 89 landmarks
  *lime_img* ⟵ *lime* (*img*) ▷ Compute LIME for original image
  *tri* ⟵ *delauny* (*L*) ▷ triangularization
  *norm_lime_img* ⟵ empty image (224×275) ▷ Create empty image
  **for each** pixel coordinate $p' \in$ *norm_lime_img* **do**
    $(v'_i, v'_j, v'_k)$ ⟵ triangle from $L'$ that contains $p'$ using *tri* triangles
    $(v_i, v_j, v_k)$ ⟵ triangle from $L$ that match $(v'_i, v'_j, v'_k)$ in $L'$
    $(c'_i, c'_j, c'_k)$ ⟵ p^' coordinates as lineal combination of $(v'_i, v'_j, v'_k)$
    $p$ ⟵ lineal combination (barycentric coordinates) of $(v_i, v_j, v_k)$ using $(c'_i, c'_j, c'_k)$ scalars
    *norm_lime_img* $[p']$ = *lime_img* $[p]$
  **return** *norm_lime_img* ▷ LIME image normalized

The merged heatmap is the average of all normalized LIME images that belong to the same training dataset, test dataset and facial expression. Then, we calculate distances between all the generated heatmaps, which is computed as one minus normalized correlation.

TABLE II. Normalization Process of the Images With LIME Applied to Merge the Important Regions for the Model

| Exp. | Original image with landmarks and triangulation | LIME with landmarks and triangulation | LIME transformed to normal face with landmarks and triangulation |
|---|---|---|---|
| Happiness |  |  |  |
| Sadness |  |  |  |
| Surprise |  |  |  |
| Fear |  |  |  |
| Disgust |  |  |  |
| Anger |  |  |  |

Distances generate a symmetric matrix that is used to cluster similar heatmaps applying the Ward's variance minimization method [45]. Finally, we visualize the dendrogram generated by clustering to analyze the arrangement of the clusters produced by the models.

### D. Procedure

We prepare three training and testing datasets: (1) a mixed dataset including a relatively gender-balanced number of images per facial expression; (2) a dataset with only images of females and (3) a dataset with only images of males.

Deep Convolutional Neural Networks (CNNs) have proved to be effective in numerous computer vision tasks [46], therefore, we use them to classify six facial expressions: anger, disgust, fear, happiness, sadness and surprise. In particular, we design our network based on the Inception v3 architecture [47]. The fully connected layer of Inception V3 network is replaced by a Global Average Pooling layer [48] and the softmax layer is modified to train the six classes (anger, sadness, fear, surprise, happiness, and disgust). Table III describes the hyper parameters of the network. The model is trained on a different dataset (mixed, only-female and only-male) and fine-tuned on imagenet [46]. We perform stratified 5-fold cross-validation, since these values have shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance [49], and report the mean classification accuracy. We highlight that the aim of the study is to analyze the influence of male and female datasets in the training, without focusing on improving the accuracy of the model. Then, the model is tested on all testing datasets (mixed, only-female and only-male) (see Table I for the details of number of images for each expression in each model). The different combinations of training and testing datasets used in this study are identified in Table IV.

TABLE III. Hyper-Parameters Used in the Inception V3 Network

| **Parameters** | |
|---|---|
| Weights (pre-trained model) | Imagenet |
| Learning Rate | $lr = 10^{-4}$ |
| Optimization algorithm | Adam [50] |
| Batch Size on training set | 128 |
| Batch Size on validation set | 32 |
| Epochs | 1 |

TABLE IV. IDs of Combinations of Training and Testing Sets

| | | Testing | | |
|---|---|---|---|---|
| | | Mixed | Female | Male |
| Training | Mixed | MI-MI | MI-FE | MI-MA |
| | Female | FE-MI | FE-FE | FE-MA |
| | Male | MA-MI | MA-FE | MA-MA |

To study differences between male and female, we apply the LIME merging procedure aforementioned to create the heatmaps to observe the face regions that are important for the model to classify images into a facial expression class. We build 36 heatmaps (3 training datasets (mixed, female and male) x 2 testing sets (female and male) x 6 expressions). In this case, LIME is configured to show the 5 most important features for the classification.

## III. Results

In this section we present the accuracy obtained and the gender differences observed regarding the face regions that influence the recognition.

### A. Accuracy

The training done with the mixed dataset achieves the best results in all cases (with accuracy around 53%-55%). Although we did not focus on improving the results, the accuracy is similar to other works that used AffectNet: Wang et al. [51] compiled several state-of-the-art methods on AffectNet with accuracies ranging from 47% to 60.23% for 7 or 8 expressions classification; Ngo et al. [52] achieved accuracies ranging from 46.07% to 60.7% using SE-ResNet-50 with different loss functions and classifying 8 expressions and Yen and Li [53] tested different architectures (ResNet-50, Xception, EfficientNet-B0, Inception, and DenseNet-121) with 8 expressions achieving accuracies ranging from 54% to 58% with class weight and data augmentation. Similar results are achieved both with the unbalanced trainings (male and female training) when tested with the mixed dataset (MA-MI, FE-MI). However, results decrease when testing with the other gender dataset, especially the training with female tested with male (FE-MA) (see Table V).

