Facial emotion recognition based on upper features and transfer learning

Ennaji Fatima Zohra[1], El Kabtane Hamada[2]
[1]Laboratory of Process Engineering, Computer Science and Mathematics (LIPIM), National School of Applied Sciences - Sultan Moulay Slimane University (ENSA-KH), Khouribga, Morocco
[2]SMARTE Systems and Applications (SSA), National School of Applied Sciences (ENSA-M) - Cadi Ayyad University, Marrakesh, Morocco

Article Info

Article history:

Received May 2, 2024
Revised Sep 11, 2024
Accepted Sep 29, 2024

Keywords:

CK+
Face mesh
Facial emotion recognition
Transfer learning
Upper face region

ABSTRACT

Facial expression recognition (FER) in the upper face focuses on the analysis and recognition of emotions based on features extracted from the upper region of the face. This region typically affects the eyes, eyebrows, forehead, and sometimes the upper cheeks. Since these areas are often less affected by face masks or other facial coverings, FER algorithms can concentrate on capturing and interpreting the relevant facial cues, such as eye movements, eyebrow positions, and forehead wrinkles, to accurately recognize and classify different emotions. By focusing on the upper face, FER systems can mitigate the impact of occlusions caused by masks and still provide meaningful insights into the emotional states of individuals. In this work, a FER approach focusing on the upper region is proposed. Several experiments have been made using the CK+ dataset in addition to a comparison between the emotion recognition scores using the upper and the entire face in order to determine whether this area can reflect the real expressed emotion. The results of our approach are promising compared to previous studies with an accuracy up to 96%.

Corresponding Author:

Ennaji Fatima Zohra
Laboratory of Process Engineering, Computer Science and Mathematics (LIPIM), National School of Applied Sciences - Sultan Moulay Slimane University (ENSA-KH)
BP 77 Bd Beni Amir, Khouribga 25000, Morocco
Email: f.ennaji@usms.ma

1. INTRODUCTION

The field of facial expression recognition (FER) has witnessed significant advancements over the years, leading to improved understanding and capabilities in the area of emotion recognition from facial expressions. Researchers identified a set of universal facial expressions, including happiness, sadness, anger, surprise, fear, and disgust. In the 1990s, advancements in computer vision and machine learning techniques paved the way for automated FER approaches. Then, researchers began exploring techniques such as feature extraction, pattern recognition, and classification algorithms to automatically detect and recognize facial expressions. Additionally, the introduction of deep learning methods, such as convolutional neural networks (CNNs) [1], enabled automatic feature learning directly from raw facial images. Thus, models like VGGNet [2], ResNet [3], and InceptionNet [4] were subsequently developed and demonstrated impressive accuracy in recognizing emotions, surpassing traditional machine learning methods. The pretrained models on large-scale image datasets, such as ImageNet, provided a transfer learning paradigm for FER tasks.

Extensive research has been conducted in the domains of computer vision and affective computing regarding face emotion recognition (FER) [5]. Fallahzadeh *et al.* [6] utilized AlexNet-DCNN model to extract high-level features associated with various emotion classes. In their research, transfer learning was employed to fine-tune the initial model using specific datasets. Notably, the CK+ dataset yielded an average recognition accuracy of 93.66%, which was identical to the accuracy achieved when using the CK dataset.

A unique approach to address intra-class variations in CNN have been proposed in [7]. The latter introduced the innovative Island Loss layer in the CNN model, in addition to three convolutional layers, each followed by parametric rectified linear unit (PReLU) and batch normalization layers. Pooling operations were applied after the first two BN layers. Two fully connected (FC) layers were added after the third convolutional layer, and the Island Loss was computed at the second FC layer. Using the CK+ dataset, the accuracy was about 94.39%.

Kusuma *et al.* [8] presented a standalone model that leveraged the VGG-16 architecture. The model was refined by incorporating global average pooling as the final pooling layer. The authors examined several factors, such as the exploration of different optimizers like Stochastic Gradient Descent and Adam, varied data distributions, layer freezing, and batch normalization. The accuracy rate has achieved 69.40%.

Zheng *et al.* [9], the authors propose a novel approach named discriminative DMTL (DDMTL). This method simultaneously integrates class label information and local spatial distribution information of the samples. To accomplish this, they introduce a siamese network with an adaptive reweighting module that considers the local spatial distribution while utilizing class label information with varying confidences. The proposed approach achieves an accuracy of 95.28% for validation and 97.63% for the test dataset.

All the research papers mentioned in Table 1, the whole face was involved in order to get the expressed emotion from facial expressions. However, the presence of facial masks (especially during situations like the COVID-19 pandemic) or occluded lower part of the face poses a considerable challenge to FER systems. This can obstruct the visibility of crucial facial regions, such as the mouth and nose, which are instrumental in accurately recognizing and interpreting facial expressions. This obstacle has prompted researchers to explore novel approaches that specifically address FER from masked faces. With the mouth region often concealed by masks, attention has shifted towards leveraging the remaining visible regions for emotion recognition.

