The challenges of emotion recognition when wearing face mask

Maria Francesca Roig-Maimó[†], Ramon Mas-Sansó[*,†], Miquel Mascaró-Oliver[†] and Esperança Amengual-Alcover[†]

[1]*University of the Balearic Islands, Carretera de Valldemossa km 7.5, Palma de Mallorca, Spain*

Abstract

The COVID-19 pandemic ushered in widespread mask mandates across many nations to curb virus transmission. Masks, shown to be cost-effective safeguards in healthcare settings, are poised to remain a societal norm. However, their use substantially obscures facial expressions, complicating emotion identification. Hence, assessing the impact of masks on facial expression recognition and emotional interpretation is imperative. Previous studies identified accuracy losses and increased confusion in emotion recognition with masks but didn't explore underlying causes. This study delves into these confusions by analyzing facial characteristics influencing each expression, using the LIME explainable AI technique. Starting with Faigin's facial feature definitions for expressions, we group similar expressions sharing visible features under masks, hypothesizing they're recognizably similar. We employ a CNN model on a masked facial expression dataset to test our hypothesis.

Keywords

emotion recognition, facial expression dataset, face mask, LIME, CNN

1. Introduction

The advent of the COVID-19 pandemic led to the widespread use of face masks to reduce virus transmission. While originally a public health measure, masks have proven to be a cost-effective mechanism for personal protection, particularly in healthcare settings, and are likely to remain common in many environments. Moreover, facial occlusion is not exclusive to pandemics: it also occurs in everyday situations such as the use of protective gear, outdoor activities, or cultural and religious practices. In all these cases, the lower part of the face—including the chin, mouth, and nose—is partially or fully hidden, making emotion recognition more challenging. Understanding how this occlusion affects emotional perception is therefore critical for the design of inclusive and affective interactive systems.

There is evidence in the literature that wearing a mask negatively impacts the ability to discern between different facial expressions as it causes the loss of a significant amount of facial information [1, 2, 3, 4, 5, 6, 7]. Consequently, it complicates the interpretation of emotional states. Pavlova and Sokolov [8], in a comprehensive review, call for further research to clarify the impact of face masks on emotion recognition.

Several studies have examined this issue in more depth. Carbon [3] examines the impact of face masks on the recognition of emotions and identifies instances of confusion between *Disgust* and *Anger*, as well as between neutral expressions and those of *Joy*, *Sadness*, and *Anger*. Other works explore the influence of face masks on children and adults [9], the role of different facial regions [10], the color of the mask [11], and cultural differences [12], with most of them reporting reduced recognition rates for *Joy*, *Disgust*, and *Sadness*. However, these studies primarily report performance outcomes without investigating the underlying reasons behind misclassifications.

In this study, we aim to go a step further by analyzing which visible facial features contribute to systematic confusion when a person wears a mask. We hypothesize that expressions which share the same visible features when masked are likely to be confused. Previous studies have shown that Convolutional Neural Networks (CNNs) can detect subtle facial features more effectively than human observers, often leading to more accurate classification of expressions [13, 14]. To test our hypothesis, we train a CNN model on a dataset of masked facial expressions and interpret the resulting classifications using an explainable artificial intelligence technique (XAI) called Local Interpretable Model-agnostic Explanation (LIME).

In order to analyze how wearing a face mask may affect the recognition of facial expressions, we begin by identifying the facial features that define each expression, based on the theoretical framework proposed by Faigin [15], who divides the face into three main zones: the upper part of the face (including forehead and eyebrows), the eyes, and the mouth. He defines similarity between facial expressions based on the configuration of facial elements within these zones. Two expressions are considered "similar" when they share configurations in one or more of these regions. Taking this definition as our baseline, we group facial expressions into sets of "similar facial expressions", where each set includes expressions that share the same visible facial features when a mask is worn (i.e., primarily those in the eye and upper face regions). Accordingly, we hypothesize that all expressions within the same set are likely to be recognized as the same facial expression, given the limited visibility caused by the mask.

From the perspective of Human-Computer Interaction (HCI), this research provides valuable insights into how facial occlusion interferes with affective computing. The findings are applicable to various interactive systems where emotional awareness is essential, such as virtual assistants, e-health, or socially assistive robots.

The paper is structured as follows: Firstly, we introduce the UIBVFED-Mask dataset, which will be employed extensively in our study. Subsequently, we formulate hypotheses regarding facial expression confusion when wearing face masks for each emotion. In Section 4, we provide an overview of our neural network model, data preprocessing procedures, and our prediction explanation methodology. Section 5 is dedicated to present our findings and engaging in discussions for each emotion category. Lastly, we summarize our conclusions.

2. The UIBVFED-Mask dataset

For the experiments, the UIBVFED-Mask dataset [16] is used. The UIBVFED-Mask dataset is an extension of the UIBVFED dataset [17]. UIBVFED is a database formed by virtual characters performing 32 facial expressions, classified based on the six universal emotions according to Gary Faigin (*Anger*, *Disgust*, *Fear*, *Joy*, *Sadness*, and *Surprise*) [15] plus the *Neutral* emotion. The UIBVFED-Mask dataset comprises the same images as the UIBVFED dataset but has been reconstructed to include face masks. Figure 1 shows an example of the *Neutral* expression in the original UIBVFED dataset and the corresponding image in the UIBVFED-Mask dataset, where we can clearly see the portion of the face hidden by the mask.

The images of the facial expressions that compose the dataset were generated following the guidelines of the Facial Action Coding System (FACS) [18]. So, the deformations applied to the 3D models have a direct correspondence with the Action Units (AUs) that are associated to each expression. This procedure ensures that images are labeled objectively. Moreover, the usage of synthetic datasets has actually proven to be a good replacement for real image datasets since they achieve recognition rates that are comparable to the real ones [19, 20].

2.1. Data description

The dataset is comprised of a total of 660 images, with 20 avatars representing 33 different facial expressions each. Table 1 displays the distribution of images per emotion.

Figure 1: Example of the *Neutral* facial expression: (a) in the UIBVFED dataset without mask, and (b) in the UIBVFED-Mask dataset including the face mask.

Table 1
Number of images per emotion of the UIBVFED-Mask dataset

	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
N° of images	80	60	80	280	20	120	20

2.2. Limitations

Despite all the advantages of synthetic datasets, the UIBVFED-Mask dataset has associated some limitations that may lead to some emotion's confusion:

- Geometry: Facial geometry of the characters can make certain values, for example those that differ in intensity, indistinguishable from each other. In addition, the pressure of the eyes is produced by a deformer that acts on the lower eyelid. The most visible effect of this are the characteristic wrinkles in the lower and exterior parts of the eyes. The UIBVFED-Mask models do not have enough geometric density to reflect these wrinkles in detail. For this reason, this feature can cause confusion.
- Texture: The UIBVFED-Mask characters have the eyebrows defined by texture, not geometrically. As in reality, depending on the shape of the eyebrow or the makeup of the character, a relaxed eyebrow may look like a furrowed brow. Therefore, certain characters can lead to confusion.
- Facial animation: Facial animation of the characters is based on blendshape deformations that affect the geometry of the skin and the lower denture. These deformations originate from the character configuration in the Autodesk Character Generator software application, and they are not inheritable by hierarchy. Therefore, the mask geometry is not affected by the animation. This might seem an inconvenience since the image will not have this information. In any case, in this study we do not consider image sequences where the movement of the mask deformation could be more helpful, but we only analyze facial expressions at their zenith.

