

# Emotion Categorization from Faces of People with Sunglasses and Facemasks

Harisu Abdullahi Shehu (✉ [harisushehu@ecs.vuw.ac.nz](mailto:harisushehu@ecs.vuw.ac.nz))

Victoria University of Wellington <https://orcid.org/0000-0002-9689-3290>

Will Browne

Victoria University of Wellington

Hedwig Eisenbarth

Victoria University of Wellington

---

## Research Article

**Keywords:** CK+, Emotion categorization, Facemasks, Facial expression, Obscure faces, Sunglasses

**Posted Date:** April 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-375319/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

---

1  
2  
3  
4  
5  
6  
7

# Emotion Categorization from Faces of People with Sunglasses and Facemasks

8  
9  
10  
11  
12  
13  
14

Harisu Abdullahi Shehu · Will Browne ·  
Hedwig Eisenbarth

15  
16  
17

Received: 30th March 2021 / Accepted: date

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

**Abstract** Emotional information is considered to convey much meaning in communication. Hence, artificial emotion categorization methods are being developed to meet the increasing demand to introduce intelligent systems, such as robots, into shared workspaces. Deep learning algorithms have demonstrated limited competency in categorizing images from posed datasets with the main features of the face being visible. However, the use of sunglasses and facemasks is common in our daily lives, especially with the outbreak of communicable diseases such as the recent coronavirus. Anecdotally, partial coverings of the face reduces the effectiveness of human communication, so would this have hampering effects on computer vision, and if so, would the different emotion categories be affected equally?

30  
31  
32

Here, we use a modern deep learning algorithm (i.e. VGG19) to categorize emotion from faces of people obscured with simulated sunglasses and face-

33  
34  
35  
36  
37

---

Harisu Abdullahi Shehu (Corresponding author)  
School of Engineering and Computer Science, Victoria University of Wellington, 6012 Wellington  
E-mail: harisushehu@ecs.vuw.ac.nz  
ORCID: 0000-0002-9689-3290

38  
39  
40  
41  
42  
43

Will Browne  
School of Engineering and Computer Science, Victoria University of Wellington, 6012 Wellington  
Tel: +64 4 886 5342  
E-mail: will.browne@vuw.ac.nz  
ORCID: 0000-0001-8979-2224

44  
45  
46  
47  
48  
49

Hedwig Eisenbarth  
School of Psychology, Victoria University of Wellington, 6012 Wellington  
Tel: +64 4 463 9541  
E-mail: hedwig.eisenbarth@vuw.ac.nz  
ORCID: 0000-0002-0521-2630

50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 masks. We found that face coverings obscure emotion categorization by up  
2 to 74%, whereby emotion categories are affected differently by different cov-  
3 coverings, e.g. clear mouth coverings have little effect in categorizing happiness,  
4 but sadness is affected badly. While an overall accuracy of up to 97% has been  
5 achieved with nothing added to the face, the achieved accuracy decreases in all  
6 other cases when the face is obscured. Notably, clear visors have only a small  
7 effect across all emotions, where the classifier achieved an accuracy of up to  
8 89.0% compared to other types of facemasks in which the achieved accuracy  
9 is less than 36%.

11 **Keywords** CK+ · Emotion categorization · Facemasks · Facial expression ·  
12 Obscure faces · Sunglasses

## 15 1 Introduction

17 The categorization of human facial expression by artificial systems has re-  
18 ceived significant attention in recent years (Huang et al., 2019). One reason  
19 behind that is due to its diverse applications such as in human-computer in-  
20 teraction (Picard, 2003; Yuan and Ip, 2018; Brave and Nass, 2009; Arunnehr  
21 and Geetha, 2017).

23 Recently, coverings that obscure faces such as facemasks and sunglasses have  
24 become widely used for various purposes. This may have a great impact on  
25 the performance of artificial systems in emotion categorization.

26 The distinction between emotion recognition and emotion categorization is  
27 made here because we affirm that it is only an approximation to understand  
28 the underlying emotion of a person from their face (Shehu et al., 2020b).  
29 This is because emotion expression can vary across individuals, for instance,  
30 while scowling and crying are perceived as an expression of anger and sadness,  
31 certain people scowl when paying too much concentration to a task (Barrett  
32 et al., 2019; Boehner et al., 2007) and various people cry when they are happy.

33 Currently, hospitals and healthcare systems are motivated to introduce  
34 robots to help doctors in critical healthcare conditions (Shehu et al., 2020a),  
35 especially with the ongoing challenge of a global pandemic. However, these  
36 robots need to identify human emotions in order to interact with humans in  
37 an intuitive way.

38 Faces can be covered for different reasons e.g. facemasks are widely being  
39 used (Jones, 2020) to prevent the spread of infectious diseases such as tuber-  
40 culosis, coronavirus, swine flu, etc. Sunglasses are used to protect the eye from  
41 the sunlight, improve appearance, and obscure the face. However, while it is  
42 anticipated that different emotions will be affected differently by different face  
43 coverings, it is not known to what extent these face coverings can affect the  
44 ability of artificial systems to categorize emotion since the effect of obscured  
45 faces on emotion categorization systems has not been tested.

47 Recently, manufacturers have introduced a solution to solve the problem  
48 of using a fully covered facemask that reduces lip-reading opportunities by  
49 designing a facemask with a transparent window that allows the mouth to be  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

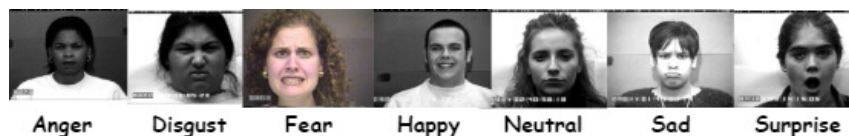
1 visible (Coleman, 2020). However, while the individual perception of a trans-  
2 parent mask might help lip reading and maybe also help emotion classification  
3 by humans, it is unknown if only adding the transparent window is enough  
4 for the artificial systems to categorize the emotion of humans as emotion is  
5 considered to be conveyed from the full face rather than only from the mouth  
6 or eyes (Coleman, 1949).  
7

8 The novelty of this research is to identify how wearing sunglasses or dif-  
9 ferent versions of facemasks affects emotion classification systems. In other  
10 words, the paper raises this important question: *To what extent can the per-*  
11 *formance of artificial emotion categorization systems be affected when people*  
12 *wear sunglasses or facemasks?*  
13

14 This question is addressed by implementing a strategy to manually add the  
15 sunglasses and facemasks to the images of the CK+ database. The research  
16 will analyze this effect by using a state-of-the-art deep learning classifier i.e.  
17 VGG19 model as an example of a deep learning algorithm (Simonyan and  
18 Zisserman, 2014). Contrary to the use case of a mask, either full or with a  
19 transparent window, this research also proposes the use of a clear (fully trans-  
20 parent) facemask, which has similar properties as a visor. The performance of  
21 an emotion classification system is compared for when manipulated images of  
22 people wearing this visor-like mask and people wearing sunglasses, fully and  
23 partially covered facemasks are included.  
24

25 The research uses the CK+ database because it is an example of emotion  
26 video-frame images and uses the last-half frames of each sequence in order to  
27 have more data to be used in the experiments.