Table 1. Comparative review of research on FER presented in the literature review

Work	Year	Region of interest	Used Approach	Dataset	Accuracy
Boughida *et al.* [10]	2022	Entire face	Gabor Filters, genetic algorithm and SVM	CK CK+	94.20% 94.26%
Fallahzadeh *et al.* [6]	2021	Entire face	AlexNet-DCNN	CK / CK+	93.66%
Cai *et al.* [7]	2018	Entire face	CNN	CK+	94.39%
Kusuma *et al.* [8]	2020	Entire face	VGG-16	FER2013	69.40%
Zheng et al. [9]	2020	Entire face	DDMTL	CK+	97.63%
Mukhopadhyay *et al.* [11]	2023	Entire face	CNN with LTP	CK+	89.2%
			CNN with LBP		79.5%
			CNN with CLBP		91%
Chowdary *et al.* [12]	2023	Entire face	Pre-trained Resnet50	CK+	97.7%
			Pre-trained VGG19		96%
			Pre-trained Inception V3		94.2%
			Pre-trained Mobile Net		98.5%
Ahadit and Jatoth [13]	2021	Entire face	LogicMax layer + VGG16	CK+	98.62%
Liu *et al.* [14]	2021	Entire face	genetic algorithm using SVM (GA-SVM)	CK+ MUG	95.85% - 97.59% 96.56%
Alsemawi *et al.* [15]	2023	Entire face	HOG + FLN algorithm	Yale face	95.04%

The impact of the absence of the bottom part of the face on emotion recognition (FER) has been the subject of several studies. Research has consistently shown that the use of face masks can impair FER, leading to a reduction in accuracy. For example, the study presented in [16] found that emotion recognition from masked faces was about 20% worse than from unmasked faces, with specific confusions observed for various emotions such as disgust, happiness, anger, sadness, and surprise. Additionally, a paper from ELsayed *et al.* [17] aiming to improve FER for masked faces highlighted the challenges posed by face masks to automatic FER, emphasizing the importance of developing better recognition approaches for both masked and unmasked faces. These findings collectively underscore the significant impact of the absence of the lower half of the face on FER and the need for robust recognition approaches.

Deducting emotion from the upper part of the face is an active research area, aiming to develop robust and accurate systems that can effectively recognize and interpret emotions even when the lower region (mouth and cheeks) are occluded. Several research conducted using visual stimuli has revealed that different parts of the face contribute varying levels of detail in conveying a person's emotional state. Specifically, when individuals experience happiness, surprise, or disgust, the lower parts of the face offer the most informative cues. On the other hand, the upper parts of the face provide crucial details for understanding emotions like fear or anger. In the case of sadness or a neutral emotional state, both the lower and upper regions of the face play an equally significant role in conveying emotional information [18].

Comparing the coverage of the lower and upper regions of the face, it has been observed that masking the lower part of the face has a greater negative impact on the recognition of happiness compared to masking the upper part. However, the effect on other emotions varies. For instance, in one study by Kotsia *et al.* [19], it was found that covering the mouth disrupted the recognition of emotions like disgust and anger more than covering the eyes. Conversely, another study by Schurgin *et al.* [20] reported the opposite trend.

In our case, the visible part of the facial is the upper one, particularly the eyes, forehead and eyebrows. Thus, to accurately infer emotions, researches prioritize extracting and analyzing features found in the visible regions, since eyes play a vital role in conveying emotions, placing emphasis on their significance.

For instance, Castellano *et al.* [21], proposed a FER approach. The pipeline starts by detecting the mask employing a CNN model. Thus, the eye area is extracted so it can be used to predict the expressed emotions. However, this work shows low results for the emotion detection for either the upper face (an average accuracy of 43%) or the entire face (an average accuracy of 56.57%).

A combination of an advanced face parsing and vision transformer, alongside a cross-attention-based approach, was employed in [22] to enhance the detection accuracy of FER in the presence of face masks. The goal is to differentiate the unobstructed area of the face as well as the region covered by the face mask from the remaining regions within a given image. The experiments used several datasets with 3 emotions for M-LFW-FER and M-KDDI-FER, and 7 emotions for M-FER-2013 and M-CK+. The obtained accuracy was in the range of 61.08% and 91.83% for the four datasets.

In another work Magherini *et al.* [23], 5 classes (from the AffectNet dataset) have been used while grouping the anger and disgust emotions in one single class, as well as the fear and the surprise emotions. The happiness, sadness and neutral emotions have been treated separately. On the other hand, the authors proposed an approach that relies on two main steps. The first one aims to filter images if the upper region is not visible using the pre-trained model ResNet50v2. The second step of the process focuses on the detection of emotions using the inception model. An accuracy of 96.92% have been found at the end of the experimental phase. Nevertheless, the model cannot differentiate between the fear and the surprise emotions as well as between the anger and the disgust emotions.

In summary, the research based facial coverage highlights the differential contribution of the lower and upper regions of the face to the perception and recognition of emotions, with specific emotions showing varying sensitivity to masking or covering specific facial areas. In Table 2, a summary of the researches mentioned in the literature review and more is presented showcasing the accuracy of the model, the used dataset and approach, as well as the region of interest.

Table 2. Comparative review of research that focuses on FER using the upper region

Work	Year	Region of interest	Used approach	Dataset	Accuracy
Mukhiddinov *et al.* [24]	2023	Upper Face	facial landmark detection + CNN	Af-fectNet	69.3%
Akhmedov *et al.* [25]	2022	Upper Face	Haar–Cascade Classifier	FER-2013	91.2%
Castellano *et al.* [21]	2021	Upper Face Entire face	VGG-16	FER-2013	56.57% 43%
Yang *et al.* [22]	2022	Masked Face	Face parsing + vision transformer	M-LFW-FER M-KDDI-FER M-FER-2013 M-CK+	90.31% 91.83% 66.53% 61.08%
Magherini *et al.* [23]	2022	Masked Face	ResNet50 + Inception	AffectNet (5 classes)	96.92%
Our work	2024	Upper Face	Mesh Extraction + Fine tuned VGG19	CK+ (7 classes)	**96.38%**

In this paper, a FER solution using only the upper part of the face is proposed in order to obtain higher accuracy than related works. It specially involves the extracting of the upper region, feature extracting using face mesh and transferring knowledge based on VGG19 pre-trained model. The paper's remaining sections are structured as follows. Section 2 is dedicated to give more details about the proposed FER approach based on the upper face. Several experimentations have been developed in section 3 in addition of a comparison between results using the upper face and the global face. To finally conclude with a summary and some future works in section 4.