3. Hypothesis: facial expression confusion when wearing face masks

In accordance with Faigin [15], facial expression recognition depends on the role of the face muscles. In his work, the author focuses on the action of muscles in three key areas of the face: (1) the forehead and brows; (2) the eyes; and (3) the mouth and chin. The same author empathizes that "an expression will only be clear and unambiguous when there is action in both the eyes/brow and the mouth at once".

Wearing a face mask occludes the area of the mouth and chin, that is, one of the three key areas. Therefore, it is almost impossible, according to the theory, to recognize a facial expression in a clear and unambiguous way. So, confusion between facial expressions is to be expected when wearing a face mask, that is to say, when the area of the mouth and chin is occluded.

In order to evaluate how wearing a face mask may affect the recognition of a facial expression, in this section the facial expressions described by Faigin are analysed. This analysis focuses on the features of the upper part of the face -i.e., the key areas of the face that remain visible when wearing a face mask: (1) the forehead and brows, and (2) the eyes.

While Faigin describes in detail the muscle movements and visible features associated with each facial expression, he does not address the impact of partial facial occlusion –such as the use of face masks– on expression recognition. The hypothesis proposed in this work applies Faigin's definitions to the masked context, aiming to predict which expressions may become visually indistinguishable when the lower face is hidden.

In the following sections, for each of the six universal emotions plus the *Neutral* emotion, we present a table summarizing, for each of its associated facial expressions according to Faigin [15] (first column), the facial features of the two key areas of the upper part of the face (second column). The facial expressions associated to the same emotion that share all their visible facial features are shown in the same row, meaning that they are plausible to be confused. Besides, for each row, we present (in the third column) the number of similar facial expressions (we define "similar" in terms of sharing the same facial features) associated to other emotions. Similarities between different emotions are grouped by colour. So, facial expressions of the same colour could be associated (correctly or incorrectly) to the same emotion.

3.1. Neutral emotion

Table 2 summarizes the facial features of the facial expression associated to the *Neutral* emotion that are visible when wearing a face mask (see second column). The third column relates, for the *Neutral* facial expression, the emotions that comprise facial expressions that share the same facial features on the upper-side of the face and, consequently, could be misled with that facial expression when wearing a face mask. For instance, the *Neutral* facial expression is shown to have the same upper-face features (relaxed eyebrows and eyes opened without pressure) as one facial expression from the *Disgust* emotion and two facial expressions from the *Joy* emotion. These facial expressions, although emotionally distinct, may appear similar when the lower part of the face is masked, leading to potential confusion. Therefore, the *Neutral* facial expression could be confused with one of the three facial expressions associated with the *Disgust* emotion (see the *Disgust* facial expression in Table 4), and with two of the fourteen facial expressions associated with the *Joy* emotion (see *False Laughter 1* and *False Smile* facial expressions in Table 6).

Table 2
Expected emotion confusion for the facial expressions associated to the *Neutral* emotion

Facial Expression	Facial Features	Similar Emotion
Neutral	Eyebrows: relaxed Eyes: opened without pressure	Disgust 1/3 (see Table 4) Joy 2/14 (see Table 6)
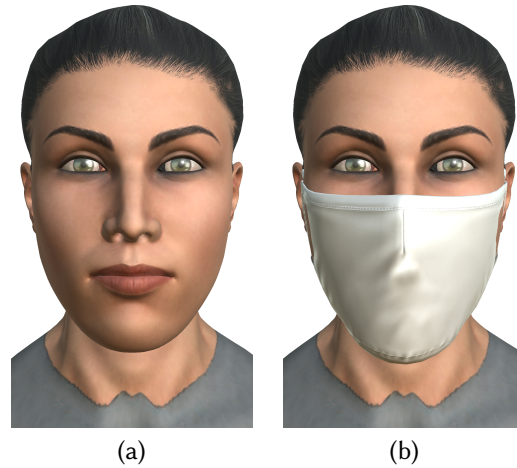
Figure 2: Example of images of the *Neutral* facial expression: (a) without wearing the face mask, and (b) when wearing the face mask.

Figure 2 shows an avatar (Isabel avatar) performing the *Neutral* facial expression without (see Figure 2a) and with a face mask (see Figure 2b). In this figure, it can be observed that the facial features corresponding to the *Neutral* facial expression that are visible when wearing the face mask are relaxed eyebrows and the eyes opened without pressure. In this scenario, the information of the facial feature of the relaxed mouth is occluded with the face mask and, therefore, this information is lost.
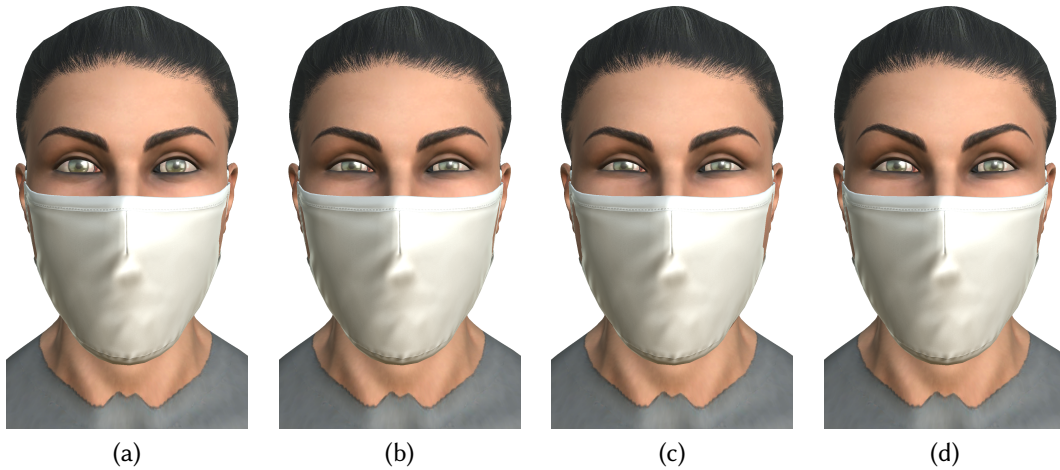
Figure 3: Example of images of the same avatar wearing the face mask and performing (a) the *Neutral* facial expression (*Neutral* emotion) and their similar facial expressions: (b) *Disgust* facial expression (*Disgust* emotion), (c) *False Laughter 1* facial expression (*Joy* emotion), and (d) *False Smile* facial expression (*Joy* emotion).