28 The rest of the paper is organized as follows; Section 2 describes how  
29 certain researchers have used machine learning and deep learning algorithms  
30 to classify emotion. Section 3 explains the properties of the CK+ database,  
31 its emotion categories as well as how and why it is used in this research. This  
32 section also explains how computer vision techniques are used to create and  
33 add sunglasses and facemasks to emotion images. In addition, the section also  
34 provides an explanation of the properties of the deep model used, how it is  
35 set up, and why it is chosen to be used in this research. Section 4 presents the  
36 obtained results. Section 5 provides a further discussion of the findings and  
37 finally, Section 6 concludes the paper with hints at future studies.  
38  
39  
40



41  
42  
43  
44  
45  
46 **Fig. 1** Sample of six basic plus neutral expressions from the CK+ database (Shehu et al.,  
47 2020a). Note that the majority are grayscale images  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 2 Related Work

A number of methods can be used to categorize emotion from images. For instance, there is a single frame-based method that categorizes emotion from a single frame and a multi frame-based method that categorizes emotion from multiple frames.

The multi frame-based method with reference frames was used to analyze emotion from the CK+ dataset (Otroshi-Shahreza, 2017). Firstly, landmark coordinates of the face in each frame were derived using the dlib library (Dlib, 2017) followed by a normalization process. Vector movements from normalized coordinates of the landmark were calculated from the initial frame where the posed emotion is neutral to the last frame where the emotion is expressed at peak. The same was done for each of the six basic emotions (see Fig. 1), whereas the detected landmarks on initial frames were used for the neutral emotion. The calculated vector movements were used to determine the facial expression using Random Forests (RF), Decision Tree (DT), and Linear Discriminant Analysis (LDA) which lead to an accuracy of 93.47%, 89.29%, and 96.08% respectively. However, there are concerns that the proposed method is sensitive to the choice of landmarks as the use of dlib to automatically detect coordinates decreases the accuracy of the method compared to when landmarks are detected manually.

Majumder et al. (Majumder et al., 2018) proposed an automatic facial expression recognition system (AFERS) that used four different layers of a neural network to classify emotion using a deep network framework. The first two layers used geometric and appearance features for better representation of the facial expression. The third layer used Kohonen’s self-organizing maps (SOM)-based classifier. The SOM-based classifier used an improved learning algorithm and a soft-threshold logic as an improvement to give higher accuracy. Several experiments performed at the last layer to demonstrate the varying performances for a different number of nodes in the last layer. The performance of the proposed deep network had been tested on the CK+ (Lucey et al., 2010) and the MMI (Pantic et al., 2005) database, which lead to accuracy of 98.95% and 97.55%. The first two layers of the proposed AFERS system work by first detecting the faces in the image using the Viola-Jones (Viola and Jones, 2004) algorithm. However, as the Viola-Jones algorithm only detects faces on fully displayed frontal face images, it is anticipated that the method will not function on images displaying only part of the face. This will limit the usability of the method to only frontal face images.

Almowallad and Sanchez (Almowallad and Sanchez, 2020) proposed a deep learning framework for label-distribution learning (EDL-LBCNN) to classify emotion from images. The proposed method enhances features extracted by the convolution neural network by forming a local binary convolutional (LBC) layer to acquire texture information from face images so as to improve the generalization of the trained model. The proposed EDL-LBCNN was evaluated on the Japanese Female Facial Expression (JAFFE) (Lyons et al., 1999) database, the results have shown superior performance compared to the various state-

1 of-the-art label distribution-learning methods when the evaluation was made  
2 on straight facial images. However, the result was not very encouraging when  
3 the analysis was made on tilted face images.

4 Harshitha et al. (Harshitha et al., 2019) proposed a convolution neural net-  
5 work (CNN) architecture to classify emotion on the six basic (anger, disgust,  
6 fear, happy, sad, surprise) expressions of the JAFFE database. The proposed  
7 CNN was developed to have two convolution, pooling, fully connected layers,  
8 and a rectified linear unit (ReLU) activation function at each layer. The pro-  
9 posed approach has achieved an accuracy of 91.6% when tested with images  
10 from the JAFFE database. The proposed CNN has only two layers, as such,  
11 questions remain unanswered as to whether the method will perform better  
12 than the current state-of-the-art CNN architectures that were developed to  
13 have 50 (e.g. ResNet50(He et al., 2016)) or more layers.

14 Fathallah et al. (Fathallah et al., 2017) proposed an architecture based on a  
15 convolution neural network (CNN) to recognize facial expression from images.  
16 Initially, the proposed CNN is trained with fine-tuning by the Visual Geome-  
17 try Group (VGG) model to improve results. In the second step, training was  
18 repeated, however, fine-tuning was carried out with the obtained first model  
19 to obtain the final model. The performance of the method was determined af-  
20 ter evaluation on three state-of-the-art databases (CK+, MUG (Aifanti et al.,  
21 2010), and RaFD (Langner et al., 2010)), which lead to an accuracy of 99.33%,  
22 87.65%, and 93.33% for the CK+, MUG, and RaFD databases respectively.  
23 The developed model was trained with only fully displayed faces that are lo-  
24 cated at the center of the images. As such, it is anticipated that the method  
25 might perform badly when tested with tilted face images, faces located in dif-  
26 ferent regions of the image, or with obscure i.e. faces that are partially covered  
27 with external devices such as sunglasses or facemasks.

28  
29  
30  
31 Deep learning algorithms have been used to classify emotion from images  
32 (Majumder et al., 2018; Almowallad and Sanchez, 2020; Harshitha et al., 2019;  
33 Fathallah et al., 2017). However, emotion classification models developed based  
34 on a fully displayed facial configuration might perform badly in classifying the  
35 emotion of people from tilted face images or when a certain portion of the  
36 face is obscured e.g. by sunglasses or a facemask. This research is needed to  
37 analyze the effect of using systematic facemasks and sunglasses applied to an  
38 emotion dataset to test artificial emotion classification systems.  
39

## 40 41 42 **3 Methodology**

### 43 44 **3.1 Database**

45  
46 The extended Cohn-Kanade (CK+) (Lucey et al., 2010) database is an ex-  
47 tended version of the CK (Kanade et al., 2000) database. It contains posed  
48 facial expressions of 201 adults between the ages of 18-50 years. The image se-  
49 quence in the CK+ database varies from 10 to 60 frames starting from neutral  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 to peak expressions. This research uses the CK+ database to analyze emotion  
2 on six basic emotions defined by Ekman (Ekman and Friesen, 1971) as well as  
3 the neutral expression.  
4

5 We know from our previous work (Shehu et al., 2020b) that using the last-  
6 half frames of each sequence of the CK+ database gives a more accurate result  
7 than using only the last few frames where the emotion is expressed at peak.  
8 Consequently, the last-half frames of each sequence are assigned the emotion  
9 label of the sequence and the first-two frames of each sequence are assigned  
10 as neutral expression. For instance, we assigned neutral expression to frame  
11 one and two and happy expression to frames starting from 11 to 20 in a given  
12 sequence where the total number of frames is 20 and the decoded emotion is  
13 happy. Therefore, a total of 3,368 images are used.

