INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 19, Issue: 1, Month: February, Year: 2024
Article Number: 5363, https://doi.org/10.15837/ijccc.2024.1.5363

CCC Publications

Emotion estimation of people wearing masks using machine learning

Naziha Hmidi, Rim Afdhal, Monia Hamdi, Ridha Ejbali and Mourad Zaied

Naziha Hmidi
Faculty of Sciences of Gabes
University of Gabes, Tunisia
naziha.hmidi1997@gmail.com

Rim Afdhal
Faculty of Sciences of Gabes
University of Gabes, Tunisia
eng.afdhal.rim@ieee.org

Monia Hamdi*
Department of Information Technology
College of Computer and Information Sciences
Princess Nourah bint Abdulrahman University
P.O.Box 84428, Riyadh 11671, Saudi Arabia
*Corresponding author: mshamdi@pnu.edu.sa

Ridha Ejbali
Research Team in Intelligent Machines
Engineering School of Gabes
University of Gabes, Tunisia
ridha_ejbali@ieee.org

Mourad Zaied
Research Team in Intelligent Machines
Engineering School of Gabes
University of Gabes, Tunisia
mourad.zaied@ieee.org

Abstract

Coronavirus (COVID-19) is the infectious agent responsible for the transmission of SARS (severe acute respiratory syndrome). When an infected person coughs, talks, or breathes, it spreads as small droplets of fluid from the mouth and nose of the infected person. Sanitary masks help prevent the spread of the virus from the person wearing the mask to others. This new behavior may cause a number of problems in interpersonal interactions. The goal of this paper is the emotion estimation of masked faces. It presents two major parts. At the first level, we created a system for identifying sanitary masks, using CNN. At the second one, we developed an emotion estimation system in order to estimate the classification rates of a masked face. It consists of three steps: face element detection, feature point localization, and classification. We used the well-known Viola and Jones algorithm in order to achieve the first step. We used several techniques to estimate emotions (SVM, KNN and deep learning). We made comparisons of the obtained results. Particular attention is also given to the effect of face masks on the performance of various methods.

Keywords: Emotion Estimation, masked face, COVID-19, SVM, KNN, Deep Learning, CNN.

1 Introduction

Facial expressions are an integral part of emotion recognition [1] [2]. They are used in the process of non-verbal communication as well as to identify people. Emotion recognition systems are commonly represented by key facial features such as eyes, nose, and mouth [9]. However, many situations, such as pandemics, laboratories, medical offices or excessive contamination, require the wearing of masks that mask or partially conceal the face. The appearance of COVID-19 has had a global impact, not only on society as a whole, but also on our lives. On the global level, many preventive measures are taken to limit the transmission of the disease, including face masks, mandates for social distancing and disinfection in public spaces, and the use of screening applications [22] [23]. One of the globally most prevalent prevention measures, introduced in response to the disease, are sanitary masks. Masked faces are a major obstacle to studying and understanding how facial recognition systems work with masked faces.

A state-of-the-art algorithm has been proposed to determine if a face is occluded, i.e. masked face recognition.

We propose in this paper an approach that consists of two models. For the first time, we created a binary classification for mask identification. For the second, it is a method inspired by an emotion recognition method called " Emotion recognition using features distances classified wavelets network and trained by fast wavelet transform" [1]. The proposed emotion estimation system contains three stages: the first consists of detecting facial elements using the Viola and Jones detectors; the second stage is used to locate feature points where each muscle will be characterized by a distance between two points. The final stage is the classification.

The contributions of this work are:

- We created a binary classification system for mask identification.
- We developed an emotion estimation system despite the lack of information lost by the sanitary masks.
- Before the start of the pandemic, there was an obvious lack of datasets, so we prepared a dataset based on the chon kanade dataset by adding the masks.

The other sections are organized as follows: Section 2 gives the related works to the emotional recognition subject. We'll cover our methodologie in Section 3. Section 4 covers the stages of the proposed system. Section 5 reflects the results of the proposed solution. Section 6 is the discussion part. And section 7 presents the conclusions.

2 Related works

Here, we quickly review some recent work. Primary emotions are the main focus of the majority of research on emotion recognition. We identified emotion recognition by facial expression as one of the many ways employed in this kind of research. Darwin [2] recognized that emotions that are expressed through facial expressions (Ekman) [3], (Izard) [4] are the easiest to recognize. The most relevant and up-to-date works in the literature on this subject focus on the use of facial expressions.

Several recent studies examining masks and facial emotion recognition have shown that emotion recognition accuracy reduces when wearing a face mask. Anyhow, this loss of precision is not the same for all facial expressions. For example, inefficiencies in facial emotion recognition have been found for happiness, sadness, disgust, and anger, but not for fear or neutral emotions [10] [16]. previously, covering lower facial features such as the mouth, cheeks, and nose with a mask, as we experimented with "bubbles" in our study, has different effects on different facial expressions [27]. Moreover, other approaches suggest that predominantly informative regions of the face differ for different expressions [17] [24]. In contrast, facial masking analyzes revealed differences in facial expression between eye and mouth part masking results [25] [26].

In [14], researchers chose 122 participants for an online test to determine whether (a) emotion recognition, (b) trust attribution, and (c) re-identification were conditioned by presence. of criteria.

A mask that covers the entire mouthpiece area or a transparent mask that restores visual access to the mouth area. For these actions, they used Karolinska Directed Emotional Faces (KDEF)25,26 and 48 different stimuli from the Chicago Face Database (CFD)27. KDEF items are used for emotion recognition and trust attribution tasks, and CFD stimuli are used for trust attribution and re-identification tasks. Transparent masks facilitate the assessment of emotion and authenticity, because perceptions of emotion and identity rely on distinct processes and it is desirable to restore (largely) visual access to the mouth area. However, re-identifying the identity is not always easy. For the emotion recognition task, they reached a rate of 90% for the neutral state, 75% for the state of fear, 80% for the state of joy and 70% for the state of sadness. In [15], The aim of this study was to analyze how facial emotion recognition is affected by mask use and how this affects gender perceptions of facial attractiveness. The results of this study show that men and women perceive basic emotions in the same way. Besides, sex did not affect emotional perception with or without a mask. However, the use of surgical masks hinders facial recognition of emotions for both men and women. In this sense, negative emotions such as sadness and anger are affected in post-use emotion recognition by surgical masks. However, the use of a surgical mask had the greatest impact on good mood compared to mood perception. On the other hand, with or without a surgical mask, we would be indifferent to surprises. This method gives the rates 59%, 26%, 59%, 74% respectively, for emotions such as anger, joy, sadness and surprise.