2. METHOD

In order to extract the expressed emotions from the upper face, several steps have been exploited as presented in Figure 1 and will be detailed in the following subsections. The process starts by extracting the upper part of the face followed by the upper face mesh extraction as a feature extraction method. Before training our fine-tuned VGG-19 model, the obtained images have been resized and augmented.

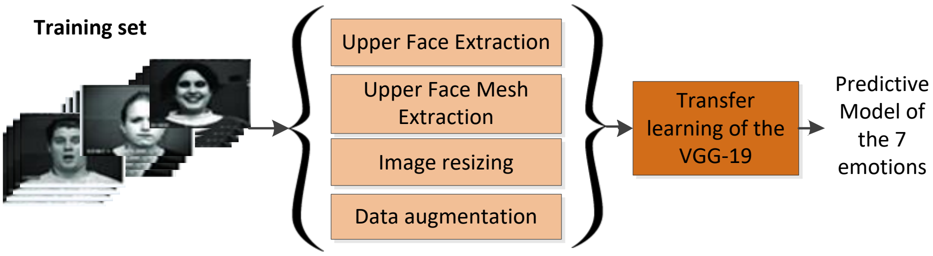

Training set
Upper Face Extraction
Upper Face Mesh Extraction
Image resizing
Data augmentation
Transfer learning of the VGG-19
Predictive Model of the 7 emotions

Figure 1. The proposed approach for emotion extraction from the upper facial region

2.1. Upper face extraction

Upper face extraction refers to the process of isolating or extracting the region of the face that encompasses the forehead, eyebrows, eyes and the upper cheeks. This extraction can be useful in various applications, such as facial analysis, emotion recognition, gaze tracking, or facial recognition systems that focus on specific facial features.

In our case, the process starts by detecting the overall face region in the input image. After detecting the face, defining the boundary or region of interest for the upper face is done. Typically, this includes the area from the top of the forehead to just upper the cheek bones (e.g., the upper half of the face).

Finally, the upper face region from the original image is extracted by cropping the image using the coordinates $upper\ Region = [(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)]$, where $x_1 = x_{box}$, $y_1 = y_{box} + 10$, $x_2 = x_{box} + W - 10$, $y_2 = y_{box} + H//2$. In this case, H and W indicate the height and the width of the box containing the face region only; x_{box} and y_{box} contains the coordinates of the lower corner of the face box. Figure 2 shows an example of the extracted Upper Face from a CK+ image.

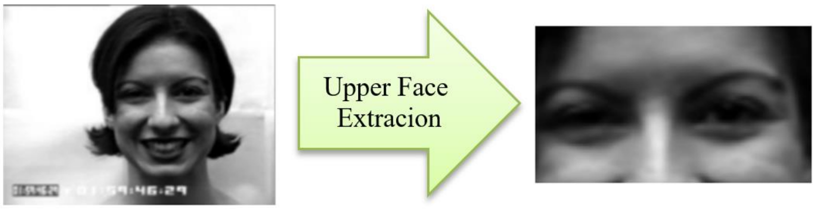


Figure 2. Example of the upper face cropping on a CK+ image

2.2. Upper face mesh extraction

MediaPipe [26] is an open-source framework developed by Google that offers various pre-built solutions for computer vision tasks. MediaPipe's face mesh is a powerful solution for real-time face mesh extraction. It provides a pre-trained deep learning model and a pipeline that enables the estimation of 3D face geometry from images or video streams. The face mesh solution is built on the MediaPipe framework, which offers a convenient way to integrate computer vision functionalities.

The upper face mesh extraction aims to pop up the expressions of the upper part of the face using the face mesh data generated by the MediaPipe face mesh solution. The face mesh data typically includes the 3D vertex positions of the face mesh vertices, along with other optional information like confidence scores or texture coordinates. It allows more detailed geometric features related to facial muscle deformations during different emotional expressions. This included parameters like vertex displacements, curvature changes, and depth variations [27].

To reach the desired result, the upper lines of the face mesh were exploited to extract the convex hull of the face. Then we eliminate the background noises while keeping the half oval shape of the face using$orient(L_1, L_2, P)$ that returns 1 if $det(\overrightarrow{L_1P}, \overrightarrow{L_2P}) \succ 0$and -1 otherwise where the convex hull represented by L_i represent a landmark and P a pixel in the image. Thus, only the upper part of the face mesh is extracted using a boolean function IOConv(P) that evaluate if the number of fragments in the convex hull is equal the rest of the equation as follows (1). At the end of the process, only the upper part of the face mesh is selected and superposed on the cropped image as shown in Figure 3.

$$IOConv(P) = (|L| == \left|orient(L_{|L|}, L_1, P) + \sum_{i=1}^{|L|} orient(L_i, L_{i+1}, P)\right| \tag{1}$$

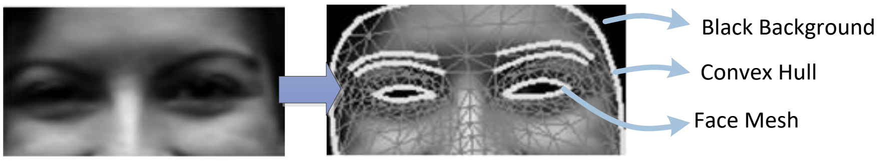


Figure 3. An upper face mesh extraction and background elimination example

2.3. Image resizing, normalization and data augmentation

Resizing is performed to ensure uniform dimensions across all images inputted into the network. This is necessary for compatibility with CNN architectures (consistent input sizes). All images were resized to the shape (200,100) to reduce the complexity of the model while maintaining the mesh's clarity as shown in Figure 4.