TABLE V. Mean Class-Wise Percentage Accuracy of the Models, Broken Down by Datasets

| | | Testing dataset | | |
|---|---|---|---|---|
| | | **Mixed** | **Male** | **Female** |
| **Training dataset** | **Mixed** | 53.83 | 52.73 | 54.84 |
| | **Male** | 47.12 | 47.80 | 46.51 |
| | **Female** | 47.61 | 42.86 | 51.98 |

Observing the confusion matrices (see Tables VI, VII and VIII), happiness is the best recognized expression for all training and testing datasets, except for FE-MA (female training dataset and male testing dataset). Although accuracy is high both for the mixed and male trainings (above 75%) with all testing datasets, female training dataset values are around 70% only with the female testing datasets and achieves lower values with the other testing datasets (lower than 56%). It is noteworthy that the higher values testing with females are achieved both with the male (MA-FE) and mixed (MI-FE) training datasets, which might be because of the higher expressiveness of females when smiling [7].

In the case of mixed training, there is a similar behavior both for the testing with male and female (MI-MA, MI-FE), obtaining the worse classifications for the fear and anger expressions. Fear is highly misclassified with surprise, and anger with disgust and sadness. Even humans have difficulties identifying facial expressions such as disgust and anger [25].

In general, both surprise and anger are not well recognized when training with males but recognizing anger for females achieves only a 25% of accuracy (MA-FE). When training with females, surprise and fear are low recognized both for the mixed and male testing (FE-MI and FE-MA) and for female testing (FE-FE), fear and sadness present the lowest accuracy.

In FE-MA (female training dataset and male testing dataset), expressions tend to be classified as angry. That means that from the total of classifications, anger is the expression mostly selected by the CNN (see last row of each confusion matrix in Table VII). On the contrary, when using the MA-FE (male training dataset and female testing dataset), expressions tend to be classified into the happiness expression (see last row of each confusion matrix in Table VIII). Lastly, the mixed training dataset tends not to classify expressions into the fear class (see last row of each confusion matrix in Table VII and VIII).

TABLE VI. Confusion Matrix (Tested With Mixed Dataset, Trained With All Datasets). Values Are Expressed as Percentages. Last Row Is the Sum of All

| **Train/Test** | | **Mix** | | | | | |
|---|---|---|---|---|---|---|---|
| | | Ha | Sa | Su | Fe | Di | An |
| **Mix** | Ha | 79.07 | 4.43 | 6.29 | 0.86 | 5.58 | 3.78 |
| | Sa | 5.35 | 50.44 | 9.98 | 4.66 | 12.85 | 16.72 |
| | Su | 8.25 | 8.25 | 54.63 | 13.70 | 6.82 | 8.35 |
| | Fe | 3.27 | 10.83 | 30.32 | 36.89 | 8.87 | 9.81 |
| | Di | 6.97 | 11.24 | 6.49 | 4.21 | 54.03 | 17.06 |
| | An | 4.04 | 16.78 | 9.37 | 3.78 | 18.09 | 47.94 |
| | | 107.0 | 102.0 | 117.1 | 64.1 | 106.3 | 103.7 |
| **Ma** | Ha | 78.19 | 5.30 | 5.46 | 2.67 | 6.41 | 1.97 |
| | Sa | 9.50 | 41.58 | 10.68 | 11.27 | 14.52 | 12.43 |
| | Su | 9.05 | 10.89 | 37.75 | 29.10 | 7.76 | 5.45 |
| | Fe | 5.01 | 10.28 | 22.05 | 47.13 | 9.20 | 6.32 |
| | Di | 12.10 | 14.24 | 6.21 | 9.63 | 45.28 | 12.54 |
| | An | 8.57 | 19.83 | 8.74 | 10.41 | 19.62 | 32.83 |
| | | 122.4 | 102.1 | 90.9 | 110.2 | 102.8 | 71.6 |
| **Fe** | Ha | 55.51 | 8.95 | 7.23 | 3.16 | 17.07 | 8.08 |
| | Sa | 1.48 | 43.50 | 8.96 | 7.20 | 16.67 | 22.18 |
| | Su | 4.64 | 10.13 | 44.94 | 20.99 | 7.58 | 11.73 |
| | Fe | 1.57 | 10.57 | 24.64 | 40.35 | 9.25 | 13.63 |
| | Di | 3.05 | 12.45 | 5.99 | 6.51 | 50.15 | 21.85 |
| | An | 1.44 | 16.77 | 6.28 | 6.50 | 17.79 | 51.22 |
| | | 67.7 | 102.4 | 98.0 | 84.7 | 118.5 | 128.7 |