3 Methodology

The primary objective of our research is to build a system capable of evaluating the emotions of mask wearers. Before predicting emotions, we propose a first system to classify the input image as a masked or unmasked face (binary classification). The workflow of our proposed system is shown in Figure 1. We implemented a binary CNN model to classify the input image ('With mask' or 'Without mask'). Then, we developed an estimation emotion system to estimate the emotions of people wearing sanitary masks. We inspired our second system from the approach [1] ''Recognition of emotions using the characteristics of distances classified by a network of wavelets and trained by FWT," which makes it possible to analyze facial expressions and facial variations based on changes in distance between certain points from a neutral state. Our principles are based on information extracted from faces that can describe human emotions. Therefore, the information is extracted by the interaction between two proposed points, a static and a dynamic point, which move with the muscular ovement. The detection of the elements extracted from a certain face is carried out by the method of Viola and Jones, and the positions of these characteristic points are made by 5 rectangles. After obtaining sufficient representation of the facial expression as well as the origin, we use the change of observed distance from the neutral state. A vector containing the distances is the input to our classifier. Figure 2 shows the different stages of this system.

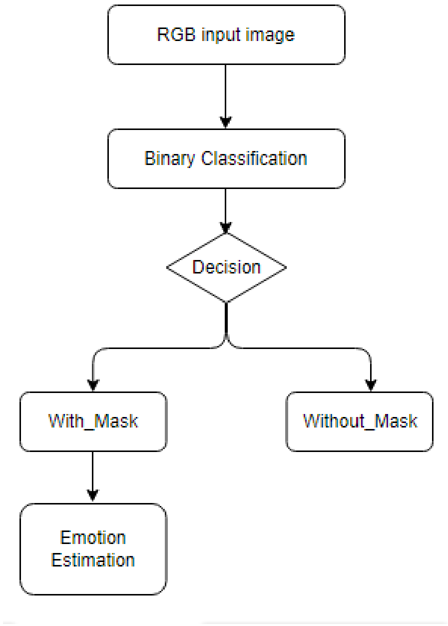

Figure 1: Proposed methodology

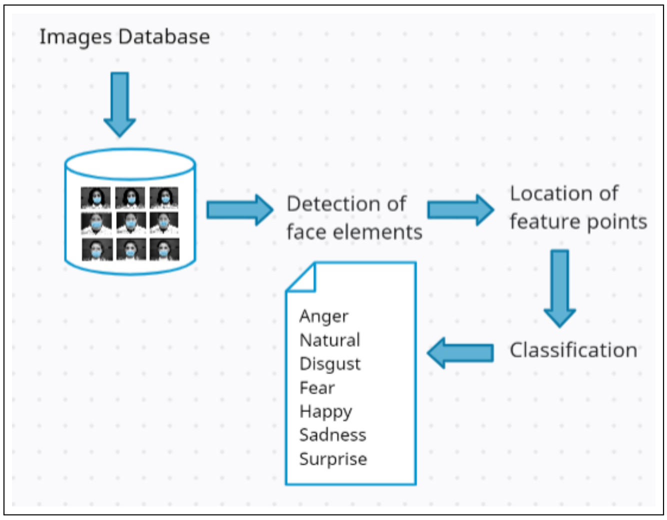

Figure 2: Descriptive diagram of the Emotion Estimation Model

3.1 Data-set preparation

Image databases are an important resource for the development of computer-based decision support systems. They help with the performance measurement tool used to test and evaluate diagnostic support systems. In this context, we have chosen the Cohn_Kanade data (CK+). It is a public data set of reference to identify units of action and emotions.

CK+ includes a total of 593 sequences from 123 subjects aged 18 to 50, with a variety of genres and heritage. Sequences range from neutral expression to the maximum expression. CK+ is a university database. It contains a set of grayscale facial expression images. The size of each image is 640 by 490 pixels; the camera orientation is frontal; and small head movements are present. This dataset is frequently used for facial expression recognition. The CK+ dataset is generally recognized as the most comprehensive laboratory-controlled database for facial expression classification. It is widely used and is used for most of the classification methods for facial expressions. As shown in the following figure, there are a total of 7 classes to recognize (A=anger, D=disgust, F=fear, H=happy, N=natural, S=sadness, U=surprised).

3.2 Adding mask to images

The goal of our approach is to estimate the emotions of people wearing masks. Given the absence of a public database to identify the emotions of masked faces, we have chosen to prepare a base of data based on the chonkanade dataset by adding the masks on test folder images. To achieve this task, we used the image processing software Adobe Photoshop. In fact, we took the pictures one by one and then we added the mask to each photo. We used the layer masking technique in order to add a mask to the images. It allows the user to hide or reveal specific parts of a layer without permanently altering the original image.

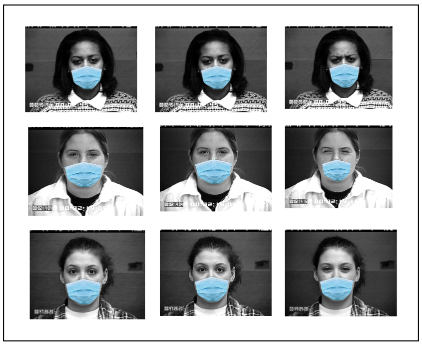

Figure 3: Examples of pictures from the Cohn_Kanade database with masks applied

4 Steps of our approach

4.1 Identification mask model

As we mentioned before, our approach is divided into two models: the first is the model that identifies the mask. Before starting the model development, we used techniques that allowed us to increase the diversity of the data.

4.1.1 Model development

In Figure 4, we've got a model made up of 2 Conv2D layers. They use 200 and 100 filters, each with a size of 3. It takes in data with a shape of (112,112,3) and uses 'Relu' as activation function for the Max-pooling. After that, it flattens features out, does a dropout check, goes through a regular dense layer, and finally gives an output. To classify items, it uses "softmax," improves with "Adam," and learns by minimizing "binary crossentropy" loss.

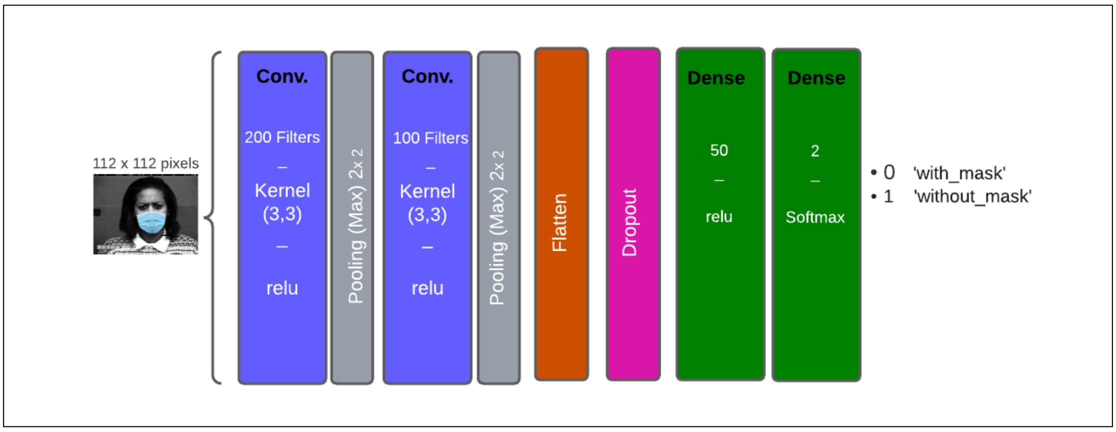

Figure 4: CNN Architecture

4.2 The emotion estimation model

This system is used to estimate and classify the emotions of people wearing masks. This method involves three main steps: Detection of facial elements, Localization of characteristic points, Classification.