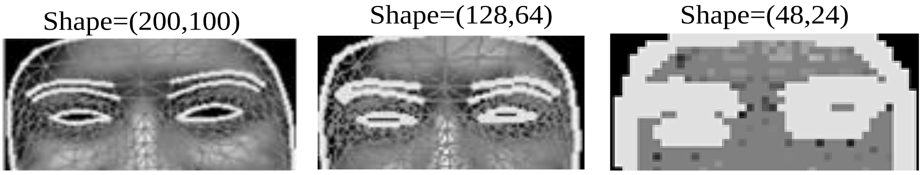


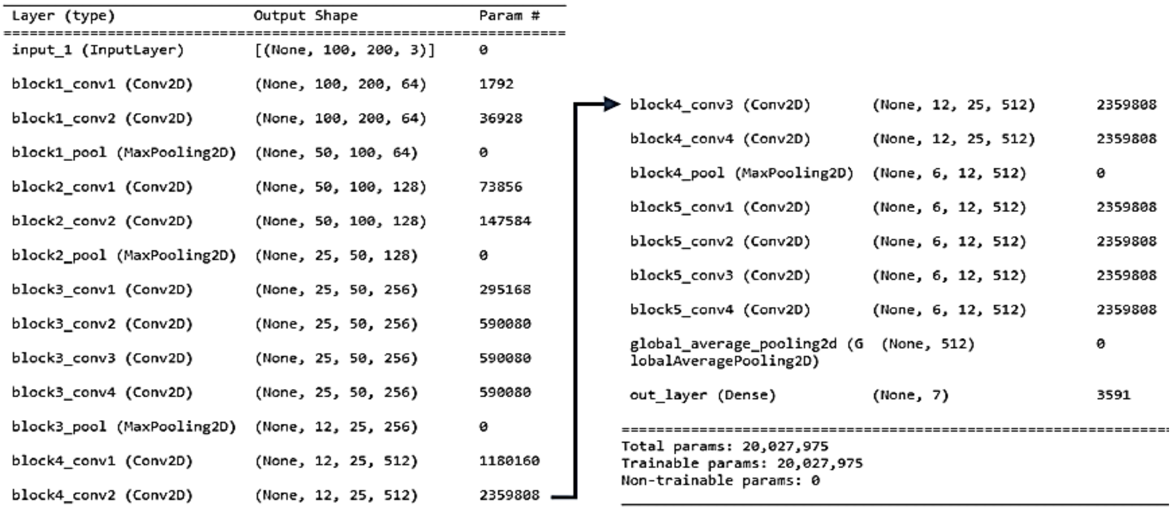
Figure 4. Results of resizing an image using different shapes 200, 128 and 48

On the other hand, image normalization involves transforming images into a standardized representation before feeding them into neural networks. Typically, images in computer vision tasks have varying pixel value ranges, such as 0-255 for 8-bit grayscale or 0-255 for each color channel in RGB images. In our scenario, we achieved normalization by dividing each pixel value by 255. As a result, the pixel values now fall within the range of [0,1]. Scaling the pixel values in this manner enables the neural network to handle inputs more effectively, since, many activation functions and optimization algorithms are specifically designed to work well with values within this range.

To enhance generalization and optimize model performance, data augmentation was employed. Data augmentation techniques are used to effectively increase the size of the training dataset, particularly when working with a limited amount of data. By applying data augmentation, the model becomes exposed to a wider range of image variations, enabling it to better generalize and perform optimally. Thus, several parameters have been used (the rotation_range = 15, width_shift_range = 0.15, height_shift_range = 0.15, shear_range = 0.15, zoom_range = 0.15, horizontal_flip = true) of each image in the training set.

2.4. Transfer learning with VGG19

Transfer learning is a popular technique in deep learning that involves pre-trained models on large-scale datasets to solve new tasks with limited labeled data. VGG19 is one of the pre-trained CNN model that has achieved remarkable performance in image classification tasks. The proposed model was assembled using the pre-trained VGG19 model, with replacing the last FC layer with a new global average pooling layer and a dense layer using a SoftMax activation. Moreover, the layers of the original VGG19 model have been changed to be trainable, which permits the weights to be updated (fine-tuned) during training process. Figure 5 gives more details about the layers of the proposed model.



```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 100, 200, 3)]     0

block1_conv1 (Conv2D)        (None, 100, 200, 64)      1792

block1_conv2 (Conv2D)        (None, 100, 200, 64)      36928

block1_pool (MaxPooling2D)   (None, 50, 100, 64)       0

block2_conv1 (Conv2D)        (None, 50, 100, 128)      73856

block2_conv2 (Conv2D)        (None, 50, 100, 128)      147584

block2_pool (MaxPooling2D)   (None, 25, 50, 128)       0

block3_conv1 (Conv2D)        (None, 25, 50, 256)       295168

block3_conv2 (Conv2D)        (None, 25, 50, 256)       590080

block3_conv3 (Conv2D)        (None, 25, 50, 256)       590080

block3_conv4 (Conv2D)        (None, 25, 50, 256)       590080

block3_pool (MaxPooling2D)   (None, 12, 25, 256)       0

block4_conv1 (Conv2D)        (None, 12, 25, 512)       1180160

block4_conv2 (Conv2D)        (None, 12, 25, 512)       2359808

block4_conv3 (Conv2D)        (None, 12, 25, 512)       2359808

block4_conv4 (Conv2D)        (None, 12, 25, 512)       2359808

block4_pool (MaxPooling2D)   (None, 6, 12, 512)        0

block5_conv1 (Conv2D)        (None, 6, 12, 512)        2359808

block5_conv2 (Conv2D)        (None, 6, 12, 512)        2359808

block5_conv3 (Conv2D)        (None, 6, 12, 512)        2359808

block5_conv4 (Conv2D)        (None, 6, 12, 512)        2359808

global_average_pooling2d (G  (None, 512)               0
lobalAveragePooling2D)

out_layer (Dense)            (None, 7)                 3591

=================================================================
Total params: 20,027,975
Trainable params: 20,027,975
Non-trainable params: 0
```