Figure 3 shows the Isabel avatar wearing the face mask and performing the *Neutral* facial expression (see Figure 3a) and their similar facial expressions according to Table 2: *Disgust* facial expression (see Figure 3b), *False Laughter 1* facial expression (see Figure 3c), and *False Smile* facial expression (see Figure 3d). All the images of Figure 3 present relaxed eyebrows and the eyes opened without pressure; the only subtle difference lies in how widely the eyes are opened. Observing theses images, it is almost impossible to distinguish which of them correspond to the neutral, disgust or joy emotion; and, therefore, they could be associated (correctly or incorrectly) with any of these emotions.
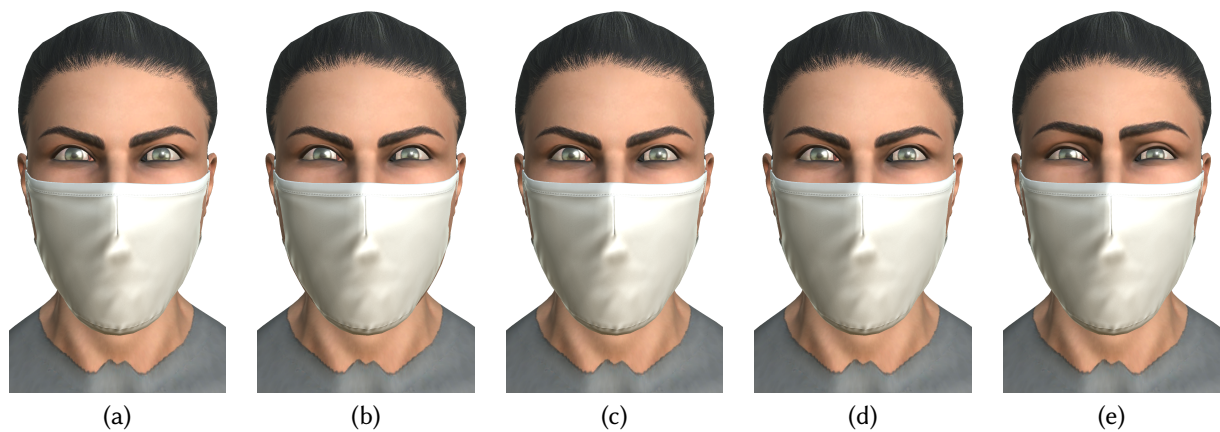
3.2. Anger emotion

Table 3 summarizes the facial features of the expressions associated to the *Anger* emotion that are visible when wearing a face mask (see second column). In the third column, it can be observed that all

the facial expressions associated to the *Anger* emotion share their visible facial features with one of the facial expressions of the *Joy* emotion: *Ingratiating Smile* (see Table 6).

Table 3
Expected emotion confusion for the facial expressions associated to the *Anger* emotion

Facial Expression	Facial Features	Similar Emotion
Enraged Compressed Lips	Eyebrows: furrowed	Joy 1/14 (see Table 6)
Enraged Shouting	Eyes: opened wide	
Mad		
Sternness Anger		

Figure 4: Example of images of the same avatar wearing the face mask and performing the (a) *Enraged Compressed Lips* facial expression (*Anger* emotion), the (b) *Enraged Shouting* facial expression (*Anger* emotion), (c) the *Mad* facial expression (*Anger* emotion), the (d) *Sternness Anger* facial expression (*Anger* emotion), and the (e) *Ingratiating Smile* facial expression (*Joy* emotion).

Figure 4 shows the Isabel avatar wearing the face mask and performing the *Enraged Compressed Lips*, *Enraged Shouting*, *Mad* and *Sternness Anger* facial expressions and its similar facial expression according to Table 6: *Ingratiating Smile* facial expression. All the images of Figure 4 present furrowed eyebrows and the eyes wide opened, making them practically indistinguishable.
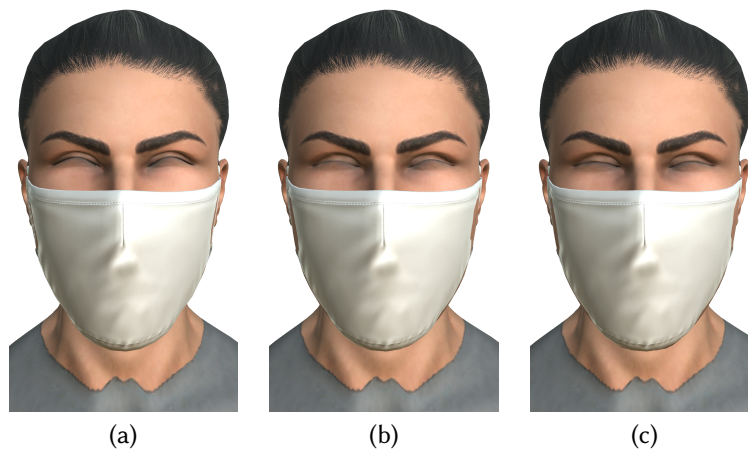
3.3. Disgust emotion

The expected confusions for the *Disgust* emotion are summarized in Table 4. Besides the *Disgust* facial expression already analyzed in the Section 3.1, we can observe in Table 7 that the *Physical Repulsion* facial expression could be confused with two of the six facial expressions associated with the *Sadness* emotion: *Crying Closed Mouth* and *Crying Open Mouthed* facial expressions.

Figure 5 shows the Isabel avatar wearing the face mask and performing the *Physical Repulsion* facial expression and its similar facial expressions of *Sadness* emotion (according to Table 4): *Crying Closed Mouth* and *Crying Open Mouthed* facial expressions. All the images of Figure 5 present furrowed eyebrows and the eyes closed with pressure, which make them indistinguishable.

Table 4

Expected emotion confusion for the facial expressions associated to the *Disgust* emotion

Facial Expression	Facial Features	Similar Emotion
Disdain	Eyebrows: relaxed Eyes: closed	
Disgust	Eyebrows: relaxed Eyes: opened without pressure	Neutral 1/1 (see Table 2) Joy 2/14 (see Table 6)
Physical Repulsion	Eyebrows: furrowed Eyes: closed with pressure	Sadness 2/6 (see Table 7)

(a)   (b)   (c)

Figure 5: Example of images of the same avatar wearing the face mask and performing the (a) *Physical Repulsion* facial expression (*Disgust* emotion), the (b) *Crying Closed Mouth* facial expression (*Sadness* emotion), and the (c) *Crying Open Mouthed* facial expression (*Sadness* emotion).
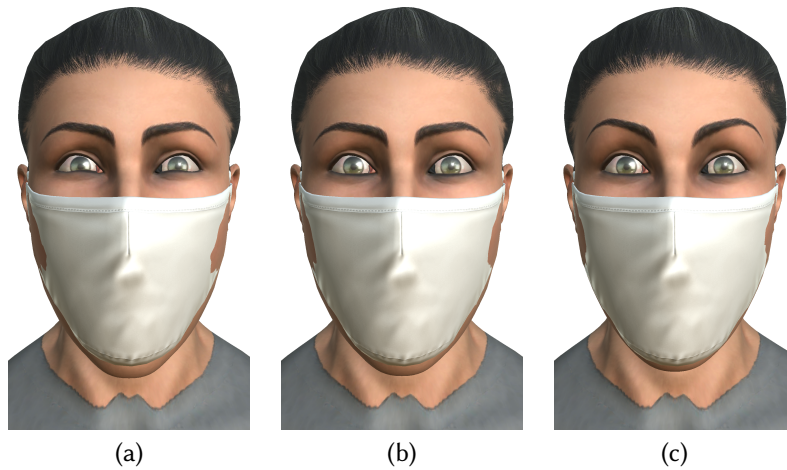
3.4. Fear emotion

In Table 5 we can see the expected confusions associated to the *Fear* emotion. Two of the facial expressions associated to the *Fear* emotion (*Terror* and *Very Frightened* facial expressions) share their visible facial features with the only facial expression of the *Surprise* emotion: *Surprise* (see Table 8).