4.2.1 Face elements detection

To estimate emotions, we first need to detect facial elements and locate them in rectangles. We realized this task with the help of Viola & Jones detectors [7] [8].

The Haar descriptors and cascade of classifiers serve as the foundation for this detector. The Haar descriptors named also Haar-like features which are rectangular filters, are very fast compared to other descriptors like HOG. Haar features operate on the hole image window instead of local patches. The cascade of classifiers structure allows for a fast rejection of non-face regions in an image, thus improving speed and efficiency. The characteristics of Haar are functions that permit recognizing the contrast differences between a number of adjacent rectangular sections in a picture. In fact, these descriptors are utilized to determine how much the sum of the pixels in the white area and the sum of the black area differ from one another. This strategy is predicated on the notion that the target object is absent from the vast majority of search windows. These phases are suggested for promptly rejecting negative samples. In fact, boosting is a strategy that is used at every level. These tags designate a positive or negative region for the area that the window's current position defines. If the region is given a negative label by the classifier, it signifies that the object was not located and that the region has been fully classified. The sliding window moves on to the next position. The region will go on to the next phase if the classifier assigns a positive label to it. A negative sample may be categorized as a false negative if one of the classifiers in the Viola and Jones detectors classifies it as a false positive.

The result of this step is five rectangles for the images without mask, figure 5 (1 for the face, 2 for the eyes, 1 for the nose, and 1 for the mouth) and three for images with mask, figure 6 (1 for the face, 2 for the eyes).

4.2.2 Location of characteristic points

The key step in this approach is locating the characteristic points on the face. This phase makes it possible to locate 38 points (Figure 7) from the rectangles extracted in the previous step. The localization of the points is done in a manual way using the points defining each rectangle. Each rectangle has the following characteristics: (1st left point of rectangle, width, length). The (x,y) coordinates of each point are determined using the coordinates of the points of the rectangles.

So from the delimiting face, we can locate the 3 points of the eyebrow, which are very close to the eye, and the same goes for the rectangle related to the right eye. Thus, they clearly drew these points with the coordinates of the points defining each rectangle.

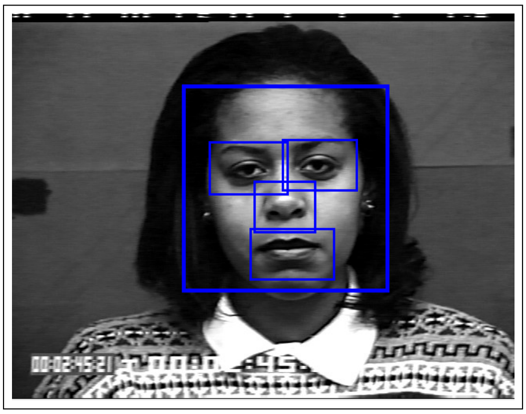

Figure 5: Detection of unmasked face elements

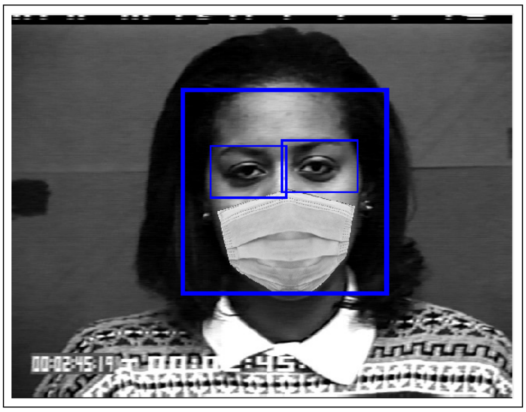

Figure 6: Detection of masked face elements

Figure 7: Localization of the characteristic points of the face without a mask

Our goal is to learn with faces without masks and then test with masked faces. So we still need to determine the points from the masked faces, i.e., only 18 points which are visible, and the remains are hidden by the mask. As shown in Figure 8, the 18 points represent the point of the upper part of the face.

Figure 8: Location of the characteristic points of the masked face

4.2.3 Calculation of distances

After the step of locating the characteristic points, we need to calculate the distances between them since each muscle will be characterized by a distance between two points. Until we come to the last step, which is to classify the variations among the distances that have been measured by the two points, and this will be done with respect to the neutral state. Figure 8 shows the biometric distances we used.

Take, for example:

- d1 is the distance between points p1 and p13, which constitutes a forehead muscle. It is calculated as follows:
  $d1 = \sqrt{((p1(1) - p13(1))^2 + (p1(2) - p13(2))^2});$

Figure 9: The distances that connect the characteristic points

The result of this task is 21 distances for unmasked faces and only 9 distances for masked faces.

4.2.4 Classification

The goal of image classification is the development of a system capable of classifying images automatically. Thus, the system enables the automation of specialized tasks that can be costly for humans to acquire. Mostly due to physical limitations such as concentration, fatigue, and the time needed for large amounts of image data.

The purpose of classification here is to assign each image a class that defines emotions. It requires two bases; a learning base created with a number of images of different variability taken from the database Chon-Kanade public data and another for the test composed of images that are images of Chon-Kanade with the addition of the masks. The purpose of this database is to demonstrate the ability to assess the emotions of people wearing masks. In this step, we used three different classifiers: KNN, SVM, and deep learning.

5 Results

In this section, we have used precision and confusion matrix plots to show the performance of the model.

5.1 Identification mask model

For this model we need a database of images with and without masks. In fact we took the same dataset explained in the previous section and rearranged the images according to mask wearing (With mask/Without mask), then we divided it 70% for training (269 images), and 30% for the test (116 images) .

Our results revealed that the model used in the analysis was very efficient, with a 100% accuracy for the binary model.