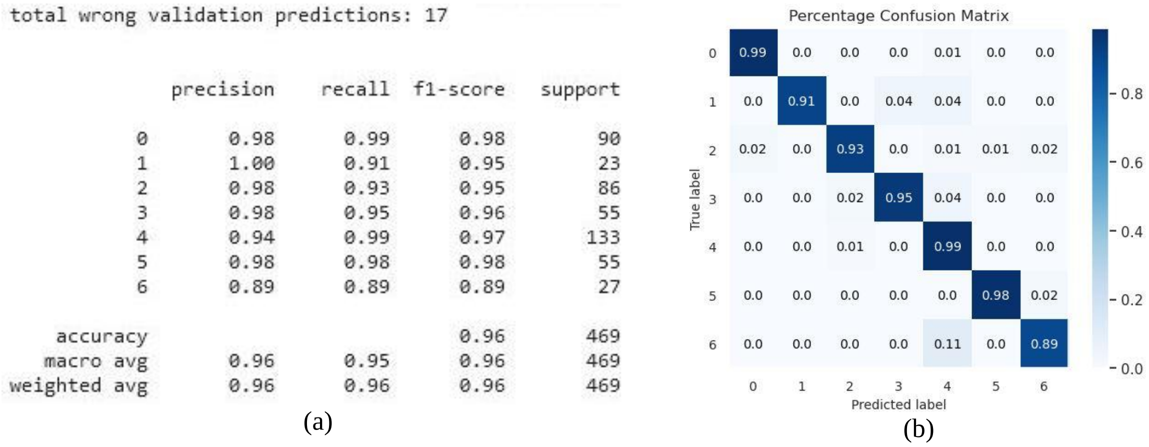
Figure 5. Layers of the proposed model

3. RESULTS AND DISCUSSION

Experiments were made using the CK+ dataset ("Extended Cohn-Kanade") [28]. It is a facial expression dataset widely used in the field of facial emotion analysis. It is an extension of the original Cohn-Kanade dataset and is designed for the purpose of emotion recognition research. It includes images representing basic emotions: anger, disgust, fear, happiness, sadness, and surprise. Each image is labeled with the corresponding emotion. CK+ includes multiple samples or instances for each subject, with variations in facial expressions (from neutral to the most expressive emotion) and other factors like lighting conditions. Thus, the neutral emotion was constructed in CK+ is Angry (19.18%), Neutral (4.98%), Disgust (18.29%), Fear (11.64%), Happy (28.41%), Sadness (11.67%), and Surprise (5.83%).

In Figure 6 the obtained results using the proposed model and preprocessing steps are promising, since the accuracy has reached 96.38% as presented in Figure 6(a). Referring to the confusion matrix in Figure 6(b), some minor mistakes have been made while relying on the emotions extracted only using the upper face. The main one was generated by the surprise emotion, since 11% of surprise expressions (3 images) have been predicted as a happy emotion. However, for the other classes, the misclassification was in the range of 1% to 4% (1 to 2 images per misclassification).

The plot in Figure 7 provides a visual depiction of the model's accuracy progression throughout the training epochs. It enables a side-by-side comparison of the training accuracy and validation accuracy as the epochs progress as shown in Figure 7(a), as well as the training and validation loss functions (shown in the right side of the same Figure 7(b)). It is evident from the plot that the model remains stable and does not exhibit signs of overfitting.



```
total wrong validation predictions: 17

              precision    recall  f1-score   support

           0       0.98      0.99      0.98        90
           1       1.00      0.91      0.95        23
           2       0.98      0.93      0.95        86
           3       0.98      0.95      0.96        55
           4       0.94      0.99      0.97       133
           5       0.98      0.98      0.98        55
           6       0.89      0.89      0.89        27

    accuracy                           0.96       469
   macro avg       0.96      0.95      0.96       469
weighted avg       0.96      0.96      0.96       469
```

(a)

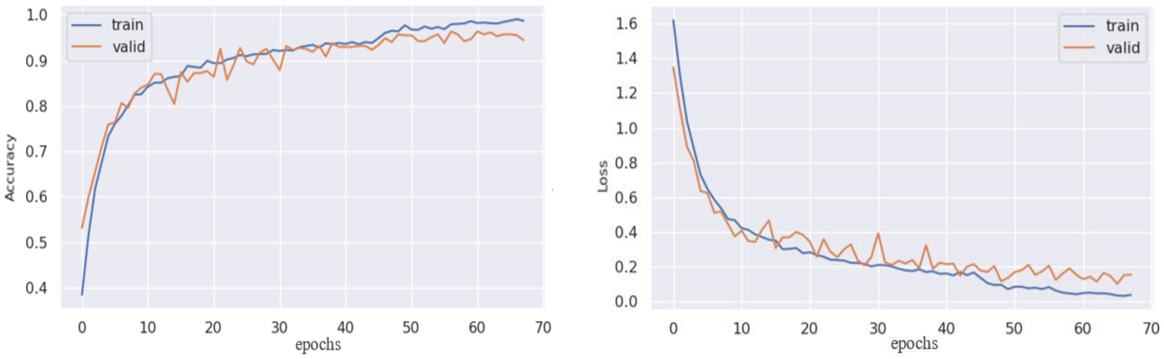
Percentage Confusion Matrix

	0	1	2	3	4	5	6
0	0.99	0.0	0.0	0.0	0.01	0.0	0.0
1	0.0	0.91	0.0	0.04	0.04	0.0	0.0
2	0.02	0.0	0.93	0.0	0.01	0.01	0.02
3	0.0	0.0	0.02	0.95	0.04	0.0	0.0
4	0.0	0.0	0.01	0.0	0.99	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.98	0.02
6	0.0	0.0	0.0	0.0	0.11	0.0	0.89