TABLE VII. Confusion Matrix (Tested With Male Dataset, Trained With All Datasets). Values Are Expressed as Percentages. Last Row Is the Sum of All

| **Train/Test** | | **Male** | | | | | |
|---|---|---|---|---|---|---|---|
| | | Ha | Sa | Su | Fe | Di | An |
| **Mix** | Ha | 75.64 | 5.58 | 5.84 | 1.10 | 6.52 | 5.33 |
| | Sa | 5.20 | 52.59 | 7.92 | 3.37 | 12.80 | 18.12 |
| | Su | 6.72 | 11.24 | 51.26 | 15.20 | 6.51 | 9.07 |
| | Fe | 3.12 | 11.24 | 28.64 | 36.48 | 9.98 | 10.54 |
| | Di | 6.50 | 12.45 | 6.49 | 4.11 | 50.96 | 19.49 |
| | An | 3.83 | 18.21 | 7.91 | 4.28 | 16.31 | 49.45 |
| | | 101.0 | 111.3 | 108.1 | 64.6 | 103.1 | 112.0 |
| **Ma** | Ha | 75.79 | 5.92 | 4.74 | 1.95 | 8.89 | 2.71 |
| | Sa | 8.68 | 41.96 | 8.87 | 8.59 | 15.02 | 16.88 |
| | Su | 8.23 | 12.00 | 35.77 | 26.70 | 8.90 | 8.41 |
| | Fe | 4.88 | 9.88 | 21.13 | 45.84 | 10.48 | 7.79 |
| | Di | 9.34 | 14.26 | 5.58 | 8.22 | 46.06 | 16.54 |
| | An | 5.53 | 17.20 | 6.67 | 9.11 | 20.07 | 41.42 |
| | | 112.5 | 101.2 | 82.8 | 100.4 | 109.4 | 93.8 |
| **Fe** | Ha | 39.86 | 13.11 | 5.33 | 4.32 | 24.62 | 12.77 |
| | Sa | 0.58 | 43.63 | 4.64 | 6.36 | 20.12 | 24.68 |
| | Su | 2.65 | 13.88 | 32.14 | 25.89 | 9.45 | 15.99 |
| | Fe | 1.27 | 11.43 | 20.45 | 40.02 | 9.88 | 16.94 |
| | Di | 1.19 | 14.09 | 4.94 | 6.32 | 47.31 | 26.15 |
| | An | 0.44 | 18.44 | 2.94 | 6.96 | 17.01 | 54.21 |
| | | 46.0 | 114.6 | 70.4 | 89.9 | 128.4 | 150.7 |

TABLE VIII. Confusion Matrix (Tested With Female Dataset, Trained With All Datasets). Values Are Expressed as Percentages. Last Row Is the Sum of All

| Train/Test | | Female | | | | | |
|---|---|---|---|---|---|---|---|
| | | Ha | Sa | Su | Fe | Di | An |
| Mix | Ha | 82.30 | 3.34 | 6.71 | 0.64 | 4.70 | 2.32 |
| | Sa | 5.50 | 48.46 | 11.89 | 5.84 | 12.89 | 15.42 |
| | Su | 9.63 | 5.58 | 57.66 | 12.34 | 7.10 | 7.69 |
| | Fe | 3.39 | 10.47 | 31.77 | 37.32 | 7.88 | 9.16 |
| | Di | 7.38 | 10.17 | 6.48 | 4.31 | 56.78 | 14.87 |
| | An | 4.21 | 15.49 | 10.71 | 3.33 | 19.71 | 46.54 |
| | | 112.4 | 93.5 | 125.2 | 63.8 | 109.1 | 96.0 |
| Ma | Ha | 80.45 | 4.71 | 6.14 | 3.36 | 4.07 | 1.28 |
| | Sa | 10.26 | 41.24 | 12.36 | 13.74 | 14.04 | 8.36 |
| | Su | 9.79 | 9.89 | 39.50 | 31.27 | 6.75 | 2.80 |
| | Fe | 5.12 | 10.64 | 22.85 | 48.29 | 8.09 | 5.01 |
| | Di | 14.55 | 14.23 | 6.75 | 10.90 | 44.61 | 8.96 |
| | An | 11.35 | 22.22 | 10.62 | 11.61 | 19.20 | 25.00 |
| | | 131.5 | 102.9 | 98.2 | 119.2 | 96.8 | 51.4 |
| Fe | Ha | 70.26 | 5.02 | 9.01 | 2.08 | 9.95 | 3.67 |
| | Sa | 2.31 | 43.38 | 12.95 | 7.97 | 13.51 | 19.88 |
| | Su | 6.42 | 6.77 | 56.38 | 16.61 | 5.90 | 7.92 |
| | Fe | 1.82 | 9.81 | 28.36 | 40.65 | 8.69 | 10.66 |
| | Di | 4.72 | 10.99 | 6.91 | 6.68 | 52.67 | 18.04 |
| | An | 2.36 | 15.26 | 9.33 | 6.08 | 18.48 | 48.49 |
| | | 87.9 | 91.2 | 122.9 | 80.0 | 109.2 | 108.7 |