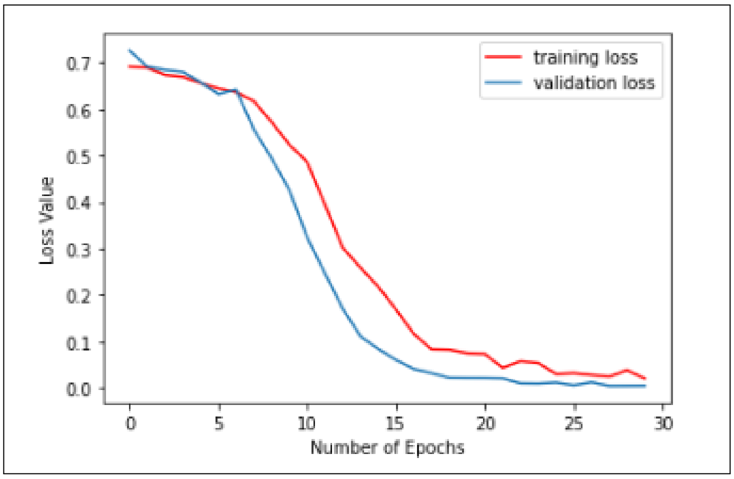

Figure 10: Training/ validation Loss

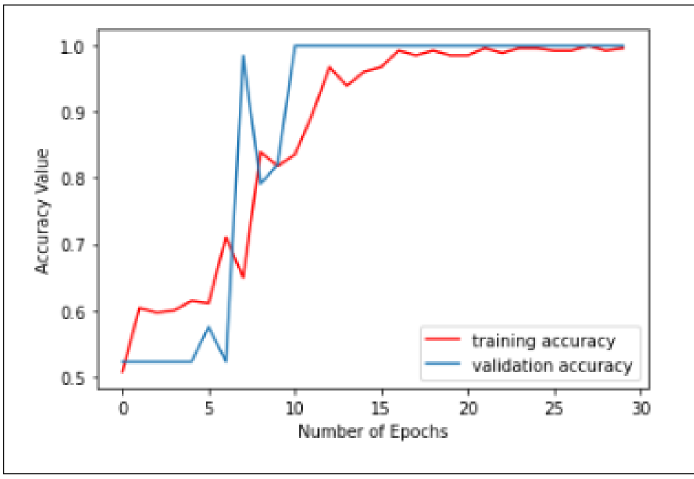

Figure 11: Training/ validation accuracy

The two figures 10 and 11 represent the precision and loss information during the training phase and for each epoch.

In Figure 10, the red curve represents the accuracy of the training data and the blue curve represents the accuracy of the test data. Also note that the accuracy of the training data continues to improve with the accuracy of the test data.

In Figure 11, the red curve represents training data loss and the blue curve represents test data loss. Additionally, we can see that training data loss continues to decrease with test data loss.

A confusion matrix, or error matrix is a special tabular format that helps visualize model/classifier performance.

According to the confusion matrix of figure 12 we found for the class 'With_mask' 65 correctly classified images (100 %) and for the class 'Without_mask' 51 correctly classified images (100 %).

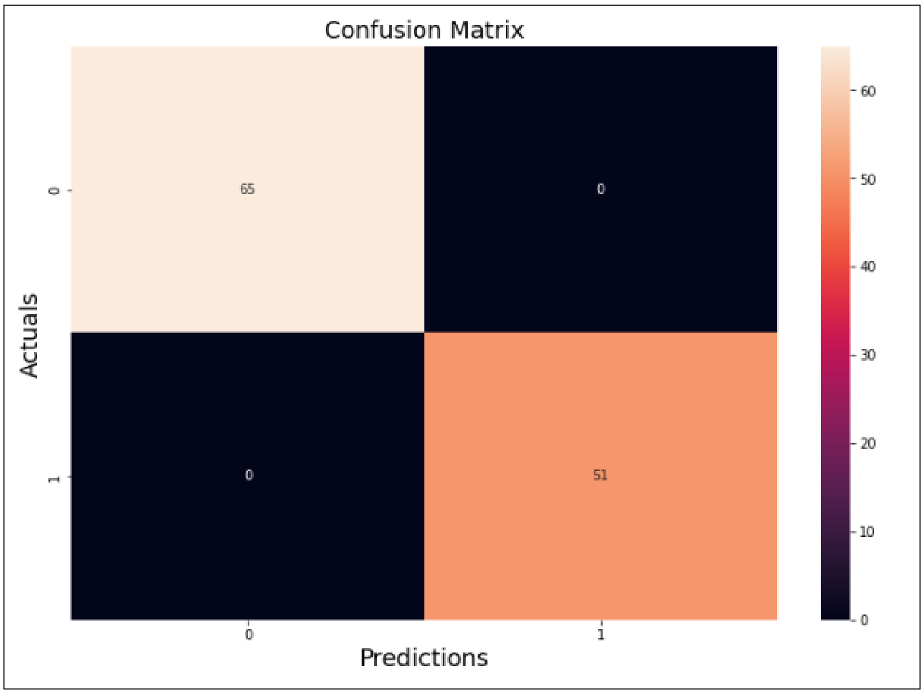

Figure 12: Confusion matrix

5.2 The emotion estimation model

The purpose of classification is to assign each image a class that defines emotions. It requires two bases, a learning base created with a number of images of different variability taken from the database public data Chon-Kanade and another for the test composed of images which are Chon-Kanade images with the addition of the masks. The purpose of this database is to demonstrate the ability to assess emotions people with masks. During this step, we used two different classifiers: KNN and SVM. The primary goal is the classification of images into 7 classes:

- A: Anger
- N: Natural
- D: Disgust
- F: Fear
- H: Happy
- S: Sadness
- U: Surprise

So for the learning phase, we used 269 images from the database and 116 for the test phase. The table below (Table 1) shows the classification results of our approach:

Classifier	accuracy
KNN	25.47%
SVM	27.82%

Table 1: classification results

From Table 1, it can be noticed that the emotion classification rates of masked faces were low.

Table 2 shows the percentage distribution of KNN classifier rates by class. This classifier classified the class: Anger with a classification rate equal to 45% and an error rate equal to 55%, the natural

Class	Classification Rate(25.47%)
A	45%
N	100%
D	0%
F	13.3%
H	20%
S	0%
U	0%

Table 2: KNN classification results

class with a classification rate equal to 100%, the Fear class with a classification rate equal to 13.3% and an error rate equal to 87% and the class Happy with a classification rate equal to 20% and an error rate equal at 80%.
In addition, he classified the Disgust, Sadness and Surprise class with a rate equal to 0%.