(b)

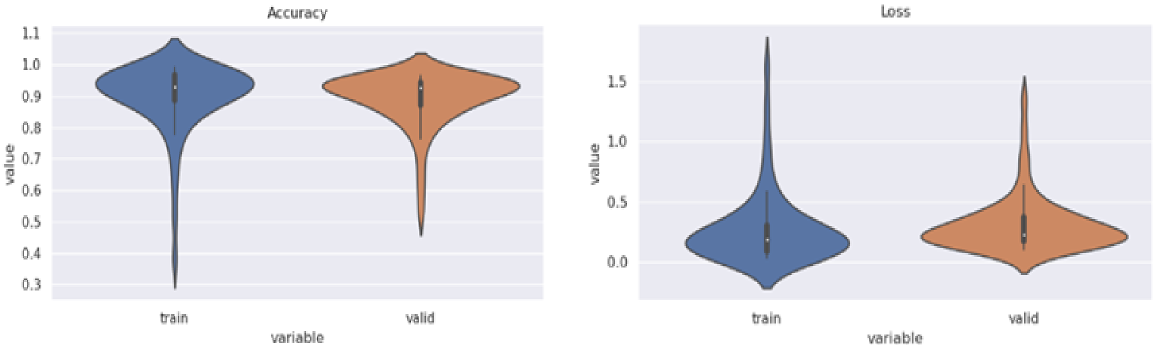
Figure 6. The classification report (a) and the confusion matrix and (b) of the proposed approach

The performance distribution graph presented in Figures 8, illustrates that the violins in both subplots exhibit similar shapes and medians, indicating that the model is effectively generalizing as shown in Figures 8(a) and 8(b). The receiver operating characteristic (ROC) graph is a commonly used as a visualization tool in machine learning for evaluating the performance of a binary classification model. It provides a comprehensive analysis of the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) at various classification thresholds. In the ROC graph, the x-axis represents the false positive rate, while the y-axis represents the true positive rate. As displayed in Figure 8(c), we can assess the model's ability to balance between true positives and false positives since the curve hugs the top-left corner of the graph, and make informed decisions about the threshold that maximizes the desired trade-off.

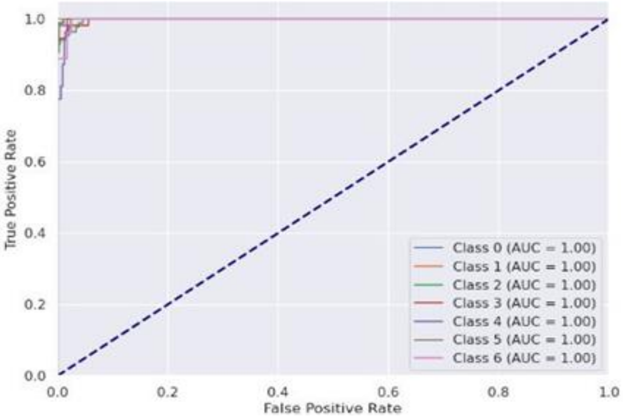


(a) (b)

Figure 7. The visual depiction of the model's accuracy (a) accuracy and (b) loss of the training and validation during epochs



(a) (b)



(c)

Figure 8. The performance distribution graphs in (a) accuracy (b) loss and (c) of the proposed approach and the ROC graph of the proposed VGG19

Using the same presented solution in this paper, prediction using, this time, the hall face has been performed. The aim of this experimentation is to compare and to conclude if the upper face can hold the emotions expressed by the global face. Figure 9 presents the results, showcasing the precision, recall, and f1-score for each emotion (happy, sad, disgust, surprise, neutral, fear, angry) when using the upper face beside the entire face. By conducting a comprehensive analysis of the scores, we can conclude that taking into consideration the hall face gives better results (as expected).