Table 5

Expected emotion confusion for the facial expressions associated to the *Fear* emotion

Facial Expression	Facial Features	Similar Emotion
Afraid	Eyebrows: slightly lifted straight up Eyes: opened wide	
Terror Very Frightened	Eyebrows: lifted straight up Eyes: opened extra wide	Surprise 1/1 (see Table 8)
Worried	Eyebrows: lifted straight up Eyes: opened wide	

Figure 6 shows the Isabel avatar wearing the face mask and performing the *Terror* and *Very Frightened* facial expressions and its similar facial expression according to Table 5: *Surprise* facial expression. All the images of Figure 6 present lifted straight up eyebrows and extra wide opened eyes. A subtle difference can be observed within the two emotions: in the facial expression of the *Surprise* emotion the eyebrows appear more raised than in the facial expressions of the *Fear* emotion.

Figure 6: Example of images of the same avatar wearing the face mask and performing the (a) *Terror* facial expression (*Fear* emotion), the (b) *Very Frightened* facial expression (*Fear* emotion), and the (c) *Surprise* facial expression (*Surprise* emotion).

3.5. Joy emotion

In previous sections, we have already analyzed the facial expressions that could be confused with the facial expressions associated with the *Joy* emotion (see Table 6): the *False Laughter 1* and the *False Smile* facial expressions have already been analyzed in the Section 3.1, and the *Ingratiating Smile* facial expression has already been analyzed in the Section 3.2.

Table 6

Expected emotion confusion for the facial expressions associated to the *Joy* emotion

Facial Expression	Facial Features	Similar Emotion
False Laughter 1 False Smile	Eyebrows: relaxed Eyes: opened without pressure	Neutral 1/1 (see Table 2) Disgust 1/3 (see Table 4)
Smiling Closed Mouth Smiling Open Mouthed Stifled Smile	Eyebrows: relaxed Eyes: opened with pressure	
Laughter Uproarious Laughter	Eyebrows: relaxed Eyes: closed with pressure	
False Laughter 2 Abashed Smile	Eyebrows: relaxed Eyes: closed without pressure	
Eager Smile	Eyebrows: furrowed Eyes: opened extra wide	
Ingratiating Smile	Eyebrows: furrowed Eyes: opened wide	Anger 4/4 (see Table 3)
Sly Smile	Eyebrows: furrowed Eyes: opened with pressure	
Melancholy Smile	Eyebrows: slightly lifted straight up Eyes: opened without pressure	
Debauched Smile	Eyebrows: lifted asymmetrically Eyes: pupil partially covered	

3.6. Sadness emotion

Table 7 shows the facial features of the facial expressions associated to the emotion of *Sadness*. All the facial expressions that could be confused with the facial expressions associated to other emotions have already been analyzed in Section 3.1: the *Crying Closed Mouth* and the *Crying Open Mouthed* facial expressions.

Table 7
Expected emotion confusion for the facial expressions associated to the *Sadness* emotion

Facial Expression	Facial Features	Similar Emotion
Crying Closed Mouth Crying Open Mouthed	Eyebrows: furrowed Eyes: closed with pressure	Disgust 1/3 (see Table 4)
Miserable	Eyebrows: slightly lifted straight up Eyes: opened with pressure	
Nearly crying Sad Suppressed sadness	Eyebrows: lifted straight up Eyes: opened with pressure	

3.7. Surprise emotion

The expected confusions for the *Surprise* emotion can be seen in Table 8. This facial expression could be confused with facial expressions associated to the *Fear* emotion: *Terror* and *Very Frightened* facial expressions (see Table 5).

Table 8
Expected emotion confusion for the facial expressions associated to the *Surprise* emotion

Facial Expression	Facial Features	Similar Emotion
Surprise	Eyebrows: lifted straight up Eyes: opened extra wide	Fear 2/4 (see Table 5)

3.8. Summary

In the tables of the previous sections, similarities between different emotions have been grouped by color, while similarities within the same emotion have been grouped in the same row. According to these criteria, all colorless expressions should be perfectly distinguishable. Table 9 summarizes the theoretical expected confusion between emotions.

Table 9
Summary of the theoretical emotion confusion when wearing face mask

|	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Anger	X	-	-	X	-	-	-
Disgust	-	X	-	X	X	X	-
Fear	-	-	X	-	-	-	X
Joy	X	X	-	X	X	-	-
Neutral	-	X	-	X	X	-	-
Sadness	-	X	-	-	-	X	-
Surprise	-	-	X	-	-	-	X

4. Methods

In this work we use a Convolutional Neural Network (CNN) to analyse the effects of wearing a face mask in the automatic recognition of facial expressions.

The main contribution of this work is not the CNN model, rather to test the factors that wearing a face mask introduces in both training and recognition of human emotions on facial expression images. Therefore, we don't pretend to define a very precise neural network to recognize facial expressions, as this target has already been successfully addressed in the literature [21], but to use a simple CNN model as a baseline to analyze our hypothesis.

This section contains a detailed description of how we pre-process the data, the procedure we follow to train and test the CNN model and, finally, how we explain the predictions obtained by the model using the Local Interpretable Model-agnostic Explanation (LIME) technique.

4.1. The convolutional neural network

We use a Convolutional Neural Network (CNN) following the scheme shown in Figure 7. CNNs are particularly well-suited for facial expression recognition due to their ability to automatically learn spatial hierarchies of features from image data.
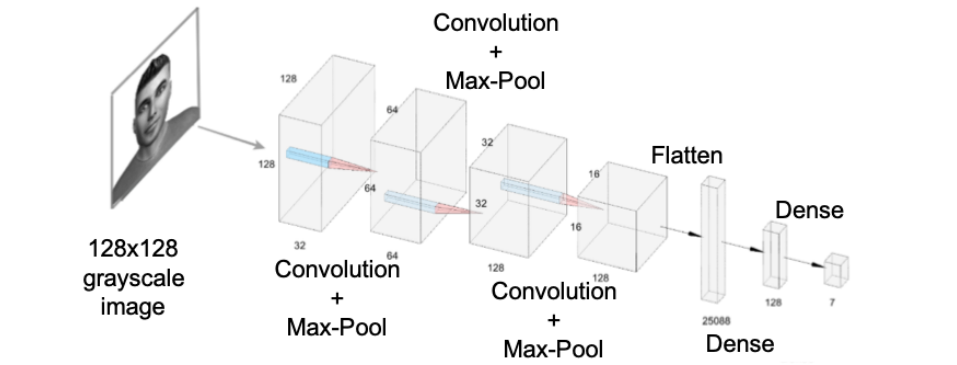
Figure 7: The Convolutional Neural Network (CNN). From a 128x128 grayscale image, the CNN provides the target class describing the identified emotion.