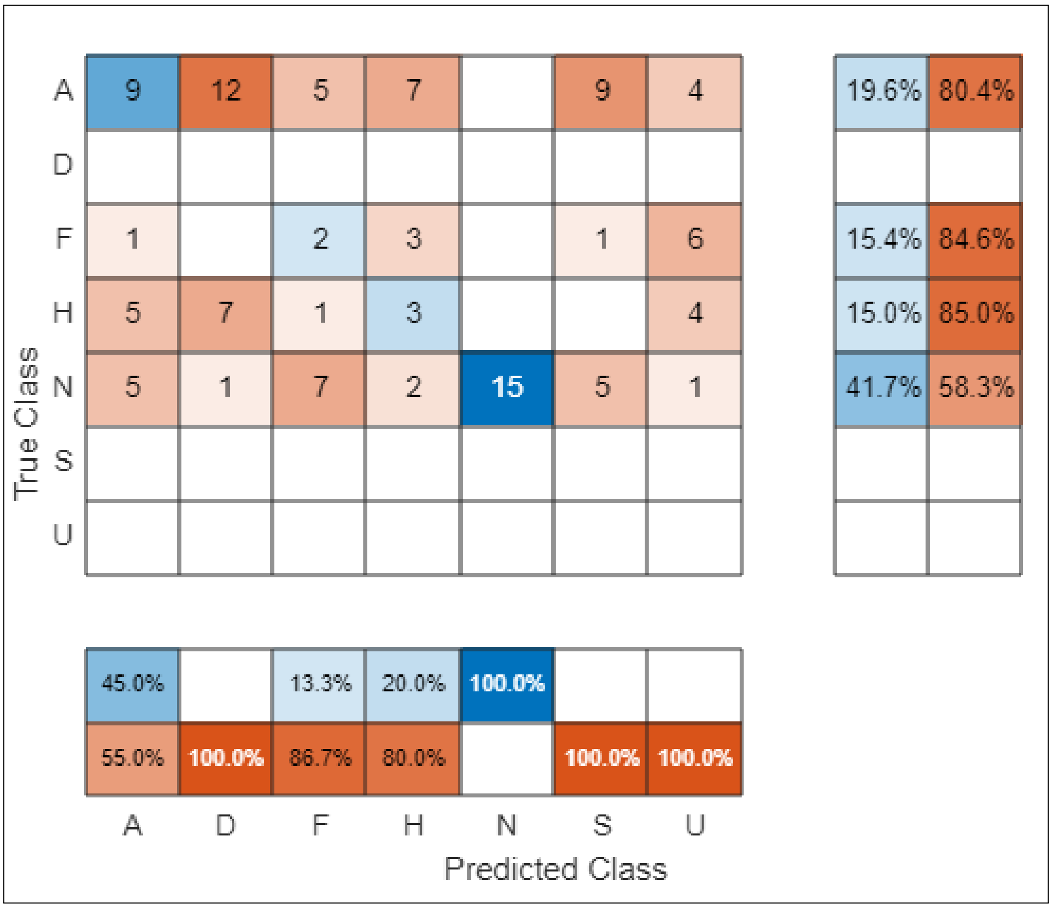

Figure 13: The confusion matrix

The KNN confusion matrix (Figure 13) shows that for class A(Anger), 9 images out of 20 are correctly classified, for class F(Fear), 2 images out of 15 are correctly classified, for class H(Happy), 3 images among 15 are correctly classified and for the class N(Natural), the 15 images of this class are correctly classified.

With the SVM classifier we found 27.82% as the overall classification rate, Table 3 shows the distribution of the percentages according to the class. This classifier has classified the class: anger with a classification rate equal to 45% and an error rate equal to 55%, the natural class with a classification rate equal to 100%, the fear class with a classification rate equal to 20% and an error

Class	Classification Rate(27.82%)
A	45%
N	100%
D	0%
F	20%
H	20%
S	0%
U	13.3%

Table 3: SVM classification results

rate equal to 80% and the class is satisfied with a classification rate equal to 20% and an error rate equal to 80%. In addition, he classified the class disgust, sadness with an equal rate at 0%, the class surprises with a classification rate equal to 13.3% and an error rate equal to 86.7%.

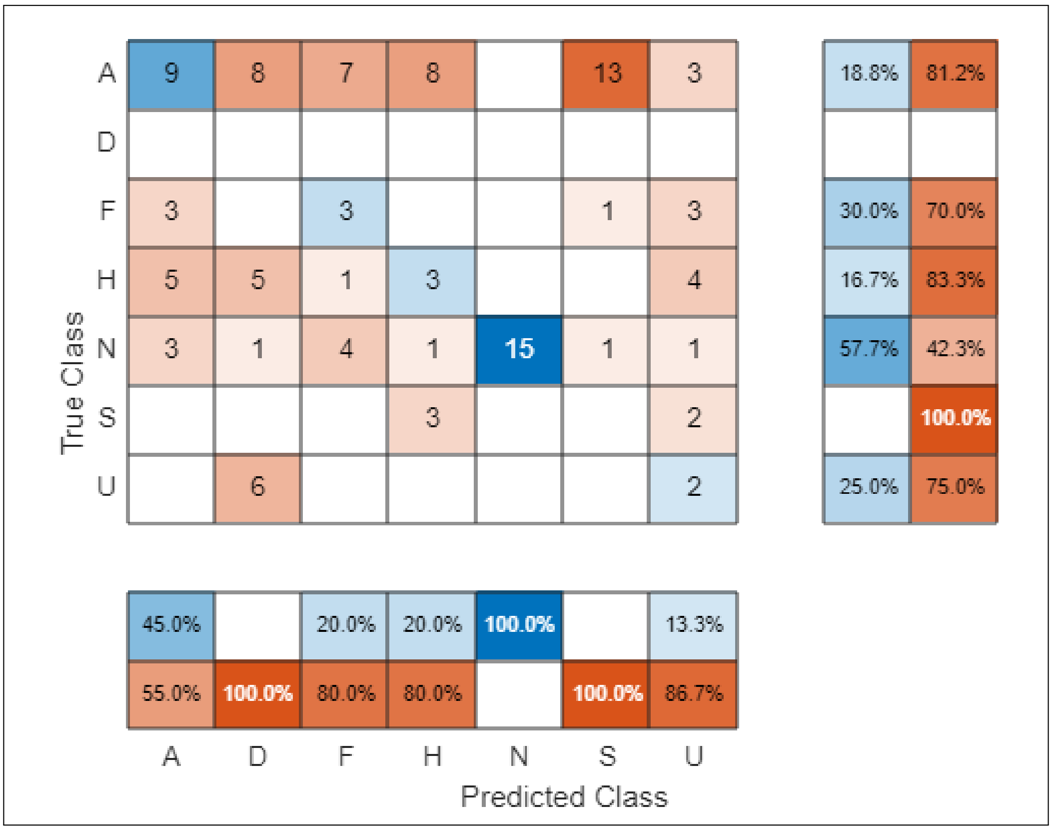

Figure 14: The confusion matrix

From Tables 2 and 3, we can see that the two classes; Disgust and Sadness have zero classification rates regardless with the KNN or SVM classifier. Our results suggest that facial emotion recognition when wearing a full-face mask is more difficult with an uncovered face. According to our hypothesis, we found that emotion recognition accuracy decreased for all tested emotions (anger, fear, happiness, sadness, disgust, and surprise) when presented with a face mask equal to 100%. This invention is consistent with previous research showing that emotion recognition is made more difficult when the area under the face is obscured.