14 Since deep learning algorithms require more data to be trained, the CK+  
15 dataset is chosen to be used to obtain more data for the experiment.  
16

### 17 *3.1.1 Train test split*

18 A total of 350 images, consisting of 50 images from each class were randomly  
19 selected as the test set and the remaining 3,018 images were used for training.  
20 The database is split to have the same number of images from each test class  
21 to avoid bias in the performance estimate.  
22  
23  
24

### 25 *3.1.2 Pre-processing*

26 Image pixels were converted to an array and normalization has been performed  
27 on the pixels of raw images to adjust values between the range of [0, 1] to enable  
28 fast computation. Labels were also converted to integers and one-hot encoded.  
29  
30  
31

## 32 *3.2 Deep learning model*

33 Deep learning models are a type of artificial neural network model that per-  
34 forms end-to-end learning. These algorithms are designed to recognize patterns  
35 in data based on an inspiration derived from neurons in the human brain.  
36 These algorithms use the layers of neural networks to extract higher-level in-  
37 formation at other layers based on the raw input data, which in this case is  
38 an image.  
39

40 In this research, VGG19 (Simonyan and Zisserman, 2014), which is an  
41 improved version of VGG16, is used as an example of a deep model to classify  
42 emotion labels from the CK+ database. VGG19 is chosen because it is a well  
43 tested standard model that achieved high performance in many studies (Ullah  
44 et al., 2019; Rassadin et al., 2017; Knyazev et al., 2017; Oloko-Oba and Viriri,  
45 2020; Dua et al., 2020). Another reason why the network is chosen is because  
46 it is deeper and has more weight layers compared to its pair (VGG16) (Zheng  
47 et al., 2018), which could lead to more flexible feature extraction.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1       The same VGG19 model with the same number of layers as proposed by  
2 the original paper is used from the Keras API (Keras API, 2021) with the  
3 following setup; the model is set to run over 200 epochs with a learning rate  
4 starting from 0.001 and reduced by 10% after 80, 120, and 160 epochs. The  
5 learning rate is set to reduce by only 5% after 180 epochs. 10% of the training  
6 data is reserved for validation and data augmentation is also used to improve  
7 the diversity of the data during training. The validation patience is set to five;  
8 meaning that training should stop if the loss on the validation set is larger or  
9 equal to the previously smallest loss for up to five consecutive times.

### 12 3.3 Dataset Modification

14       This procedure is followed to add an obscure artifact to a face, which in this  
15 case is sunglasses or a facemask: Initially, the foreground of the glasses or mask  
16 is placed on top of an overlay image. The overlay is blank and is resized to have  
17 the same size as the width and height of the input image. The alpha channel,  
18 which controls transparency in a given region is added and resized to have the  
19 same size as the input image. However, the alpha channel only contains the  
20 foreground mask. Alpha blending is performed to merge the alpha channel,  
21 foreground, and the background, which returns the output image.

#### 24 3.3.1 *Sunglasses*

26       Sunglasses are created to cover the two eyes of the participants in the image  
27 (see Fig. 2d). They are created by converting a particular glasses image sourced  
28 from the internet into a transparent mask (Rosebrock, 2018).

29       Algorithm 1 adds sunglasses to each test image of the CK+ database as  
30 well as generates a prediction for each image.

#### 33 3.3.2 *Facemasks*

35       Three different facemasks are applied to the images to analyze how different  
36 obscured faces affect the performance of an emotion classification system; one  
37 of which is a fully transparent visor type design, which has similar properties  
38 to a visor in terms of adding a small amount of visual noise to the image as in  
39 Fig. 2a. Another type has a transparent window that has an obscured noise,  
40 chin, and cheeks, but leaves the eyes, eyebrows, mouth, and forehead visible  
41 (see Fig. 2b). Finally, we included the common mask used by the individuals  
42 in multiple public settings which allows only the eyebrows, eyes, and forehead  
43 to be visible (see Fig. 2c).

#### 46 3.3.3 *Application*

48       This section provides a step-by-step explanation of how the artificial sunglasses  
49 and facemasks are applied to the face images.



**Algorithm 1** Procedure adopted to add sunglasses and generate prediction

---

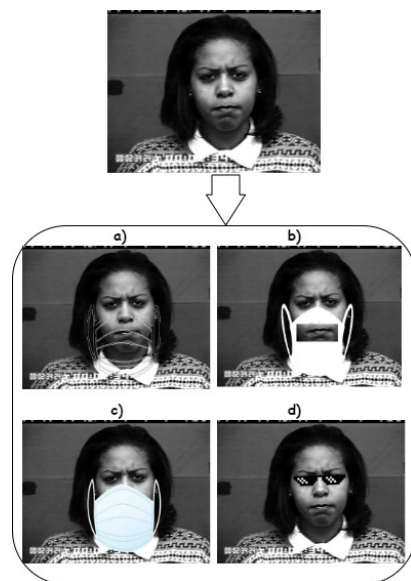
```

1: procedure SUNGLASSES
2:    $L_{t_i} \leftarrow$  list of test images
3:    $Prediction_{list} \leftarrow []$ 
4:   load CAFFE model, Haar cascade, VGG19 model # CAFFE and Haar cascade
5:   model are used to detect faces
6:   Begin
7:   for each test image  $i$  in  $L_{t_i}$  do
8:     flag = false
9:     repeat
10:      blob = blobFromImage(i) # construct blob from image
11:      detector = CAFFE(blob) # obtain detections from blob
12:      face = np.argmax(detector) # since we know that a particular images can
13:      have only one participant, we only get the detection with largest probability
14:      confidence = detector confidence # get confidence level from detector
15:      if confidence < 0.5 then # if confidence is not reliable
16:        flag = true
17:      end if
18:      if flag == true then
19:        detector = Cascade(i) # detect face using Haar cascade classifier
20:      end if
21:      detect  $(x,y)$  coordinates of facial landmarks
22:      compute bounding box from face
23:      construct rectangle from blob
24:      get landmarks of the left and right eye
25:      compute center of mass
26:      angle = compute eye centroid's angle
27:      sunglasses = rotate(sunglasses, angle) # rotate sunglasses based on calculated
28:      angle to align with tilted face
29:      sunglasses = sunglasses width * .9 # reduce the size of sunglasses since we
30:      do not want it to cover the whole face
31:      output_image = add(face, sunglass, mask) # function that adds the sunglasses
32:      to the image
33:      prediction = vgg19.predict(output) # predict the class of image using vgg19
34:      model
35:       $Prediction_{list}.append(prediction)$  # append prediction to prediction list
36:    until  $i == \text{len}(L_{t_i})$  # all images are processed
37:  end for
38: end procedure