### *B. Gender Differences in FER: Regions of Influence for Classification*

Table IX and X present the merged heatmaps for the LIME images which help us understand those image regions that impact the classification of the image into a class. It is important to remark that even if heatmaps are similar between expressions, the network could be observing different features in those zones (e.g., the presence of frown or not). To focus on the differences in the important zones between male and female, we used the heatmaps of FE-FE and MA-MA to calculate their subtraction. We obtained three images per expression showing the absolute difference, the difference between FE-FE and MA-MA, and the difference between MA-MA and FE-FE (see Table XI).

Observing the difference in the heatmaps, the disparity between female and male datasets for the sadness, fear and anger expressions is lower than happiness, disgust and surprise, which are the most different regarding gender.

In the happiness expression, the network trained both with males and females give importance to the lower face region (see Table IX), but in female training, this zone is more highlighted, and in male training, cheeks are also important.

In the case of disgust, the female dataset (FE-FE) focuses on the lower region face (the mouth and chin), whereas male datasets (MA-MA) accentuate the central zone comprised between the mouth and the eyes (up to the temples) (see Table X).

In the surprise expression, similar heatmaps are achieved with both male and female datasets (see Table IX), however, female datasets (FE-FE) highlight with more intensity the upper-middle face.

Heatmaps for surprise and fear are quite similar (see Table IX), this could be the reason for their misclassifications (see Fig. 3,

TABLE IX. Heatmaps of Merged LIME Explanations for Each Expression, Training and Testing Datasets

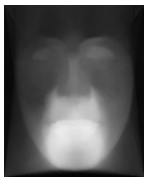
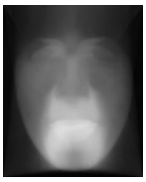
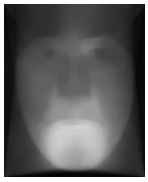
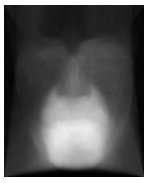
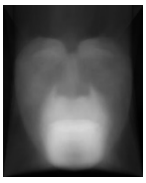
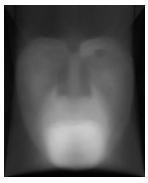
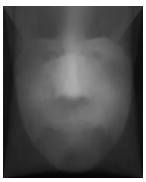
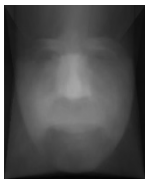
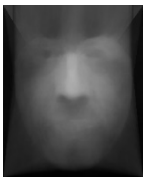
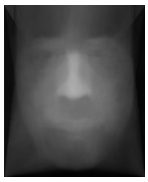
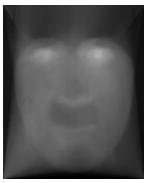
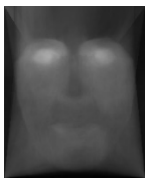
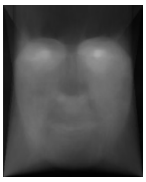
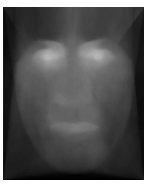
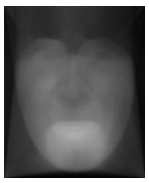
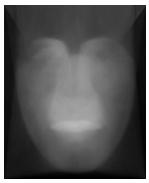
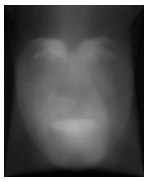
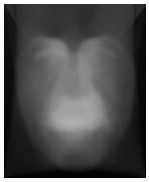
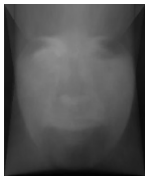
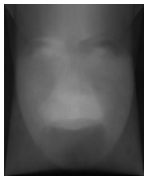
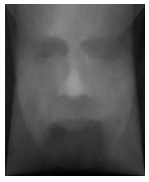
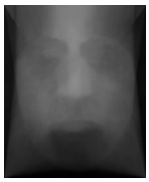
| | | Training Female | Training Mixed | Training Male |
|---|---|---|---|---|
| Happiness | Test Fe |  |  |  |
| | Test Ma |  |  |  |
| Sadness | Test Fe |  |  |  |
| | Test Ma |  |  |  |
| Surprise | Test Fe |  |  |  |
| | Test Ma |  |  |  |
| Fear | Test Fe |  |  |  |
| | Test Ma |  |  |  |