As input, we have a grayscale image with a resolution of 128x128 pixels. The characteristics of the input image are extracted using a three-level combination of convolution and max-pooling. A Rectified Linear Unit (ReLU) activation function is also applied to activate only the nodes that serve our purpose. The two final layers of the CNN are fully connected layers. In the first dense layer, we flatten the last output tensor from the convolutional base and, in the final dense layer, we obtain the seven outputs corresponding to the classification classes of the emotions (*Anger, Disgust, Fear, Joy, Neutral, Sadness, Surprise*).

4.2. Data pre-processing

To pre-process each image, we began by cropping the face to reduce the impact of the background on the result. Then, we converted the image to grayscale and we adjusted its resolution to fit the required dimensions of the data input to our CNN (128x128 pixels).

4.3. Procedure

After completing the pre-processing step for all the images in the UIBVFED-Mask dataset, we prepared the training and testing datasets.

For the training dataset, we collected a 80% of the data, and we took the remaining 20% for the testing dataset. Both datasets contained a class distribution that was representative of the complete UIBVFED-Mask dataset (see Table 10).

Table 10
Class distribution of the complete UIBVFED-Mask, training and testing datasets

Dataset	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise	Total
UIBVFED-Mask	80	60	80	280	20	120	20	660
Training	64	48	64	224	16	96	16	528
Testing	16	12	16	56	4	24	4	132
Proportion	0.12	0.09	0.12	0.42	0.03	0.18	0.03	1

We trained the CNN model previously described with the training dataset. Then, the model was tested with the testing dataset and the evaluation metrics in terms of global accuracy and confusion matrix were computed.

As a final step, and to try to obtain an explanation of the model's outcome, we apply the LIME technique over the predictions (see Section 4.4).

4.4. The predictions' explanations: the LIME technique

We visually analyze the predictions obtained by the model using the Local Interpretable Model-agnostic Explanation (LIME) [22] technique. LIME is a widely used method to obtain local explanations of black-box models [23, 24] because it can be applied to any machine learning model due to its agnostic nature.

When explaining the classification on images, LIME depicts the main parts of the image input that contribute to the prediction using a simple approach: it perturbs the inputs of the model and observes how the new predictions behave, then it learns how the model works using a linear model through the weighting of the perturbations. The obtained explanation is not globally valid but it is locally accurate around the perturbed inputs.

What LIME highlights in their explanations are the superpixels, or collections of pixels that cover a connected area of the image, which best justify the selection of a given class. The superpixels should correspond to specific patterns of the image but normally the user can only specify the resolution of the considered areas. This poses an additional difficulty as significant features may lie within different superpixels. However, we can exert some control if we know the relative size of the affected areas, allowing us to adjust the number of superpixels and, consequently, their size (see Figure 8). In this study we set the number of superpixels to 50.

N° superpixels = 25   N° superpixels = 50
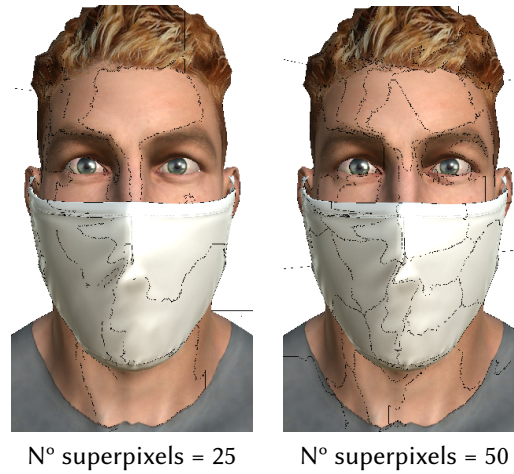
Figure 8: Example of the effect of the configuration of the number of superpixels to segment into on a facial expression image of the UIBVFED-Mask dataset.

5. Results and discussion

The Convolutional Neural Network model trained with face masks had an overall accuracy of 0.65. As can be seen in the confusion matrix depicted in Figure 9, we obtained perfect results for *Anger* (100%), very good results for *Joy* (89.3%) and good results for *Fear* (75%). Notable are the cases of *Disgust* and *Neutral* emotion, in which none of the images were correctly classified.

It is also interesting the information of the most misled emotions: *Neutral*, *Disgust* and *Sadness* are confused with *Joy* in the 100%, 91.7% and 70.8% of the cases, respectively. This confusion agrees with the theoretical emotion confusion summarized in Table 9, as some of the facial expressions associated to the emotions *Neutral*, *Disgust*, *Joy* and *Sadness* share the same visible facial features. Also the misled of 50% between the emotions *Fear* and *Surprise* is inline with the theoretical expected confusion.

Below there is a detailed discussion of the predictions obtained by emotion.
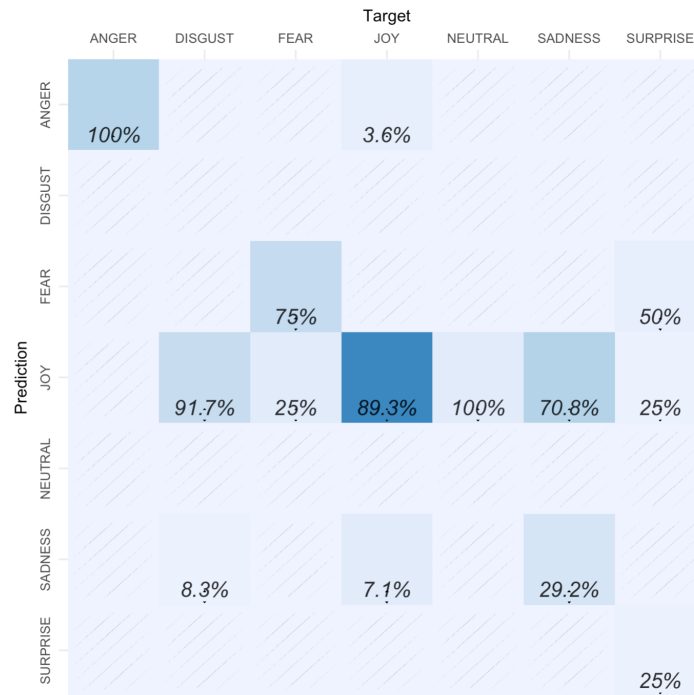
Figure 9: Confusion matrix of the emotions' classification performed by the CNN model.

5.1. Neutral emotion

As already commented, all the images of the *Neutral* emotion were labeled as *Joy* emotion.

Theoretical confusion of emotions related to the *Neutral* expression, as shown in Table 2, suggests that the *Neutral* facial expression (associated with the *Neutral* emotion) may be confused with the facial expressions *False Laughter 1* and *False Smile*, which are associated with the *Joy* emotion, as well as with one facial expression linked to the *Disgust* emotion. Consequently, the behavior of the CNN model aligns with the hypothesis we have formulated in Section 3. This observation is also consistent with the findings presented in Carbon [3].
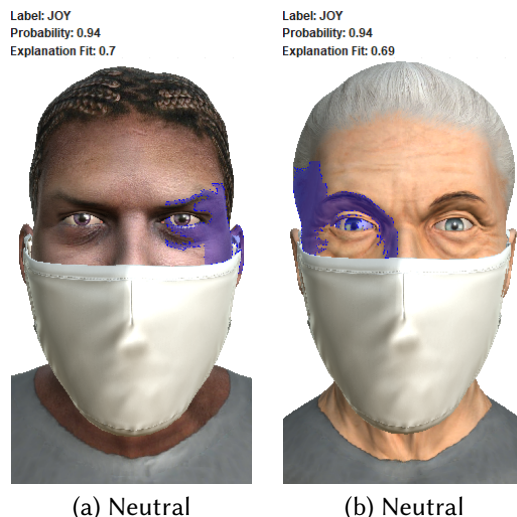
(a) Neutral   (b) Neutral

Figure 10: Results of applying LIME to two images of *Neutral* emotion labeled as the emotion of *Joy*.