```

---

The face is first detected using a pre-trained convolution architecture for fast feature embedding (*CAFFE*) (Jia et al., 2014), the *res10-300x300-ssd-iter-140000* model followed by constructing a dlib (Dlib, 2017) rectangle object, which is used to detect facial landmarks within faces. If the *CAFFE* model failed to detect a face in an image, the *Haar cascade* classifier (Viola and Jones, 2001) is used as an alternative to detect a face. The coordinates of the left and right eye are extracted from the detected landmarks followed by computation of the angle between the eye centroids and the centre of mass for each eye. Since certain faces of participants in the CK+ database are tilted (see Fig. 3), the computed angle is used to rotate the sunglasses to make sure that the sunglasses are aligned with how the face is tilted in the image. Furthermore, the width of the glasses is reduced to 90% to make sure that the glasses are not



**Fig. 2** Sample sunglasses and facemasks added to an anger image from the CK+ database. Note that the image in a), b), c), and d) represents the same image with a fully transparent visor, a facemask with a transparent window, a fully covered facemask, and sunglasses added.

covering the entire face before the glass is added to the face according to the detected landmarks of the left and right eye.

A similar procedure is performed to add a facemask onto the face except that we are more interested in where the chin is located in the detected face, rather than where eyes are located. The landmark of the jaw is initially extracted using dlib followed by extracting the landmark of the chin. In addition, the mask is also rotated according to how a particular face is tilted. Also, since the mask should cover the entire face, the mask is resized to have the same size as the width of the image before adding it to the face.



**Fig. 3** Sample of adapted images with tilted faces from the CK+ database.

## 4 Results

In the *Expressions* section of Table 1, 2, 3, 4, and 5, *An* represents anger, *Di* represents disgust, *Fe* represents fear, *Neu* represents neutral, *Ha* represents happy, *Sa* represents sadness and *Su* represents surprise expression. Also, \* refers to the average accuracy achieved from all categories of emotions. We chose to analyze emotion for the six basic plus neutral expression as they are considered to be universal expressions that can be understood by many people (Nummenmaa et al., 2007).

The VGG19 model is trained from scratch and evaluated on the validation set. As the model is not deterministic, the program is set to run 30 times and the performance of each run is assessed by testing the model on the test set. For that reason, results are provided in two different forms. \* represents average accuracy obtained from testing the best model on the test set, which can be visualized from the confusion matrix whereas \*\* represents the overall average accuracy obtained from 30 runs with upper and lower bound of a 95% confidence interval.

**Table 1** Confusion matrix obtained from testing original images

Expression	<b>An</b>	<b>Di</b>	<b>Fe</b>	<b>Ha</b>	<b>Neu</b>	<b>Sa</b>	<b>Su</b>
<b>An</b>	<b>98.0</b>	0	0	0	2.0	0	0
<b>Di</b>	0	<b>100</b>	0	0	0	0	0
<b>Fe</b>	0	0	<b>98.0</b>	0	2.0	0	0
<b>Ha</b>	0	0	0	<b>98.0</b>	2.0	0	0
<b>Neu</b>	0	2.0	2.0	0	<b>96.0</b>	0	0
<b>Sa</b>	0	0	0	0	0	<b>100</b>	0
<b>Su</b>	0	0	0	0	0	0	<b>100</b>

\* 98.57% accuracy is achieved from the best model

\*\* 97.15%±0.23 average accuracy is achieved from 30 runs

Table 1 shows the confusion matrix obtained from testing images of the CK+ database with no changes made to them. The model achieved an accuracy of up to 100% in three different classes (disgust, sad, and surprise). The model achieved a very high accuracy across all classes, with the lowest accuracy of 94.0% when the prediction is made for neutral class images.

Table 2 presents the confusion matrix obtained from predicting images of people with sunglasses. While the accuracy of up to 100% is achieved when the prediction is made for anger class images, the average accuracy achieved from the best model is 84.86%. This is because the model was not able to achieve a classification accuracy of more than 92% in all other classes except for the anger class compared to the minimum accuracy of 94% with nothing added to the image.

The sunglasses are placed to only cover the eyes, however, a drastic fall (e.g. from 100% to 80% in surprise class and from 100% to 72% in sad) in the

**Table 2** Confusion matrix obtained from testing images with sunglasses

Expressions	An	Di	Fe	Ha	Neu	Sa	Su
An	100	0	0	0	0	0	0
Di	10.0	80.0	4.0	0	6.0	0	0
Fe	8.0	4.0	86.0	0	2.0	0	0
Ha	2.0	4.0	0	92.0	2.0	0	0
Neu	12.0	4.0	0	0	84.0	0	0
Sa	16.0	0	0	0	12	72.0	0
Su	8.0	4.0	0	2.0	6.0	0	80.0

\* 84.86% accuracy is achieved from the best model

\*\* 66.69%±6.89 average accuracy is achieved from 30 runs

accuracy achieved by the majority of the class implies that the eyes provide salient information in categorizing emotion.

**Table 3** Confusion matrix obtained from testing images with a fully covered facemask

Expressions	An	Di	Fe	Ha	Neu	Sa	Su
An	84.0	2.0	0	0	14.0	0	0
Di	80.0	18.0	0	0	2.0	0	0
Fe	46.0	2.0	22.0	6.0	24.0	0	0
Ha	74.0	10.0	0	14.0	2.0	0	0
Neu	68.0	8.0	0	2.0	22.0	0	0
Sa	56.0	10.0	2.0	0	20.0	12.0	0
Su	34.0	4.0	2.0	4.0	42.0	2.0	12.0

\* 26.29% accuracy is achieved from the best model

\*\* 22.75%±0.86 average accuracy is achieved from 30 runs

Table 3 shows the confusion matrix obtained from predicting images of people wearing a fully covered facemask. Here, the model performs only a little better than random guessing (in this case 14.29% since we have seven classes) by achieving an average accuracy of 26.29% from the best model and an overall average accuracy of 22.75%±0.86 from 30 runs. Apart from the 84% accuracy achieved by the anger class, no other class achieves an accuracy of up to 23%.