TABLE X. Heatmaps of Merged LIME Explanations for Each Expression, Training and Testing Datasets (Cont.)

| | | Training Female | Training Mixed | Training Male |
|---|---|---|---|---|
| Disgust | Test Fe |  |  |  |
| | Test Ma |  |  |  |
| Anger | Test Fe |  |  |  |
| | Test Ma |  |  |  |

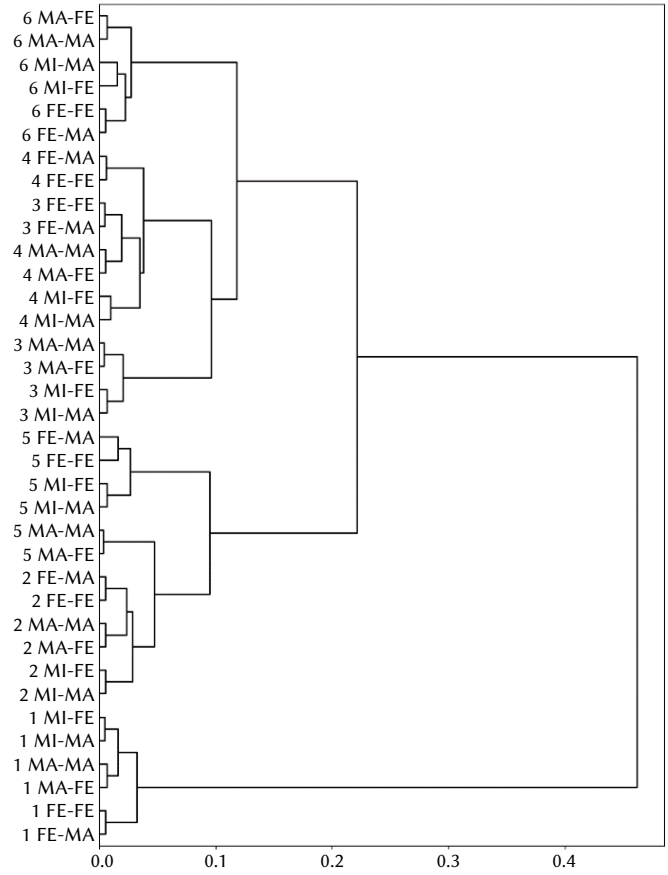
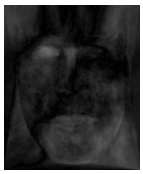
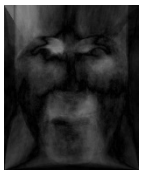
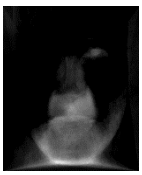
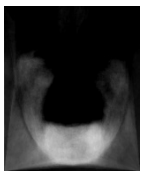
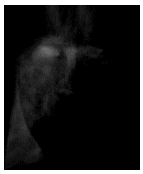
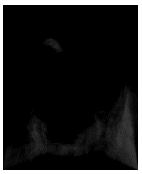



Fig. 3. Dendrogram with expressions happiness (1), sadness (2), surprise (3), fear (4), disgust (5) and anger (6). Ward's method is used to join clusters. Distance between clusters is computed as one minus normalized correlation.

TABLE XI. Differences Between Heatmaps for Each Expression. First Row: Absolute Value of the Difference Between the Heatmaps MA-MA and FE-FE. Second Row: Difference Between the Heatmaps FE-FE and MA-MA. Third Row: Difference Between the Heatmaps MA-MA and FE-FE. The Scalar Is Just Used to Improve the Visualization of the Differences. FE-FE Is Represented With F, MA-MA Is Represented With M

| | Happiness | Sadness | Surprise | Fear | Disgust | Anger |
|---|---|---|---|---|---|---|
| 3*\|f-m\| |  |  |  |  |  |  |
| 3*Max(f-m, 0) |  |  |  |  |  |  |
| 3*Max(m-f, 0) |  |  |  |  |  |  |

where the heatmaps for these expressions are grouped hierarchically based on similarity). Further, observing the grouped clusters in the dendrogram, happiness is the most different expression, which means that the important facial regions for learning are different to the other expressions and may help in the recognition.

The dendrogram also highlights the similarities of the heatmaps built using the mixed, male or female datasets for each expression. In general, expressions created with all datasets focus on similar regions, except for disgust trained with male datasets and surprise trained with female datasets. The first union of branches is mainly between those heatmaps belonging to the same training dataset and expression, meaning that the influence of the LIME images provided by the testing datasets is scarce. The next unions of branches depend mainly on the expression (with some exceptions, disgust in male training and surprise in female training), meaning that independently of the datasets, the expression influences the heatmap created. The similarities between heatmaps indicate that fear and surprise are the most similar and they share regions with anger, then sadness and disgust also coincide in face regions. And as already commented, happiness is the most different expression regarding the important face regions.