5.3 Deep learning neural network

In order to improve the classification rates obtained in Tables 2 and 3, we tried to create and train a deep learning network for the classification of deep learning feature data, since one can train a deep network training using a feature input layer if one has a dataset of numeric features, such as a set of digital data without spatial or temporal dimensions. Figure 15 shows the network architecture. First, we apply data normalization, starting with a feature input layer that allows feature data to be input into the network. It then goes through a fully connected layer that multiplies the input by a weight matrix and then adds a bias vector. The third level is the batch normalization level, which normalizes each mini-batch of all observations for each channel separately. The fourth layer uses the Relu layer. Tasks in this layer perform a threshold operation on each element of the input, setting values less than zero to zero. Then we applied a softmax function to the input. Finally, we completed this model with a classificationLayer, to calculate the cross-entropy loss for the classification task.

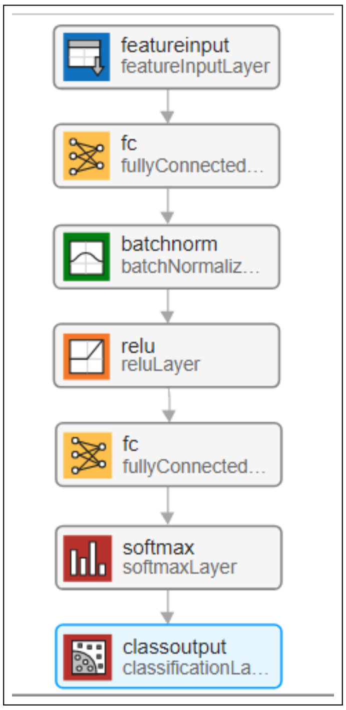

Figure 15: network architecture

This method gave us as a result an accuracy equal to 32.2% for masked faces.

Class	Classification Rate(32.2%)
A	5%
N	100%
D	5%
F	26.7%
H	13.3%
S	0%
U	93.3%

Table 4: Classification results with Deep Learning

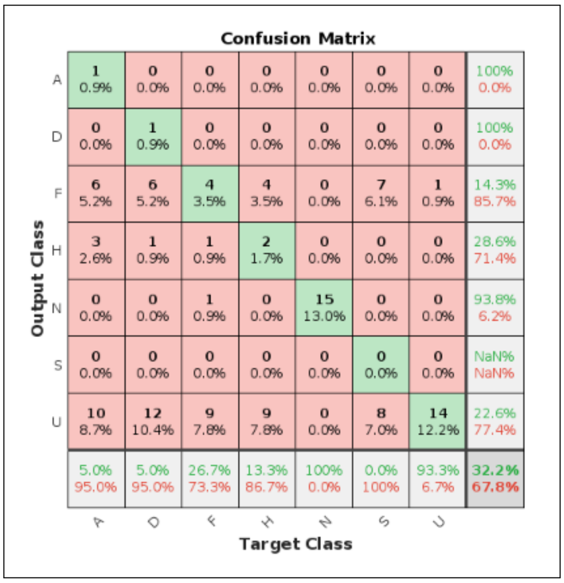

Figure 16: Confusion Matrix

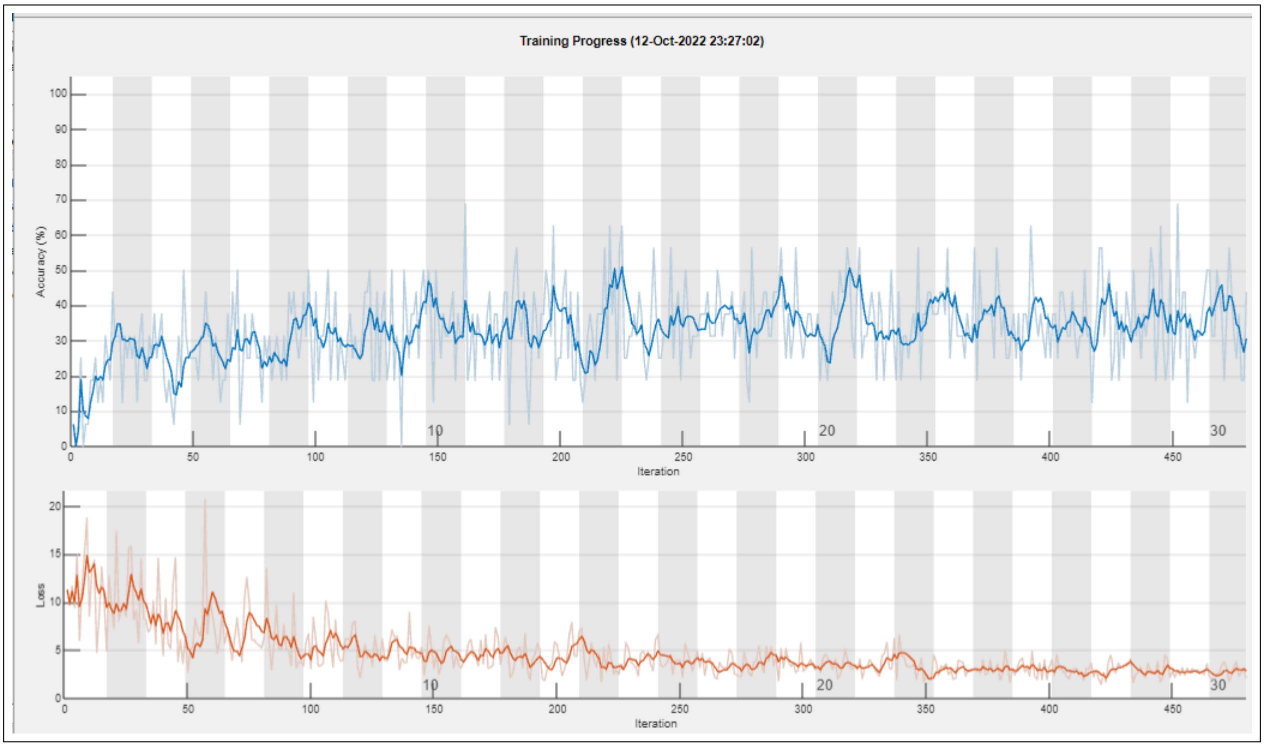

Figure 17: Training results

Returning to Figure 16; This confusion matrix shows the result of the classification of this model with its different columns. This model gives for the two columns 1 and 2, a classification rate equal to 5% (only 1 image for each class is correctly classified among 20 images), for class F, the classification rate is equal to 26.7% (4 images out of 15 are correctly classified) and the loss rate is equal to 73.3%, for class H, the classification rate is equal to 13.3% (2 images out of 15 are the correctly classified images), However, for column 5 we have 15 images(100%) of class N correctly classified. Finally for class U, 14 images out of 15 are correctly classified or a classification rate equal to 93.3%.

The results of this study suggest that emotions in images where the masked face covers the mouth and nose are more difficult to detect in the unmasked face. According to our hypothesis, we found that the accuracy of emotion recognition decreased when the face was presented with a mask. This was especially true for sad and disgusting faces. This finding is consistent with previous research showing that emotion recognition is made more difficult when the area under the face is obscured.