There is certain evidence that the eyebrows, eyes, and forehead makes a very strong contribution towards the frowning of the face (Russell, 1994). As such, one reason why the anger class in Table 3 might have achieved a higher accuracy result compared to the other classes could be because the fully covered facemask still allows the eyebrows, eyes, and forehead to be visible.

Table 4 presents the confusion matrix obtained from testing images of people wearing a facemask designed with a transparent window. Interestingly, the model achieved an accuracy of 0% when the prediction is made for disgust and sad class images. This has a great impact on the average accuracy achieved by the best model (48.86%) compared to the accuracy achieved when nothing is added to the image.

**Table 4** Confusion matrix obtained from testing images of people wearing a mask with transparent window

Expressions	An	Di	Fe	Ha	Neu	Sa	Su
<b>An</b>	<b>58.0</b>	0	10.0	22.0	6.0	0	4.0
<b>Di</b>	46.0	<b>0</b>	6.0	22.0	24.0	0	2.0
<b>Fe</b>	16.0	2.0	<b>46.0</b>	18.0	8.0	0	10.0
<b>Ha</b>	2.0	0	0	<b>94.0</b>	4.0	0	0
<b>Neu</b>	24.0	0	6.0	22.0	<b>46.0</b>	0	2.0
<b>Sa</b>	12.0	0	20.0	6.0	44.0	<b>0</b>	18.0
<b>Su</b>	2.0	0	0	0	0	0	<b>98.0</b>

\* 48.86% accuracy is achieved from the best model

\*\* 35.37%±1.59 average accuracy is achieved from 30 runs

What is striking in this table is the accuracy achieved when the prediction is made for images of the surprise class. While only 12% of images from the surprise class are classified correctly when the prediction is made for images with a fully covered mask, the achieved accuracy is up to 98% (an increase of 86%) when the prediction is made for images of people wearing a mask with a transparent window. It can be inferred from this result that the mouth is making a very strong contribution in categorizing surprise expressions.

**Table 5** Confusion matrix obtained from testing images of people wearing a fully transparent visor

Expressions	An	Di	Fe	Ha	Neu	Sa	Su
<b>An</b>	<b>86.0</b>	0	0	0	14.0	0	0
<b>Di</b>	0	<b>98.0</b>	0	0	2.0	0	0
<b>Fe</b>	0	0	<b>98.0</b>	0	2.0	0	0
<b>Ha</b>	0	0	0	<b>96.0</b>	2.0	0	2.0
<b>Neu</b>	0	0	4.0	0	<b>94.0</b>	0	2.0
<b>Sa</b>	0	0	4.0	0	8.0	<b>88.0</b>	0
<b>Su</b>	0	0	0	0	0	0	<b>100</b>

\* 94.29% accuracy is achieved from the best model

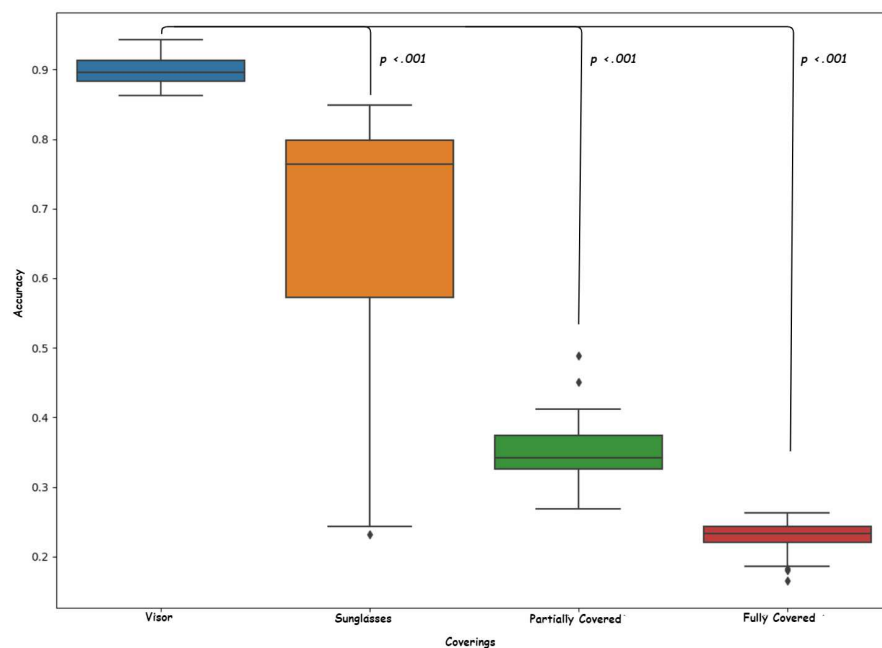
\*\* 89.95%±0.77 average accuracy is achieved from 30 runs

Table 5 shows the confusion matrix obtained from the prediction of images of people wearing a fully transparent visor. Apart from the anger and sad classes that achieved an accuracy of 86% and 88%, all other classes achieved an accuracy of more than 93%. An overall accuracy of up to 94% has been achieved from the prediction of images of people wearing a fully transparent visor.

There is a large difference between the accuracy achieved from predicting images with a fully transparent visor compared to when there are glasses on the face, a fully covered facemask, or a partially covered facemask with a transparent window is used.

A one-way ANOVA was conducted to compare these effects of face coverings on the achieved categorization accuracy for sunglasses, fully and partially

covered mask, as well as, the fully transparent mask conditions. It was found that face coverings have significant effect on the accuracy of emotion categorization systems at the  $\alpha < .05$  for the four conditions [ $F(3, 116) = 277.1, p < .001$ ]. Post hoc comparison of two-sample unpaired t-test with *Bonferroni correction* between three ( $t(58) = 114.2773, p < .001$  (Fully covered),  $t(58) = 60.6151, p < .001$  (Partially covered), and  $t(58) = 6.5776, p < .001$  (Sunglasses)) different groups at the  $\alpha = .017$  (see Fig. 4) all showed that there was a significant difference (Transparent Mask  $M = 89.95$ , Fully Covered Mask  $M = 22.75$ , Mask with Transparent Window  $M = 35.37$ , Sunglasses  $M = 66.69$ ) between the results obtained when the face is covered with a fully transparent mask compared to other types of face coverings.



**Fig. 4** Box plot showing accuracy achieved by each method and their significant interaction. Note that the significance interaction compares the accuracy achieved by the visor type covering to all other types of coverings

Also, ( $t(58) = 13.6831, p < .001$ ) showed that the achieved accuracy when wearing a mask with a transparent window is significantly (Mask with Transparent Window  $M = 35.37$ , Fully Covered Mask  $M = 22.75$ ) better than the achieved accuracy when wearing a fully covered facemask.