In Figure 10 we can observe the results of applying LIME to the predictions of two images of *Neutral* emotion labeled as *Joy* emotion. The areas highlighted in blue correspond to the regions on which the model relies to make predictions for these expressions. In both cases, the superpixels highlighted of the images coincide with the area of the eyes and they depict the relaxed eyebrows and opened eyes without pressure.

5.2. Anger emotion

Our trained CNN model correctly classifies all *Anger* facial expressions as *Anger* emotion (100%). But, as anticipated in Kastendieck et al. [9], a subset of *Joy* facial expressions exhibits confusion with the *Anger* emotion (3.6% in our confusion matrix). Our hypothesis attributes this confusion specifically to the *Ingratiating Smile*, one of the fourteen possible *Joy* facial expressions.
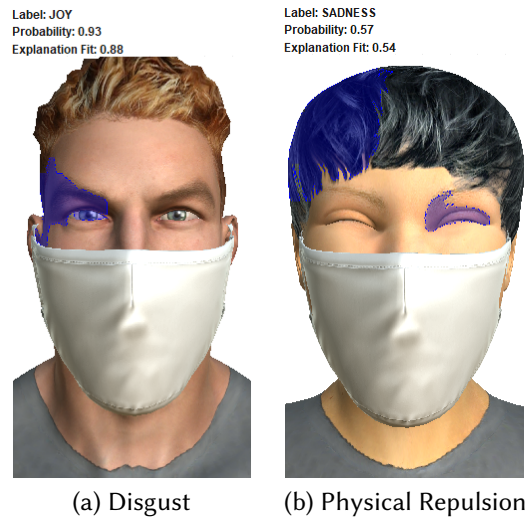
Figure 11 shows the results of applying LIME to the predictions of two images of *Joy* emotion labeled as *Anger* emotion. The facial expressions involved are *Eager Smile* and *Ingratiating Smile*, these expressions exhibit the furrowing of the eyebrows and the wide aperture of the eyes.

(a) Eager Smile (Joy)   (b) Ingratiating Smile (Joy)

Figure 11: Results of applying LIME to the two images of *Joy* emotion labeled as the emotion of *Anger*: (a) *Eager Smile* facial expression and (b) *Ingratiating Smile* facial expression.

5.3. Disgust emotion

If we analyze the predictions obtained for the *Disgust* emotion, we realize that it is completely confused with the emotions of *Joy* and *Sadness*. Such results are coherent, as the *Disgust* facial expression shares all their visible facial features with the *False Laughter 1* and *False Smile* facial expressions of *Joy*, and the *Physical Repulsion* expression shares all its visible facial features with the facial expressions *Crying Closed Mouth* and *Crying Open Mouthed* of the emotion of *Sadness* (see Table 4). Figure 12 underlines the areas of the eyebrows (relaxed for *Disgust* and the commented facial expressions of *Joy*; furrowed for *Physical Repulsion* and its similar facial expressions of *Sadness*) and eyes (opened without pressure for *Disgust* and the commented facial expressions of *Joy*; closed with pressure for *Physical Repulsion* and its similar facial expressions of *Sadness*) of the predictions.

(a) Disgust   (b) Physical Repulsion

Figure 12: Results of applying LIME to images of *Disgust* emotion labeled as the emotion of (a) *Joy*: *Disgust* facial expression and the emotion of (b) *Sadness*: *Physical Repulsion* facial expression.

5.4. Fear emotion

Concerning the *Fear* emotion, the results of the prediction are correct in a 75% of the cases as can be seen in the confusion matrix (see Figure 9). However, in a 25% of the cases, the emotion of *Fear* is labeled as *Joy*. This could be due to the wide opening of the eyes, as this facial feature is also present in the expressions of *Eager* and *Ingratiating Smile* corresponding to the emotion of *Joy*. But, probably, the incorrect classification is due to one of the limitations stated in Section 2.2 about the texture limitation where a relaxed eyebrow can look like furrowed. In the case of Figure 13c, it seems that the model interprets the shadow of the eyebrow as if it were the eyebrow itself. The images in the Figures 13a and 13b, corresponding to the explanation provided by LIME, show that the areas of interest are precisely located on the eyebrows.
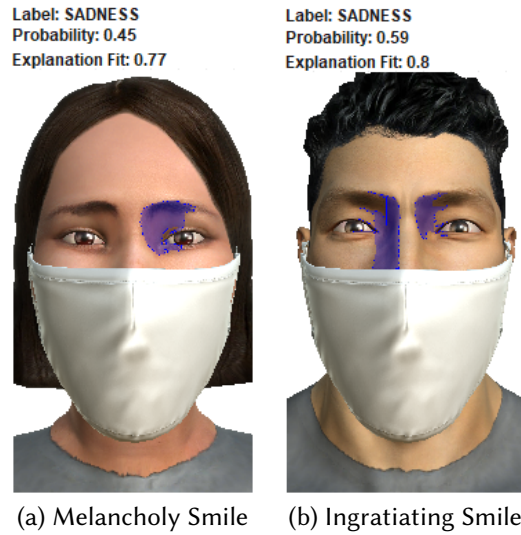

(a) Afraid   (b) Terror   (c) Worried
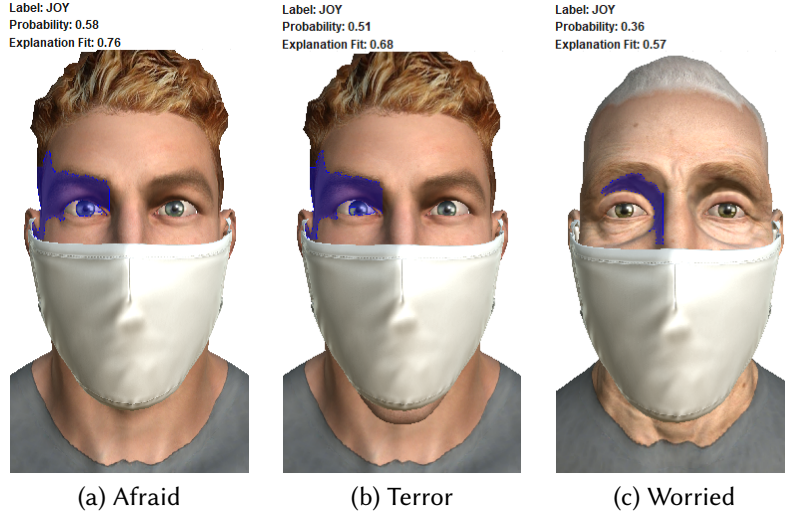
Figure 13: Results of applying LIME to images of *Fear* emotion labeled as the emotion of *Joy*: (a) *Afraid* facial expression, (b) *Terror* facial expression and (c) *Worried* facial expression.