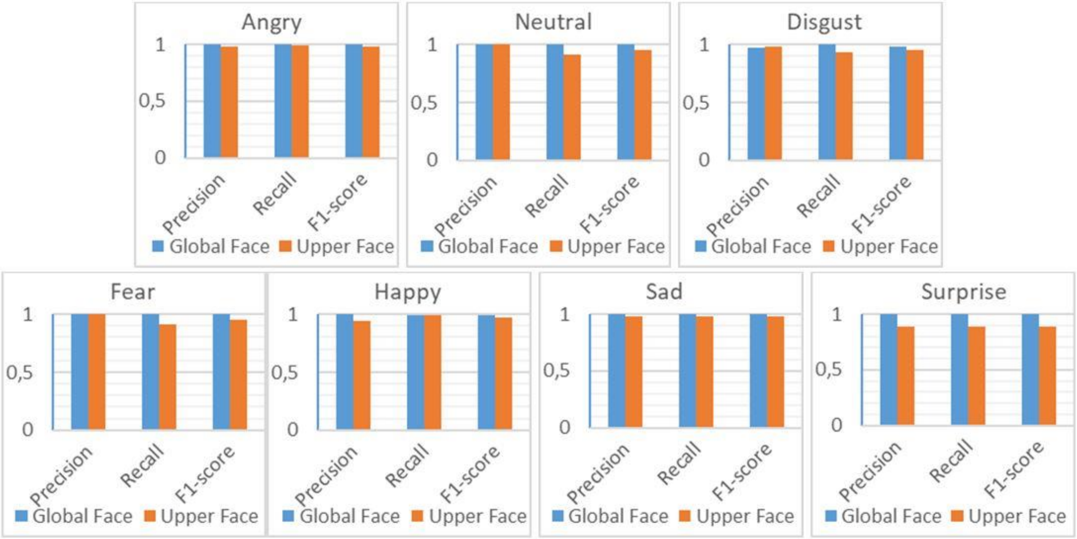


Figure 9. Comparing the metrics using the upper and the global face of each emotion

However, referring to each score, the both cases are pretty much close. For example, the precision of the angry emotion using only the upper part of the face is 98% and on the other side a precision of 100% have been found using the global face. For the neutral emotion, the precision has achieved 100% for the both cases. The main emotions that rely on the bottom part of the face are the happy and surprise emotion. In this case, a precision of 94% and 89% have been found respectively for the happy and surprise emotion using the upper part, while a 100% precision has been reached using the global face [29].

Based on Table 3, several architectures were evaluated for their performance combined with the proposed preprocessing steps in this work. The deep convolutional neural network (DCNN) achieved an accuracy of 82%, with balanced precision and recall scores of 83% and 82%, respectively, and an F1-score of 81%. ResNet50 and InceptionV3, demonstrated higher performance, with ResNet50 achieving an accuracy of 93% and InceptionV3 reaching 95%. These models also showed impressive precision, recall, and F1-score metrics, indicating robust performance across all evaluated measures. On the other hand, our work surpassed all others with an accuracy of 96.38%, demonstrating consistent high precision, recall, and F1-score metrics all at 96%. Moreover, it achieved the lowest loss value of 0.0165, underscoring its efficiency and effectiveness in the task compared to the other methodologies evaluated. These results highlight the significant advancement and superior performance the proposed solution over established models like ResNet50 and InceptionV3 that were used in previous works (Table 2). The achieved accuracy scores surpass those reported in prior research, indicating the efficacy of the proposed preprocessing steps since we tried to test the use of face mesh and additional steps combined with the models that were presented section 1. The findings presented in this section demonstrate that the upper face alone can be a robust feature set for emotion recognition, particularly in scenarios where the lower face is obscured or masked.

Table 3. Comparative review of research that focuses on FER using the upper region

Method	Accuracy	Precision	Recall	F1-score	Loss
DCNN + Preprocessing	82%	83%	82%	81%	0.5232
ResNet50 + Preprocessing	93%	95%	93%	94%	0.1945
InceptionV3 + Preprocessing	95%	95%	95%	95%	0.1313
Our work	**96%**	**96%**	**95%**	**96%**	**0.1215**

4. CONCLUSION

Occlusion of the lower part of the face presents a significant challenge in the accurate detection of sentiment. It has become widespread and can significantly impact the visibility of facial features that are crucial for accurate emotion recognition. Addressing the impact of occluded lower face parts on facial sentiment detection rather than evaluating the totality of the face region is the core of this work. Face mesh extraction and fine-tuning techniques using the VGG19 architecture have been exploited. As a result, by focusing solely on the upper face region, the model is able to accurately classify and predict emotions (accuracy of 96.38%) better than the previous works. The evaluation of the solution, as indicated by the ROC curve and other relevant metrics, suggests that the model's discriminatory ability is commendable, with a high true positive rate and a low false positive rate. This indicates that the model effectively distinguishes between different emotions based solely on the upper face. Comparing the results using the global face and the upper face have been done to highlight the potential of using the upper face as a reliable and informative feature for emotion prediction, offering practical advantages such as reduced computational complexity and improved interpretability. It is important to note that while the upper face provides valuable information for emotion prediction, it may not capture the complete range of facial expressions. Still, in situation when only the upper part of the face is available (masked face for example) we can certainly rely on this particular area.

REFERENCES

[1] M. Sam'an, Safuan, and M. Munsarif, "Convolutional neural network hyperparameters for face emotion recognition using genetic algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 442–449, 2024, doi: 10.11591/ijeecs.v33.i1.pp442-449.

[2] S. Karen and Z. Andrew, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1409–1556.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Jun. 2016, doi: 10.1109/cvpr.2016.90.

[4] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2015, vol. 07-12-June, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[5] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences (Switzerland)*, vol. 9, no. 21, p. 4678, 2019, doi: 10.3390/app9214678.

[6] M. R. Fallahzadeh, F. Farokhi, A. Harimi, and R. Sabbaghi-Nadooshan, "Facial expression recognition based on image gradient and deep convolutional neural network," *Journal of AI and Data Mining*, vol. 9, no. 2, pp. 259–268, 2021.

[7] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. Oreilly, and Y. Tong, "Island loss for learning discriminative features in facial expression recognition," in *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 2018, pp. 302–309, doi: 10.1109/FG.2018.00051.

[8] G. P. Kusuma, A. Jonathan, and P. Lim, "Emotion recognition on FER-2013 face images using fine-tuned VGG-16," *Advances in Science, Technology and Engineering Systems*, vol. 5, pp. 315–322, 2020, doi: 10.25046/aj050638.

[9] H. Zheng *et al.*, "Discriminative deep multi-task learning for facial expression recognition," *Information Sciences*, vol. 533, pp. 60–71, 2020, doi: 10.1016/j.ins.2020.04.041.