While a comparison of Table 3 with Table 4 reveals that the fully covered facemask deeply affects the performance of an emotion categorization system, a comparison of Table 5 with the result obtained from both Table 3 and 4 when a fully covered facemask and a facemask with a transparent window is

1 used reveals that having a fully covered facemask or a facemask with a trans-  
2 parent window on the face has a great impact in the accuracy of an emotion  
3 classification system. However, wearing a fully transparent visor compared  
4 with other coverings might help increase the achieved accuracy of the emotion  
5 classification systems.  
6

## 7 8 9 **5 Discussion**

10 Taken together, these results suggest that while the use of a facemask with a  
11 transparent window might help the artificial systems to categorize the emo-  
12 tional facial expression of people by achieving an overall accuracy that is al-  
13 most twice the accuracy achieved when the prediction is made for images of  
14 people with a fully covered facemask, the achieved accuracy is still not very  
15 promising ( $< 50\%$ ).  
16

17 Conversely, an overall average accuracy of up to 89% achieved from pre-  
18 dicting images of people wearing a fully transparent visor suggests that the  
19 artificial systems have a better chance of categorizing people's emotional facial  
20 expressions correctly while wearing a fully transparent visor than a partially  
21 or a fully covered facemask. In addition, two-sample unpaired t-tests showed  
22 that the overall accuracy achieved when wearing a fully transparent visor is  
23 significantly better than the achieved accuracy when wearing other types of  
24 coverings.  
25

26 Hence, the use of a fully transparent visor is beneficial if those are con-  
27 sidered as protective, as it will not only provide protection against infectious  
28 disease but will at the same time help in improving the interaction between  
29 artificial systems and humans since people like to interact with social robots  
30 which can identify emotion (Breazeal et al., 2008; Onyeulo and Gandhi, 2020)  
31 than with robots that cannot identify emotions.  
32

33 While it is understandable to see a decrease in the accuracy of all or a  
34 particular class when sunglasses or facemasks are added to the images, the  
35 accuracy achieved from predicting images of the anger class after sunglasses  
36 have been added to the image is somewhat counter-intuitive. The accuracy  
37 increases from 98% with nothing added to the image to 100% after sunglasses  
38 have been added to the image. This possibly occurred due to the stochastic  
39 nature of processes while training the deep learning algorithm or noisy features  
40 where covered by the sunglasses.  
41

42 It is also worth mentioning that the aim of this research is to utilize artificial  
43 intelligence technique to analyze changes in the performance of an emotion  
44 categorization model when the face is covered with and without sunglasses  
45 and facemasks. The research does not aim to develop a new technique to  
46 improve the performance of an emotion categorization model from obscured  
47 faces. VGG19 has been chosen to be used as an example of a deep model  
48 as it is one of the most commonly used network nowadays, which has been  
49 extensively tested.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 6 Conclusion

To the best of our knowledge, this is the first study that analyzes emotion categorization from faces of people wearing simulated sunglasses and facemasks. This paper analyzes emotions from the faces of people wearing facemasks and sunglasses. The research adds sunglasses as well as different kinds of masks to the images of the CK+ databases. The work analyzes emotion from faces of people wearing these sunglasses and facemask, comparing the performance of an emotion categorization system on faces of people wearing sunglasses, a fully covered facemask, a transparent facemask with a transparent window, and a fully transparent visor. The achieved accuracy on the fully transparent visor is relatively larger compared with the accuracy achieved when glasses, a fully covered facemask, or a facemask with a transparent window is used.

Before this study, evidence that the use of a fully covered facemask obscure communication was purely anecdotal. However, after a comprehensive investigation that was made in the study, we can now conclude that not only using the fully covered facemask obscure communication but also the newly produced partially covered facemask with transparent window also affect the performance of an artificial emotion categorization system by a significant amount.

In addition, the empirical findings in this study provide evidence that the fully transparent facemask, which has a similar property to a visor is the easiest to understand of all coverings by the artificial emotion categorization system (see Section 4) as the achieved accuracy when wearing the fully transparent visor is significantly better than the accuracy achieved with all other types of coverings.

Despite the promising result obtained when emotion is analyzed from faces of people wearing a fully transparent visor, questions remain unanswered as to whether adding the fully transparent visor on images from a different database will also lead to a higher accuracy result. In addition, it is unknown as to how human classifiers will perform when categorizing these images. As such, future work should analyze the performance of human classifiers in categorizing these data as well as apply the same approach to a different dataset.

## 7 Declarations

### Funding

This research was not funded by any organization.

### Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Availability of data and material**

Not applicable

**Code availability**

Code will be sent on request to the corresponding author. Researchers wishing to use our code will agree with the terms that they will cite our work.

**Authors' contributions**

Harisu Abdullahi Shehu: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization, Project administration Will Browne: Supervision, Writing – Review & Editing Hedwig Eisenbarth: Supervision, Writing – Review & Editing

**Ethics approval**

Not applicable

**Consent to participate**

Not applicable

**Consent for publication**

All authors have approved the manuscript and agree with submission to the Neural Computing & Applications

**References**

- Aifanti, N., Papachristou, C., and Delopoulos, A. (2010). The mug facial expression database. *11th IEEE International Workshop on. Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4.
- Almowallad, A. and Sanchez, V. (2020). Human emotion distribution learning from face images using cnn and lbc features. *International Workshop on Biometrics and Forensics (IWBF)*, page 1–6.
- Arunnehr, J. and Geetha, M. (2017). Automatic human emotion recognition in surveillance video. *Intelligent Techniques in Signal Processing for Multimedia Security*, pages 321–342.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68.
- Boehner, K., DePaula, R., Dourish, P., and Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4):275–291.
- Brave, S. and Nass, C. (2009). Emotion in human-computer interaction. *Human-computer interaction fundamentals*, 20094635:53–68.