5.5. Joy emotion

The emotion of *Joy* has an overall recognition rate of 89.3%, which is quite acceptable and is barely confused with *Anger* (3.6%) (already commented in Section 5.2) and *Sadness* (7.1%).
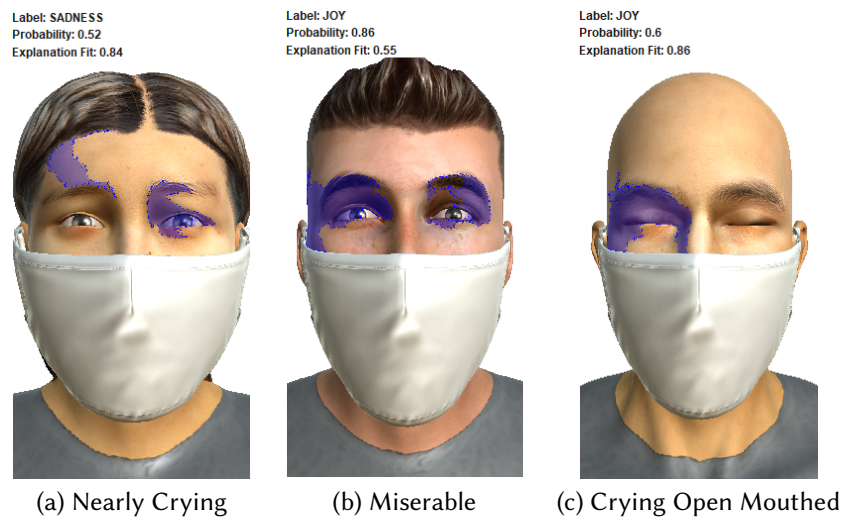
Our hypothesis formulated in Section 3 does not expect the confusion between facial expressions associated to the emotions of *Joy* and *Sadness*. Figure 14 gives some insight on the superpixels that most influence the confusion of the facials expressions *Melancholy Smile* and *Ingratiating Smile* with *Sadness* emotion. In the case of the image of the *Melancholy Smile* expression, the highlighted area corresponds to the eyebrow slightly lifted straight up and the inner part of the eye, which appears opened without pressure. The highlighted feature of the eyebrow (slightly lifted straight up) is a facial feature present in four of the six *Sadness'* facial expressions, and the only difference resides in the eye opening pressure. This no-detection of a difference in the eye opening pressure could be explained by the inherent limitations present in the geometry of our dataset (see Section 2.2). The case of the image of the *Ingratiating Smile* expression could be explained, in addition to the difference in the eye opening pressure, by the texture limitation that may cause that a furrowed eyebrow looks like an slightly lifted straight up eyebrow.

Label: SADNESS
Probability: 0.45
Explanation Fit: 0.77

Label: SADNESS
Probability: 0.59
Explanation Fit: 0.8

(a) Melancholy Smile (b) Ingratiating Smile

Figure 14: Results of applying LIME to two images of *Joy* emotion labeled as the emotion of *Sadness*: (a) *Melancholy Smile* facial expression and (b) *Ingratiating Smile* facial expression.

5.6. Sadness emotion

The emotion of *Sadness* has an overall recognition rate of 29.2%, that means that this emotion is difficult to recognize when wearing a mask. In our experiment, this emotion is highly confused with the emotion of *Joy* (70.8%). As previously commented in Section 5.5, this possibility is considered as plausible because of the limitations of the geometric density of the characters of the dataset and the definition of the eyebrows as texture (see Section 2.2). These limitations could lead to a difficult distinction in the furrowing of the eyebrows and the pressure of the eyes opening. LIME explanations depicted in Figure 15b and Figure 15c seem to confirm this extend as these areas are highlighted as the superpixels of interest. Figure 15a depicts a facial expression of *Sadness* correctly labelled, where it can be observed that the highlighted areas are inline with the highlighted areas in Figure 15b and Figure 15c.

Label: SADNESS
Probability: 0.52
Explanation Fit: 0.84

Label: JOY
Probability: 0.86
Explanation Fit: 0.55

Label: JOY
Probability: 0.6
Explanation Fit: 0.86

(a) Nearly Crying (b) Miserable (c) Crying Open Mouthed

Figure 15: Results of applying LIME to images of *Sadness* emotion correctly labeled as the emotion of *Sadness*: (a) *Nearly Crying* facial expression; and incorrectly labeled as the emotion of *Joy*: (b) *Miserable* facial expression and (c) *Crying Open Mouthed* facial expression.

5.7. Surprise emotion

The *Surprise* emotion, as predicted, is mostly confused with the *Fear* emotion (50%). There is also a non-expected confusion with the *Joy* emotion (25%). This confusion can be explained in the same terms as stated in section 5.4 and in consonance with the LIME explanation (see Figure 16).

(a) Surprise   (b) Surprise
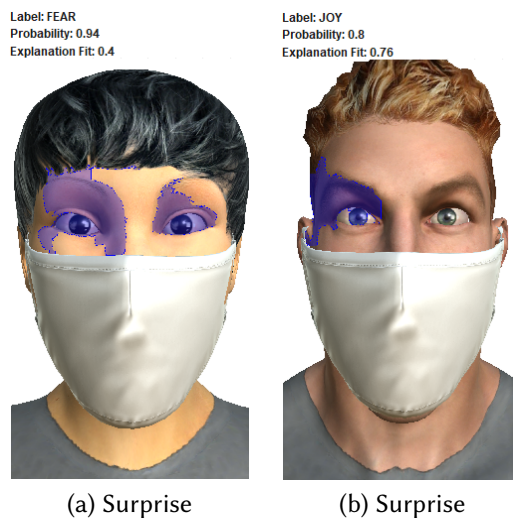
Figure 16: Results of applying LIME to images of *Surprise* emotion labeled as the emotion of (a) *Fear* and (b) *Joy*.

5.8. Summary

Based on the confusion patterns observed across all emotions, we can confirm that the hypothesis proposed at the beginning of this study (see Section 3) holds in most cases. Specifically, facial expressions that share the same visible features in the upper part of the face tend to be misclassified as one another when the lower part is masked. This expected confusion was particularly evident among expressions associated with *Joy, Disgust*, and *Neutral*, which frequently overlapped with others from different emotional categories. However, a few unexpected results also emerged (see Table 11). These outliers may be influenced by subtle geometric and textural limitations in the avatar representation. Overall, the alignment between the predicted and observed confusion patterns reinforces the validity of our hypothesis, while the divergences point to directions for future investigation in masked facial expression recognition.