Although in some expressions an influence of the face regions indicated by Ekman [9] are shared (e.g. lower face for happiness), in general there is no direct relationship with them or even with AUs.

However, as Disgust is the most different expression regarding face regions, when planning a new study with limited resources, endeavors to achieve extra Disgust images from both genders can benefit FER. Summarizing, as highlighted regions are different in several expressions, the lack of diversity in the training dataset may impact the misclassification of facial expressions.

## IV. Conclusion

Currently FER is a relevant area of research due to its wide range of applications, and frequently new approaches are validated using well-known datasets or models. However, especially in web-scrapped datasets, we cannot assure that they include an evenly distributed number of images of individuals in terms of sensitive attributes such as gender. Further, analyzing gender differences will help improve FER systems and understand the source of any inequality o misclassification. Considering this, we questioned: What are the gender differences when training a ML system for FER?

The aim of the present study was to study how gender impacts in the learning of a CNN. We explained comprehensively with XAI techniques the differences and similarities of the important face regions for the model to classify an expression into a class, considering gender.

A first contribution of the work is the novel explanation technique used for the comparison, that is, the creation of a unique heatmap based on the individual results of applying LIME on each image. By merging the instance-level information in a normalized image, we acquire global knowledge about the functioning of the CNN. Then, heatmaps can be used to calculate and analyze the similarities and differences among expressions and gender. This technique can be transferred to other similar FER studies.

Regarding FER, as expected, gender-balanced training datasets improve FER accuracy, achieving similar likelihood for positive and negative outcomes. However, unbalanced datasets do not affect in the same manner all the face expressions. Results show that training with male datasets achieve better results than training with female datasets, this could be related with the claims of women being generally more expressive than men and being better senders of nonverbal information [3], [54], therefore, the training with less exaggerated expressions could transfer better. However, the expression of anger is an exception, which can relate with the literature than men pose anger more intensely.

Analyzing the performance, the expression of happiness is globally well recognized, with the exception of female training tested on males. In addition, it is interesting how female training tends to classify frequently male images into the anger class, whereas, in the opposite way, male training tends to classify expressions by women as happiness. And the mixed training, decides frequently to classify images into other classes before using the fear expression.

Lastly, the findings of the comprehensive study of the important face regions for the neural network show that there exist common regions in some expressions both for females and males with different intensities (e.g., happiness); however, in expressions like disgust, face regions are different. Therefore, when datasets are not balanced, these differences can impact the correct classification of facial expressions. In addition, regarding the differences between gender and expression in the face regions important for the model, we observed that the expression is more influential than the training or testing datasets.

As Xu et al. [21] commented, there is a need for the research community to invest effort in creating facial expression datasets with explicit labels regarding sensitive attributes. The gender labelled file used in this study is available in https://github.com/josebambu/AffectNetGenderLabelling/.

We note that the results obtained are achieved with a particular relevant dataset (AffectNet dataset) and model (Inception), therefore, future work lines are to study if different datasets and neural networks behave similarly as this study and analyze if important facial regions for the network coincide with human perception, in order to build more human-based models. Further, it would be of interest to apply the methodology to other types of images (e.g. thermal [55] or depth [56]), and other fields of study such as face identification [57] or pain detection [27].

## Acknowledgment

This work has been supported by the Agencia Estatal de Investigación, project PID2019-104829RA-I00 / MCIN/ AEI / 10.13039/501100011033, EXPLainable Artificial INtelligence systems for health and well-beING (EXPLAINING).

## References

[1] D. McDuff, E. Kodra, R. el Kaliouby, and M. LaFrance, "A large-scale analysis of sex differences in facial expressions," *PLOS ONE*, vol. 12, no. 4, pp. 1–11, 2017.

[2] L. Cattaneo, V. Veroni, S. Boria, G. Tassinari, and L. Turella, "Sex Differences in Affective Facial Reactions Are Present in Childhood," *Frontiers in Integrative Neuroscience*, vol. 12, 2018.

[3] U. Dimberg and L.-O. Lundquist, "Gender differences in facial reactions to facial expressions.," *Biological Psychology*, vol. 30, no. 2. Elsevier Science, Netherlands, pp. 151–159, 1990.

[4] M. A. Kraines, L. J. A. Kelberer, and T. T. Wells, "Sex differences in attention to disgust facial expressions," *Cognition and Emotion*, vol. 31, no. 8, pp. 1692–1697, 2017.

[5] K. AM and G. AH., "Sex differences in emotion: expression, experience, and physiology," *J Pers Soc Psychol.*, vol. 74, no. 3, pp. 686–703, 1998.