We see that the common point between the three results presented in tables 2, 3 and 4 is that they all manifest the emotion of sadness and disgust with a classification rate equal to 0%.

The greatest decline in recognition performance is observed through face masks expressing sadness and disgust. This confirms previous reports that different units of facial movement are important for sensing different emotions and that the lower part of the face, including the nose region, is particularly relevant for sensing sadness and disgust [20].

All three classification methods correctly classified the neutral emotion with a rate equal to 100%. The distances between the detected characteristic points are not modified so the vector of the pixel difference is zero, so we have no confusion about it of emotion.

For the other emotions (Anger, Disgust, Fear, Happy, Surprise) we unfortunately cannot generalize since we found different rates. In general, expressions are easily confused due to similarity in shape and appearance characteristics, and individual variations for the same expression especially for masked faces since the area from which we will detect emotions is reduced. (we do not work on the whole face but only on the upper part of the face).

Test on multiple datasets

To show the generalization capacities of the proposed solution, we tried to test the algorithm and the techniques on multiple datasets. To test with other datasets we need other emotion recognition datasets, during our search we found the two public datasets KDEF and JAFFE but the problem that the images are images of unmasked people is why we we had to add the masks a few images to do the test. The following table shows the information from this experiment.

Dataset	Total number of images	accuracy
JAFFE[27]	185	20%
KDEF[28]	490	33%

Table 5: Test on JAFFE, KDEF datasets

For the JAFFE dataset the number of images that we could find equals 185 since we need an image that includes a face in the neutral state and we could not find many images that satisfy this condition. Whereas for the KDEF we were able to find 490 images, for each dataset we took 70% of the images for the train and 30% for the test. The KDEF dataset gave us a classification rate equal to 33% while JAFFE gave us a rate equal to 20%.

5.4 Comparison with existing approach

We compared our work with the existing approach [1]. Different ranking criteria must be taken into account to compare the results (number of test images, number of training images) as indicated in Table 5 (presentation of the different recognition system evaluation parameters):

	Dataset
Number of classes	7
Number of training images	269
Number of test images	116
Total number of images	385

Table 6: Overview of various system evaluation parameters of recognition

In this section, we want to compare our results with the results of the approach [1],"Emotion recognition using features distances classified by wavelets network and trained by fast wavelets transform". As shown in Table 6, we can see that the classification rates of our approach are very low compared to the rates of the approach [1], this decrease in classification rates is explained by the loss of information generated by the port of mask. For the Anger class, we found a classification rate equal to 5% on the other hand for faces without masks the classification rate of this class is equal to 100%, this drop in the values of the rates is logical since the expression of anger emotion is usually based on the mouth part. For the Surprise class, we can notice that the classification rate on images that include masked faces (93.3%) is very close to the classification rate on images of unmasked faces (100%), this can be explained by making it easier to detect the emotion of surprise since it is mainly characterized by wide-open eyes and raised eyebrows. For the Natural class, we found the same classification rate as for the non-masked faces (classification rate equal to 100%) (since the distances between the detected characteristic points are not modified so the pixel difference vector is nothing). For the other emotions, the classification rate of images of faces wearing masks is always lower than the classification rate of images of faces without masks.

Class	Our approach	The approach [1]
A	5%	100%
N	100%	100%
D	5%	66.66%
F	26.7%	100%
H	13.3%	100%
S	0%	33%
U	93.3%	100%

Table 7: Comparison of our results with the results [1]

synthesis

At the level of this section we will synthesize the five approaches that have has already quoted in related work section.

Approach	Discription	accuracy
Rim Afdhal, Ridha Ejbali, & Mourad Zaied[1]	Detection of face elements (V&J) / Localization of characteristic points / Tracking of characteristic points / Classification (FWT)	93.05%
Rim Afdhal, Ridha Ejbali, & Mourad Zaied[5]	Face Feature Detection(v&J) / Location of wrinkle regions / Information Extraction / Classification(FWT)	90.75%
Rim Afdhal et al [6]	Combination of two approaches [1] and [5]	98.05%
Chil-Che-Chang [28]	Classification (CNN)	72.16%
Andera Francesco Abatel [29]	Residual Masking Network / Recognition CNN Facial Expression / Amended Representation (CNN)/	without mask=90,42 % with mask=45.45 %

Table 8: Synthesis

From the previous table (Table 8) and since most of the emotion recognition systems that we have studied are based on the three following steps: Detection of face elements, localization of points characteristics and classification and which subsequently gives significant classification rates, for this we have chosen to test these three steps on images that include masked faces.

6 Discussions

The face as an anatomical map can be divided into three parts: The upper, lower and middle levels each play a major role in the expression of feelings and emotions. For example, behaviors such as smiling and frowning involve facial substructures such as the mouth, lips, and cheeks that are common in everyday speech. Facial expressions for different emotions are accompanied by dramatic changes in facial appearance. Each of these facial actions is broken down into action units (or unit actions, AU) in the well-known FACS facial expression decoding system. (Facial Action Coding System) [21].

Returning to our approach, the low rates (Table 2 and Table 3) that we get at test level when adding mask are due to information lost due to lower face coverage by the mask. Based on these results, Protective masks cover the lower part of the face and we can say that they prevent Recognize the specific facial characteristics required for the analysis and recognition of emotional expressions.
We can also observe this from the first step of our approach which is the location step characteristic points: for unmasked faces, we can detect 38 points, on the other hand for the masked faces only 18 points are detected, which makes it possible to reduce the number of distances with which does the classification hence the decrease in classification rates. This is why some researchers around the world have recently been interested in determining to what extent the masks present in a face can affect the recognition of the emotion it expresses.

Finally, facial expression recognition is possible, but you know it changes when you wear a mask.

7 Conclusion

The goal of this work is to estimate the emotions of masked faces. We used two models. The first one is based on binary classification in order to identify the mask from the face image. The achieved results proved the effectiveness of the proposed model. The second model for primary emotions estimation is based on deep neural network. The low classification rates which are obtained in the test phase are due to the loss of information because the lower part of the face is covered by the mask. We contribute to three levels: the development of a binary classification system for mask identification, the development of an emotion estimation system, as well as the construction of a dataset based on the images of the Chon-Kanade dataset by adding the masks.

Availability of data and material

The used database is publicaly available.

Competing interests

No competing interests to declare.

Authors' contributions

Conceptualization, N.H. and R. A; methodology, N.H; software, R. E; validation, M. Z.; data curation, R. A.; writing—original draft preparation, N. H and R. E; writing—review and editing, M. H and M. Z; visualization, R. E; supervision, R. A.; project administration, M. H.; funding acquisition, M. H.

Funding

This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R125), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

Aknowledgement

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB.