- 1 Breazeal, C., Takanishi, A., and Kobayashi, T. (2008). Social robots that  
2 interact with people. In: *Siciliano B., Khatib O. (eds) Springer Handbook*  
3 *of Robotics*.
- 4 Coleman, J. C. (1949). Facial expressions of emotion. *Psychological Mono-*  
5 *graphs: General and Applied*, 63(1):i.
- 6 Coleman, T. J. (2020). Coronavirus: Call for clear face masks to be ‘the norm’.  
7 *BBC News*. Accessed: 26 May, 2020.
- 8 Dlib (2017). Frame-based face emotion recognition using linear discriminant  
9 analysis. *Proceedings - 3rd Iranian Conference on Signal Processing and*  
10 *Intelligent Systems (ICSPIS)*, pages 141–146.
- 11 Dua, M., Singla, R., Raj, S., Jangra, A., et al. (2020). Deep cnn models-based  
12 ensemble approach to driver drowsiness detection. *Neural Computing and*  
13 *Applications*, pages 1–14.
- 14 Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face  
15 and emotion. *Journal of personality and social psychology*, 7(2):124.
- 16 Fathallah, A., Abdi, L., and Douik, A. (2017). Facial expression recognition  
17 via deep learning. *IEEE/ACS 14th International Conference on Computer*  
18 *Systems and Applications (AICCSA)*, pages 745–750.
- 19 Harshitha, S., Sangeetha, N., Shirly, A. P., and Abraham, C. D. (2019). Hu-  
20 man facial expression recognition using deep learning technique. *Inter-*  
21 *national Conference on Signal Processing and Communication (ICSPPC)*,  
22 pages 339–342.
- 23 He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image  
24 recognition. *IEEE conference on computer vision and pattern recognition*,  
25 pages 770–778.
- 26 Huang, Y., Chen, F., Lv, S., and Wang, X. (2019). Facial expression recogni-  
27 tion: A survey. *Symmetry*, 11(10):1189.
- 28 Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., and  
29 Darrell, T. (2014). Caffe: Convolutional architecture for fast feature em-  
30 bedding. *22nd ACM international conference on Multimedia*, pages 675–  
31 678.
- 32 Jones, R. I. (2020). How to stop your glasses steaming up with a face mask  
33 on. *BBC News*. Accessed: 14 July, 2020.
- 34 Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for  
35 facial expression analysis. *IEEE International Conference on Automatic*  
36 *Face and Gesture Recognition, (FG 2000)*, page 46–53.
- 37 Keras API (2021). Layer activation functions.
- 38 Knyazev, B., Shvetsov, R., Efremova, N., and Kuharenko, A. (2017). Convo-  
39 lutional neural networks pretrained on large face recognition datasets for  
40 emotion classification from video. *arXiv preprint arXiv:1711.04598*.
- 41 Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., and Van Knip-  
42 penberg, A. (2010). Presentation and validation of the radboud faces  
43 database. *Cognition & Emotion*, 24(8):1377–1388.
- 44 Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews,  
45 I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset  
46 for action unit and emotion-specified expression. *IEEE Computer Society*  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 1            *Conference on Computer Vision and Pattern Recognition - Workshops,*  
2            *(CVPRW)*, page 94–101.
- 3 Lyons, M. J., Budynek, J., and Akamatsu, S. (1999). Automatic classifica-  
4            tion of single facial images. *IEEE Transactions on Pattern Analysis and*  
5            *Machine Intelligence*, 21(12):1357–1362.
- 6 Majumder, A., Behera, L., and Subramanian, V. K. (2018). Automatic fa-  
7            cial expression recognition system using deep network-based data fusion.  
8            *IEEE Transactions on Cybernetics*, 48(1):103–114.
- 9 Nummenmaa, L., Saarimäki, H., Glereana, E., Gotsopoulos, A.,  
10            Jääskeläinen, I. P., Harib, R., Schwartz, H. A., and Ekman, P. (2007).  
11            Emotions revealed. recognizing faces and feelings to improve communica-  
12            tion and emotional life. *Neurobiology Essentials for Clinicians*, (1).
- 13 Oloko-Oba, M. and Viriri, S. (2020). Pre-trained convolutional neural network  
14            for the diagnosis of tuberculosis. In *International Symposium on Visual*  
15            *Computing*, pages 558–569. Springer.
- 16 Onyeulo, E. B. and Gandhi, V. (2020). What makes a social robot good  
17            at interacting with humans? In: *Siciliano B., Khatib O. (eds) Springer*  
18            *Handbook of Robotics*, 11(1):43.
- 19 Otroshi-Shahreza, H. (2017). Frame-based face emotion recognition using lin-  
20            ear discriminant analysis. *Proceedings - 3rd Iranian Conference on Signal*  
21            *Processing and Intelligent Systems (ICSPIS)*, pages 141–146.
- 22 Pantic, M., Valstar, M. F., Rademaker, R., and Maat, L. (2005). Web-based  
23            database for facial expression analysis. *IEEE Int. Conf. Multimedia Expo*  
24            *(ICME)*, page 317–321.
- 25 Picard, R. W. (2003). Affective computing: challenges. *International Journal*  
26            *of Human-Computer Studies*, 59(1-2):55–64.
- 27 Rassadin, A., Gruzdev, A., and Savchenko, A. (2017). Group-level emotion  
28            recognition using transfer learning from face identification. *19th ACM*  
29            *International Conference on Multimodal Interaction*, pages 544–548.
- 30 Rosebrock, A. (2018). Creating gifs with opencv. *Sunglasses*. Accessed: 5  
31            November 2018.
- 32 Russell, J. A. (1994). Is there universal recognition of emotion from facial  
33            expression? a review of the cross-cultural studies. *Psychological Bulletin*,  
34            115(1):102–141.
- 35 Shehu, H. A., Browne, W. N., and Eisenbarth, H. (2020a). An adversarial  
36            attacks resistance-based approach to emotion recognition from images  
37            using facial landmarks. *IEEE International Conference on Robot and*  
38            *Human Interactive Communication*.
- 39 Shehu, H. A., Browne, W. N., and Eisenbarth, H. (2020b). Emotion categoriza-  
40            tion from video-frame images using a novel sequential voting technique.  
41            *International Symposium on Visual Computing*.
- 42 Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for  
43            large-scale image recognition. *ICLR, arXiv preprint arXiv:1409.1556*.
- 44 Ullah, A., Jami, A., Aziz, M. W., Naeem, F., Ahmad, S., Anwar, M. S., and  
45            Jing, W. (2019). Deep facial expression recognition of facial variations  
46            using fusion of feature extraction with classification in end-to-end model.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 1                    *International Conference on Emerging Trends in Engineering, Sciences*  
2                    *and Technology (ICEEST)*, pages 1–6.
- 3                    Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cas-  
4                    cade of simple features. *IEEE Computer Society Conference on Computer*  
5                    *Vision and Pattern Recognition*, (1).
- 6                    Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *Internat-*  
7                    *ional Journal of Computer Vision*, 57(2):137–154.
- 8                    Yuan, S. and Ip, H. (2018). Using virtual reality to train emotional and so-  
9                    cial skills in children with autism spectrum disorder. *London journal of*  
10                    *primary care*, 10(4):110–112.
- 11                    Zheng, Y., Yang, C., and Merkulov, A. (2018). Breast cancer screening using  
12                    convolutional neural network and follow-up digital mammography. *Inter-*  
13                    *national Society for Optics and Photonics, Computational Imaging III*,  
14                    (10669):1066905.
- 15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