Table 11
Expected and unexpected confusions according to the proposed hypothesis (see Section 3)

Original expression	Expected confusion	Unexpected confusion	Comment
Neutral	Joy (100%)		Aligns with the hypothesis
Anger	Joy (3.6%)		Aligns with the hypothesis
Disgust	Joy (91.7%), Sadness (8.3%)		Aligns with the hypothesis
Fear		Joy (25%)	Probably due to texture limitations
Joy	Anger (3.6%)	Sadness (7.1%)	Probably due to geometry
Sadness		Joy (70.8%)	Probably due to geometry
Surprise	Fear (50%)	Joy (25%)	Probably due to texture limitations

6. Conclusions

With the ongoing COVID-19 pandemic and the widespread use of face masks for protection, the difficulty of understanding people's emotions has become a prominent issue. Empirical evidence and previous studies have unequivocally demonstrated that wearing masks can impair the recognition of facial expressions. The loss of crucial facial information, particularly the concealment of the mouth and chin regions, poses a substantial obstacle to accurately discerning emotions. This loss of information can lead to confusion when trying to recognize an emotion. Such confusions are particularly evident in cases where facial expressions of different emotions share visible features in the upper part of the face, such as the eyes and eyebrows.

In this work we have used the theoretical framework proposed by Faigin, where he specifies the facial features that describe each facial expression, to give some insight in the predictable confusions between emotions when the lower part of the face is occluded. We have used explainable artificial intelligence techniques to verify that the areas considered for the prediction of an emotion correspond to the facial features that can cause a confusion. We have not only confirmed that some emotions, such as *Anger* and *Joy*, are reliably recognized, while others, like *Disgust*, *Sadness* and *Neutral* emotions, consistently lead to confusion but we have also determined the main factors leading to confusion by analyzing the facial characteristics that influence each of the expressions.

These findings are particularly relevant for the design of interactive systems intended for use in contexts where mask-wearing is prevalent, such as healthcare or public service environments. Incorporating knowledge about likely emotion confusions can inform more robust emotion-aware interfaces, improve adaptive responses from interactive agents, and support the development of compensatory mechanisms (e.g., multimodal emotion detection or user feedback loops) to enhance user experience and communication effectiveness.

Acknowledgments

Declaration on generative AI

The authors have not employed any Generative AI tools.

References

[1] B. Yang, J. Wu, G. Hattori, Facial expression recognition with the advent of face masks, in: 19th International Conference on Mobile and Ubiquitous Multimedia, MUM '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 335–337. doi:10.1145/3428361.3432075.

[2] E. Freud, A. Stajduhar, R. S. Rosenbaum, G. Avidan, T. Ganel, The COVID-19 pandemic masks the way people perceive faces, Scientific reports 10 (2020) 1–8. doi:10.1109/ICAACI50733.2020.00021.

[3] C.-C. Carbon, Wearing face masks strongly confuses counterparts in reading emotions, Frontiers in psychology 11 (2020) 566886.

[4] P. Barros, A. Sciutti, I only have eyes for you: the impact of masks on convolutional-based facial expression recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Nashville, TN, USA, 2021, pp. 1226–1231.

[5] R. Golwalkar, N. Mehendale, Masked-face recognition using deep metric learning and FaceMaskNet-21, Applied Intelligence (2022) 1–12. doi:10.1007/s10489-021-03150-3.

[6] F. Grundmann, K. Epstude, S. Scheibe, Face masks reduce emotion-recognition accuracy and perceived closeness, Plos one 16 (2021) e0249792. doi:10.1371/journal.pone.0249792.

[7] M. Marini, A. Ansani, F. Paglieri, F. Caruana, M. Viola, The impact of facemasks on emotion recognition, trust attribution and re-identification, Scientific Reports 11 (2021) 1–14. doi:10.1038/s41598-021-84806-5.

[8] M. A. Pavlova, A. A. Sokolov, Reading covered faces, Cerebral Cortex 32 (2021) 249–265. URL: https://doi.org/10.1093/cercor/bhab311. doi:10.1093/cercor/bhab311.

[9] T. Kastendieck, N. Dippel, J. Asbrand, U. Hess, Influence of child and adult faces with face masks on emotion perception and facial mimicry, Scientific Reports 13 (2023) 14848. doi:10.1038/s41598-023-40007-w.

[10] M. Ventura, A. Palmisano, F. Innamorato, G. . Tedesco, V. Manippa, A. Caffò, D. Rivolta, Face memory and facial expression recognition are both affected by wearing disposable surgical face masks, Cognitive Processing 24 (2023) 43–57. doi:10.1007/s10339-022-01112-2.

[11] S. Gil, L. Le Bigot, Emotional face recognition when a colored mask is worn: a cross-sectional study, Scientific Reports 13 (2023) 174. doi:10.1038/s41598-022-27049-2.

[12] T. Saito, K. Motoki, Y. Takano, Cultural differences in recognizing emotions of masked faces, Emotion 23 (2023) 1648.

[13] G. Carreto Picón, M. F. Roig-Maimó, M. Mascaró Oliver, E. Amengual Alcover, R. Mas-Sansó, Do machines better understand synthetic facial expressions than people?, in: Proceedings of the XXII International Conference on Human Computer Interaction, Interacción '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 1–5. doi:10.1145/3549865.3549908.

[14] M. F. Roig-Maimó, M. Mascaró Oliver, E. Amengual Alcover, R. Mas-Sansó, Sobre el reconocimiento de emociones y la precisión de los clasificadores, Revista de la Asociación Interacción Persona Ordenador (AIPO) 3 (2022) 55–66.

[15] G. Faigin, The artist's complete guide to facial expression, Watson-Guptill, New York, 2012.

[16] M. Mascaró-Oliver, R. Mas-Sansó, E. Amengual-Alcover, M. F. Roig-Maimó, UIBVFED-Mask: a dataset for comparing facial expressions with and without face masks, Data 8 (2023). doi:10.3390/data8010017.

[17] M. Mascaró Oliver, E. Amengual Alcover, UIBVFED: virtual facial expression dataset, PLOS ONE 15 (2020) 1–10. doi:10.1371/journal.pone.0231266.

[18] P. Ekman, W. V. Friesen, Facial Action Coding System, Environmental Psychology & Nonverbal Behavior (1978).

[19] L. Colbois, T. d. Freitas Pereira, S. Marcel, On the use of automatically generated synthetic image datasets for benchmarking face recognition, in: 2021 IEEE International Joint Conference on Biometrics (IJCB), IEEE, Shenzhen, China, 2021, pp. 1–8. doi:10.1109/IJCB52358.2021.9484363.

[20] J. del Aguila, L. M. González-Gualda, M. A. Játiva, P. Fernández-Sotos, A. Fernández-Caballero, A. S. García, How interpersonal distance between avatar and human influences facial affect recognition in immersive virtual reality, Frontiers in Psychology 12 (2021). doi:10.3389/fpsyg.2021.675515.

[21] M. Marini, A. Ansani, F. Paglieri, F. Caruana, M. Viola, The impact of facemasks on emotion recognition, trust attribution and re-identification, Scientific Reports 11 (2021) 5577. doi:10.1038/s41598-021-84806-5.

[22] M. T. Ribeiro, S. Singh, 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, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. doi:10.1145/2939672.2939778.

[23] P. R. Magesh, R. D. Myloth, R. J. Tom, An explainable machine learning model for early detection of parkinson's disease using LIME on DaTSCAN imagery, Computers in Biology and Medicine 126 (2020) 104041. doi:10.1016/j.compbiomed.2020.104041.

[24] S. Sahay, N. Omare, K. K. Shukla, An approach to identify captioning keywords in an image using LIME, in: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 648–651. doi:10.1109/ICCCIS51004.2021.9397159.