References

[1] Rim Afdhal, Ridha Ejbali , Mourad Zaied and Chokri Ben Amar. Emotion recognition using features distances classified by wavelets network and trained by fast wavelets transform. *International Conference on Hybrid Intelligent Systems(HIS'14) proceedings, Golf University for Science and Technologie*, Kuwait. pp.238,241. December 14-16., 14h.

[2] C. Darwin. *The expression of the emotions in man and animals.*1872.

[3] P. Ekman. Are there basic emotions? *Psychological Review*, 99 :550–553,1992.

[4] C.E Izard. Human emotions. *Plenum Press* New York, 1977.

[5] R. Afdhal, R. Ejbali and M. Zaied. Emotion recognition using the shapes of the wrinkles. *International Conference on Computer and Information Technology*, December 18-20, 2016.

[6] Rim Afdhal, Ridha Ejbali and Mourad Zaied. A Hybrid System Based on Wrinkles Shapes and Biometric Distances for Emotion Recognition. *International Conference on Advances in Computer-Human Interactions*,ACHI 2017 .

[7] Ahmed A. Elngar, Mohamed Arafa, Abd El Rahman Ahmed Naeem, Ahmed Rushdy Essa, Zahra Ahmed shaaban. The Viola-Jones Face Detection Algorithm Analysis: A Survey. *Journal of Cybersecurity and Information Management (JCIM)*,April 2021.

[8] Tanmoy Paul, Ummul Afia Shammi, Mosabber Uddin ahmed, Rashedur Rahman, Syoji Kobashi and Md Atiqur Rahman Ahad. A Study on Face Detection Using Viola-Jones Algorithm for Various Backgrounds, Angels and Distances. *Human Activity Recognition*,May 2018.

[9] Wang Kay Ngai, Haoran Xieb, Di Zouc, Kee Lee Chou. Emotion Recognition based on Convolutional Neural Networks and Heterogeneous Bio-signal Data Source. *Journal Pre-proof*, 25 July 2021.

[10] Noyes, E.; Davis, J.P.; Petrov, N.; Gray, K.L.H.; Ritchie, K.L. *The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers.*R. Soc. Open Sci. 2021, 8, 201169. [Google Scholar] [CrossRef].

[11] Bassili, J. N.Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*,37(11), 2049–2058, (1979).

[12] Melina Grahlow , Claudia Rupp3, Birgit Derntl. The impact of face masks on emotion recognition performance and perception of threat. *Journal of Personality and Social* ,10.1371/journal.pone.0262840, 2022; 17(2): e0262840.

[13] Claus-Christian Carbon. Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions. September 2020.

[14] Marco Marini1, AlessandroAnsani, Fabio Paglieri1 , Fausto Caruana & MarcoViola. The impact of facemasks on emotion recognition, trust attribution and re-identifcation. 10 March 2021.

[15] Pamela Parada-Fernandez, David Herrero-Fernandez, Richard Jorge, Patricia Comesana. Wearing mask hinders emotion recognition, but enhances perception of attractiveness . *Personality and Individual Differences* Volume 184, January 2022, 111195.

[16] Carbon, C.-C. *Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions.* . Psychol. 2020, 11, 566886. [Google Scholar] [CrossRef].

[17] Blais, C.; Roy, C.; Fiset, D.; Arguin, M.; Gosselin, F. *The eyes are not the window to basic emotions.* Neuropsychologia 2012, 50, 2830–2838. [Google Scholar] [CrossRef].

[18] Henrik Nordström. Emotional Communication in the Human Voice. *Doctoral Thesis in Psychology at Stockholm University, Sweden 2019* , Sweden 2019.

[19] Ram Kumar Madupu; Chiranjeevi Kothapalli; Vasanthi Yarra; Sonti Harika; Cmak Zeelan Basha. Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network. *International conference of Electronics, Communication and Aerospace Technology (ICECA)* , 05-07 November 2020.

[20] Melina Grahlow , Claudia Rupp3, Birgit Derntl. The impact of face masks on emotion recognition performance and perception of threat. *Journal of Personality and Social* ,10.1371/journal.pone.0262840, 2022; 17(2): e0262840.

[21] Ekman, P., Friesen, W. V., & Hager. *Facial Action Coding System:* facial action coding system: the manual: on CD-ROM. Salt Lake City, UT, USA: Research Nexus J. C. (2002).

[22] Ahmad Alzu'bi, Firas Albalas, Tawfik AL-Hadhrami, Lojin Bani Younis and Amjad Bashayreh. *Masked Face Recognition Using Deep Learning: A Review* IoT Based Intelligent and Green Communications: Modelling, Practice and Applications. 31 October 2021.

[23] Mukhriddin Mukhiddinov, Oybek Djuraev, Farkhod Akhmedov, Abdinabi Mukhamadiyev and Jinsoo Cho. *Masked Face Emotion Recognition Based on Facial Landmarks and Deep Learning Approaches for Visually Impaired People* Application of Semantic Technologies in Sensors and Sensing Systems. 17 January 2023.

[24] Wegrzyn, M.; Vogt, M.; Kireclioglu, B.; Schneider, J.; Kissler, J. *Mapping the emotional face. How individual face parts contribute to successful emotion recognition.* PLoS ONE 2017, 12, e0177239. [Google Scholar] [CrossRef].

[25] Beaudry, O.; Roy-Charland, A.; Perron, M.; Cormier, I.; Tapp, R. *Featural processing in recognition of emotional facial expressions.* Cogn. Emot. 2013, 28, 416–432. [Google Scholar] [CrossRef] [PubMed].

[26] Schurgin, M.W.; Nelson, J.; Iida, S.; Ohira, H.; Chiao, J.Y.; Franconeri, S.L. *Eye movements during emotion recognition in faces.* J. Vis. 2014, 14, 14. [Google Scholar] [CrossRef] [PubMed].

[27] Gosselin, F.; Schyns, P.G. Bubbles: *A technique to reveal the use of information in recognition tasks.* Vis. Res. 2001, 41, 2261–2271. [Google Scholar] [CrossRef].

[28] Chih-Che Chung, Wei-Ting Lin, Rong Zhang, Kai-Wen Liang, and Pao-Chi Chang. Emotion Estimation by Joint Facial Expression and Speech Tonality Using Evolutionary Deep Learning Structures. *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)* .

[29] Andrea Francesco Abate, Lucia Cimmino, Bogdan-Costel Mocanu, Fabio Narducci, Florin Pop. The limitations for expression recognition in computer vision introduced by facial masks. *Multimedia Tools and Applications* 14 July 2022.

Cite this paper as:

Naziha, H. ; Rim A.; Monia H.; Ridha E. ; Mourad Z.; (2024). Emotion estimation of people wearing masks using machine learning, *International Journal of Computers Communications* & *Control*, 19(1), 5363, 2024.
https://doi.org/10.15837/ijccc.2024.1.5363