[6] M. Thunberg and U. Dimberg, "Gender Differences in Facial Reactions to Fear-Relevant Stimuli," *Journal of Nonverbal Behavior*, vol. 24, no. 1, pp. 45–51, 2000.

[7] G. E. Schwartz, S. -L Brown, and G. L. Ahern, "Facial Muscle Patterning and Subjective Experience During Affective Imagery: Sex Differences.," *Psychophysiology*, vol. 17, pp. 75–82, 1980.

[8] C. Evers, A. H. Fischer, and A. S. R. Manstead, "Gender and emotion

regulation: a social appraisal perspective on anger.," in *Emotion regulation and well-being*., Evers, Catharine: Department of Clinical and Health Psychology, Utrecht University, P.O. Box 80140, Utrecht, Netherlands, 3508 TC, c.evers@uu.nl: Springer Science + Business Media, 2011, pp. 211–222.

[9] P. Ekman, "Universals and cultural differences in facial expressions of emotion," *Nebraska Symposium on Motivation*, vol. 19, pp. 207–283, 1971.

[10] P. Ekman and W. . Friesen, *Facial action coding system: manual*. Palo Alto, Calif.: Consulting Psychologists Press. OCLC: 5851545, 1978.

[11] Y. Fan, J. C. K. Lam, and V. O. K. Li, "Demographic effects on facial emotion expression: an interdisciplinary investigation of the facial action units of happiness," *Scientific Reports*, vol. 11, no. 1, p. 5214, 2021.

[12] O. Houstis and S. Kiliaridis, "Gender and age differences in facial expressions," *European Journal of Orthodontics*, vol. 31, no. 5, pp. 459–466, 2009.

[13] T. S. H. Wingenbach, C. Ashwin, and M. Brosnan, "Sex differences in facial emotion recognition across varying expression intensity levels from videos," *PloS one*, vol. 13, no. 1, pp. e0190634–e0190634, Jan. 2018.

[14] B. Montagne, R. P. C. Kessels, E. Frigerio, E. H. F. de Haan, and D. I. Perrett, "Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?," *Cognitive Processing*, vol. 6, no. 2, pp. 136–141, 2005.

[15] R. Campbell *et al.*, "The classification of 'fear' from faces is associated with face recognition skill in women," *Neuropsychologia*, vol. 40, no. 6, pp. 575–584, 2002.

[16] A. K. Vail, J. F. Grafsgaard, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Gender Differences in Facial Expressions of Affect During Learning," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 2016, pp. 65–73.

[17] J. C. Borod, E. Koff, and B. White, "Facial asymmetry in posed and spontaneous expressions of emotion," *Brain and Cognition*, vol. 2, no. 2, pp. 165–175, 1983.

[18] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689–694, 2020.

[19] A. Domnich and G. Anbarjafari, "Responsible AI: Gender bias assessment in emotion recognition," *arXiv*, pp. 1–19, 2021.

[20] M. Deramgozin, S. Jovanovic, H. Rabah, and N. Ramzan, *A Hybrid Explainable AI Framework Applied to Global and Local Facial Expression Recognition*. 2021.

[21] T. Xu, J. White, S. Kalkan, and H. Gunes, "Investigating Bias and Fairness in Facial Expression Recognition," in *ECCV Workshops 2020*, 2020.

[22] Z. Wang *et al.*, "Towards Fairness in Visual Recognition: Effective Strategies for Bias Mitigation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8916-8925.*, 2020.

[23] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese Female Facial Expression (JAFFE) Dataset." Zenodo, 1998.

[24] A. Heimerl, K. Weitz, T. Baur, and E. Andre, "Unraveling ML Models of Emotion with NOVA: Multi-Level Explainable AI for Non-Experts," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 1–13, 2020.

[25] D. Schiller, T. Huber, M. Dietz, and E. André, "Relevance-Based Data Masking: A Model-Agnostic Transfer Learning Approach for Facial Expression Recognition," *Frontiers in Computer Science*, vol. 2, p. 6, 2020.

[26] P. Prajod, D. Schiller, T. Huber, and E. Andr'e, "Do Deep Neural Networks Forget Facial Action Units? - Exploring the Effects of Transfer Learning in Health Related Facial Expression Recognition," *ArXiv*, vol. abs/2104.0, 2021.

[27] K. Weitz, T. Hassan, U. Schmid, and J.-U. Garbas, "Deep-learned faces of pain and emotions: Elucidating the differences of facial expressions with the help of explainable AI methods," *Technisches Messen*, vol. 86, no. 7–8, pp. 404–412, 2019.