# Figures

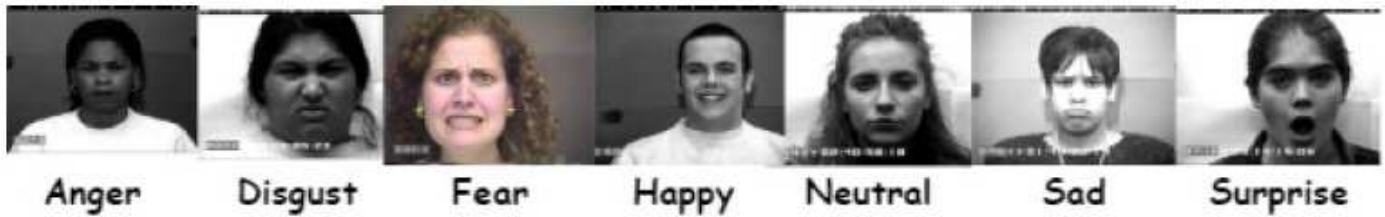


Figure 1

Sample of six basic plus neutral expressions from the CK+ database (Shehu et al., 2020a). Note that the majority are grayscale images

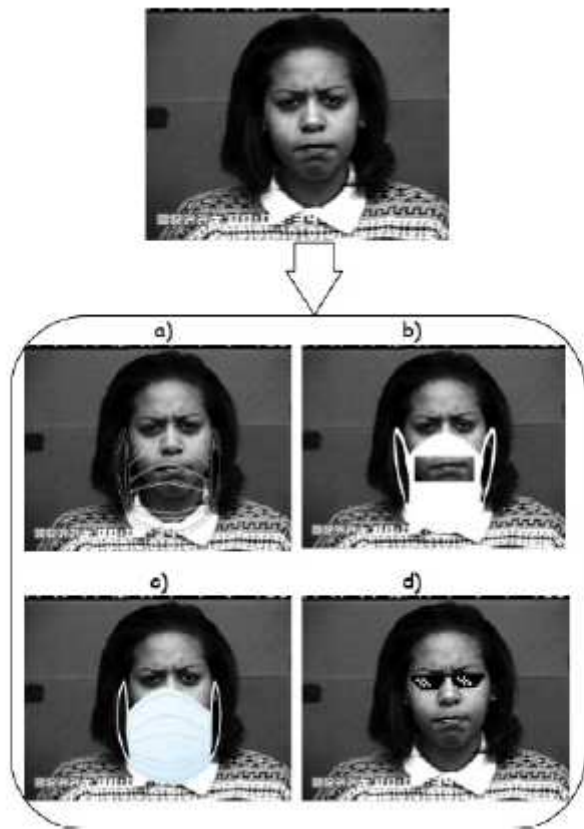


Figure 2

Sample sunglasses and facemasks added to an anger image from the CK+ database. Note that the image in a), b), c), and d) represents the same image with a fully transparent visor, a facemask with a transparent window, a fully covered facemask, and sunglasses added.



Figure 3

Sample of adapted images with tilted faces from the CK+ database.

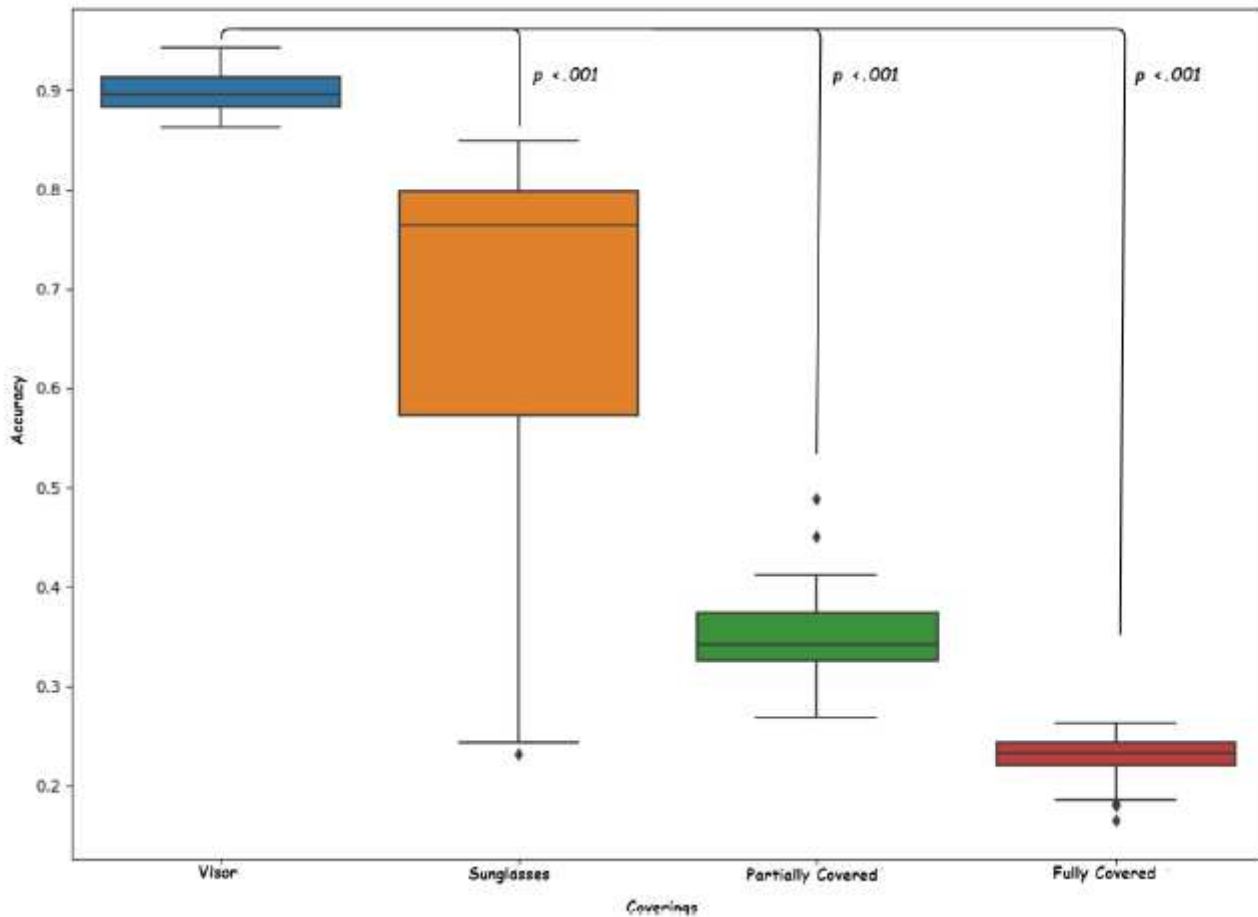


Figure 4

Box plot showing accuracy achieved by each method and their significant interaction. Note that the significance interaction compares the accuracy achieved by the visor type covering to all other types of

coverings