[10] A. Boughida, M. N. Kouahla, and Y. Lafifi, "A novel approach for facial expression recognition based on Gabor filters and genetic algorithm," *Evolving Systems*, vol. 13, no. 2, pp. 331–345, 2022, doi: 10.1007/s12530-021-09393-2.

[11] M. Mukhopadhyay, A. Dey, and S. Kahali, "A deep-learning-based facial expression recognition method using textural features," *Neural Computing and Applications*, vol. 35, pp. 6499–6514, 2023, doi: 10.1007/s00521-022-08005-7.

[12] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications," *Neural Computing and Applications*, vol. 35, no. Special issue on Human-in-the-loop Machine Learning and its Applications, pp. 23311–23328, 2023, doi: 10.1007/s00521-021-06012-8.

[13] A. B. Ahadit and R. K. Jatoth, "A novel dual CNN architecture with LogicMax for facial expression recognition," *Journal of Information Science and Engineering*, 2021, doi: 10.6688/JISE.202101_37(1).0002.

[14] X. Liu, X. Cheng, and K. Lee, "GA-SVM-based facial emotion recognition using facial geometric features," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11532–11542, 2021, doi: 10.1109/JSEN.2020.3028075.

[15] M. R. M. Alsemawi, M. H. Mutar, E. H. Ahmed, H. O. Hanoosh, and A. H. Abbas, "Emotions recognition from human facial images based on fast learning network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 3, pp. 1478–1487, 2023, doi: 10.11591/ijeecs.v30.i3.pp1478-1487.

[16] M. Rinck, M. A. Primbs, I. A. M. Verpaalen, and G. Bijlstra, "Face masks impair facial emotion recognition and induce specific emotion confusions," *Cognitive Research: Principles and Implications*, vol. 7, no. 83, 2022, doi: 10.1186/s41235-022-00430-5.

[17] Y. ELsayed, A. ELSayed, and M. A. Abdou, "An automatic improved facial expression recognition for masked faces," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14963–14972, 2023, doi: 10.1007/s00521-023-08498-w.

[18] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. How individual face parts contribute to successful emotion recognition," *PLoS ONE*, vol. 12, no. 5, pp. 1–15, 2017, doi: 10.1371/journal.pone.0177239.

[19] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Image and Vision Computing*, vol. 26, pp. 1052–1067, 2008, doi: 10.1016/j.imavis.2007.11.004.

[20] M. W. Schurgin, J. Nelson, S. Iida, H. Ohira, J. Y. Chiao, and S. L. Franconeri, "Eye movements during emotion recognition in faces," *Journal of Vision*, vol. 14, no. 13, pp. 1–16, 2014, doi: 10.1167/14.13.14.

[21] G. Castellano, B. De Carolis, and N. Macchiarulo, "Automatic emotion recognition from facial expressions when wearing a mask," in *CHItaly '21: Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter*, 2021, pp. 1–5.

[22] B. Yang *et al.*, "Face-mask-aware facial expression recognition based on face parsing and vision transformer," *Pattern Recognition Letters*, vol. 164, pp. 173–182, 2022, doi: 10.1016/j.patrec.2022.11.004.

[23] R. Magherini, E. Mussi, M. Servi, and Y. Volpe, "Emotion recognition in the times of COVID19: Coping with face masks," *Intelligent Systems with Applications*, vol. 15, 2022, doi: 10.1016/j.iswa.2022.200094.

[24] M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and C. Jinsoo, "Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people," *Sensors*, vol. 23, no. 3, p. 2023, 2023.

[25] F. Akhmedov, A. B. Abdusalomov, M. Mukhiddinov, and Y.-I. Cho, "Development of real-time landmark-based emotion recognition CNN for masked faces," *Sensors*, vol. 22, no. 22, p. 8704, 2022.

[26] "MediaPipe library," https://ai.google.dev/edge/mediapipe/solutions/guide.

[27] D. Ciraolo, M. Fazio, R. S. Calabrò, M. Villari, and A. Celesti, "Facial expression recognition based on emotional artificial intelligence for tele-rehabilitation," *Biomedical Signal Processing and Control*, vol. 92, p. 106096, 2024, doi: 10.1016/j.bspc.2024.106096.

[28] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kande dataset (CK+): A complete facial expression dataset for action unit and emotion specified expression," in *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, 2010, no. July, pp. 94–101, doi: 10.1109/ISIEA.2010.5679500.

[29] E. F. Zohra and E. K. Hamada, "Transfer learning based face emotion recognition using meshed faces and oval cropping: a novel approach," *Optical Memory and Neural Networks*, vol. 33, pp. 178–192, 2024, doi: 10.3103/S1060992X24700073.

BIOGRAPHIES OF AUTHORS

Ennaji Fatima Zohra is a computer science engineer, graduated from the National School of Applied Sciences Marrakesh/Morocco at 2014. Then got a Ph.D. in Computer Science from Cadi Ayyad University, Marrakech, Morocco. Actually, she is a Professor in the National School of Applied Sciences of Khouribga, Morocco. She is interested in Machine learning applications and algorithms especially in the Topics of emotion recognition. She can be contacted at email: f.ennaji@usms.ma.

El Kabtane Hamada received the M.S. degree from the Faculty of Sciences Ibn Tofail of Kenitra, Morocco, in 2012 and then he holds a Ph.D. degree in computer science in the Laboratory (LISI), Cadi Ayyad University of Marrakech, Morocco. Since 2021, he is currently a professor of computer science in National School of Applied Sciences. He is interested on the virtual learning environments, the 3D image processing, virtual reality and augmented reality, in addition to machine learning areas. He can be contacted at email: h.elkabtane@uca.ma.