[28] G. del Castillo Torres, M. F. Roig-Maimó, M. Mascaró-Oliver, E. Amengual-Alcover, and R. Mas-Sansó, "Understanding How CNNs Recognize Facial Expressions: A Case Study with LIME and CEM," *Sensors*, vol. 23, no. 1, 2023.

[29] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[30] M. Alber *et al.*, "iNNvestigate Neural Networks!," *Journal of Machine Learning Research*, vol. 20, no. 93, pp. 1–8, 2019.

[31] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," *PLOS ONE*, vol. 10, no. 7, pp. 1–46, 2015.

[32] C. Manresa-Yee and S. Ramis, "Assessing Gender Bias in Predictive Algorithms Using EXplainable AI," in *Proceedings of the XXI International Conference on Human Computer Interaction*, 2021.

[33] K. Grabowski *et al.*, "Emotional expression in psychiatric conditions: New technology for clinicians," *Psychiatry and Clinical Neurosciences*, vol. 73, no. 2, pp. 50–62, 2019.

[34] A. M. Barreto, "Application of facial expression studies on the field of marketing," *Emotional expression: the brain and the face*, vol. 9, no. June, pp. 163–189, 2017.

[35] S. Medjden, N. Ahmed, and M. Lataifeh, "Adaptive user interface design and analysis using emotion recognition through facial expressions and body posture from an RGB-D sensor," *PLoS ONE*, vol. 15, no. 7, p. e0235908, 2020.

[36] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 01, pp. 18–31, 2019.

[37] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.

[38] S. Li and W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[39] World Health Organization, "Gender," *Available online: https://www.who.int/europe/health-topics/gender.* .

[40] Y. Chen and J. Joo, "Understanding and Mitigating Annotation Bias in Facial Expression Recognition," in *ICCV 2021*, 2021.

[41] J.-L. Lisani, S. Ramis, and F. Perales, "A Contrario Detection of Faces: A Case Example," *SIAM Journal on Imaging Sciences*, vol. 10, pp. 2091–2118, Jan. 2017.

[42] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge," in *2013 IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403.

[43] S. Ramis, J. . Buades, F. J. Perales, and C. Manresa-Yee, "A Novel Approach to Cross dataset studies in Facial Expression Recognition," *Multimedia Tools and Applications*.

[44] A. Barredo Arrieta *et al.*, "Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, no. October 2019, pp. 82–115, 2020.

[45] J. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[46] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[47] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[48] M. Lin, C. Qiang, and Y. Shuicheng, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[49] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, New York, 2013.

[50] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, 2015.

[51] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6896–6905, 2020.

[52] Q. T. Ngo and S. Yoon, "Facial expression recognition based on weighted-cluster loss and deep transfer learning using a highly imbalanced dataset," *Sensors (Switzerland)*, vol. 20, no. 9, 2020.

[53] C.-T. Yen and K.-H. Li, "Discussions of Different Deep Transfer Learning Models for Emotion Recognitions," *IEEE Access*, vol. 10, pp. 102860–102875, 2022.

[54] H. G. Wallbott, "Big girls don't frown, big boys don't cry—Gender differences of professional actors in communicating emotion via facial expression.," *Journal of Nonverbal Behavior*, vol. 12, no. 2, pp. 98–106, 1988.

[55] N. K. Benamara, E. Zigh, T. B. Stambouli, and M. Keche, "Towards a Robust Thermal-Visible Heterogeneous Face Recognition Approach Based on a Cycle Generative Adversarial Network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 4, pp. 132–145, 2022.

[56] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino, and J. Torresen, "A facial expression recognition system using robust face features from depth videos and deep learning," *Computers & Electrical Engineering*, vol. 63, pp. 114–125, 2017.

[57] A. Alcaide, M. A. Patricio, A. Berlanga, A. Arroyo, and J. J. Cuadrado-Gallego, "LIPSNN: A Light Intrusion-Proving Siamese Neural Network Model for Facial Verification," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 4, pp. 121–131, 2022.

Cristina Manresa-Yee

Cristina Manresa-Yee received her degree in Computer Science and her Ph. D. in Computer Science from the University of Balearic Islands. She is currently an Associate Professor at the University of the Balearic Islands. Her research interests include human-computer interaction, computer vision and explainable AI.

Silvia Ramis

Silvia Ramis, Ph. D. in Information and Communications Technologies from the UIB (since 2019). She has participated in several projects in the field of Computer Vision, Artificial Intelligence, Explainable Artificial Intelligence and Human-Robot Interaction. Her research experience focuses on artificial intelligence applied to human-robot interaction, especially in face detection and facial expression recognition.

Jose M. Buades

Jose Maria Buades Rubio received his degree in Computer Science and his Ph. D. in Computer Science from the University of Balearic Islands. He is currently an Associate Professor at the University of the Balearic Islands. His research interests include computer graphics, computer vision and artificial intelligence.