Emotion Detection of People Wearing Face Mask

Naga Sai Tejo Pratardanudu Pedaprolu
GRIET, Hyderabad, Telangana
tejopratardanudu@gmail.com

Y. Jayanth[2]
GRIET, Hyderabad, Telangana
jayanth.yerrapatruni@gmail.com

S. Madhur[3]
GRIET, Hyderabad, Telangana
madhurmahi4@gmail.com

D. Sai Sumanth[4]
GRIET, Hyderabad, Telangana
d.sumanth4u@gmail.com

G. Sri Vishruth[5]
GRIET, Hyderabad, Telangana
gvishruth60@gmail.com

Dr. P. Chandra Sekhar Reddy[6]
GRIET, Hyderabad, Telangana
pchandra1369@grietcollege.com

2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-9199-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCDS56580.2023.10104665

Abstract-Though many facial emotion recognition models exist, after the Covid-19 pandemic, majority of such algorithms are rendered obsolete as everybody is compelled to wear a facemask to protect themselves against the deadly virus. Face masks can hinder emotion recognition systems, as crucial facial features are not visible in the image. This is because facemasks cover essential parts of the face such as the mouth, nose, and cheeks which play an important role in differentiating between various emotions. This study intends to recognize the emotional states of anger-disgust, neutral, surprise-fear, joy, sadness, of the person in the image with a face mask. In the proposed method, a CNN model is trained using images of people wearing masks. To achieve higher accuracy, the classes in the dataset are combined. Different combinations of clubbing are performed, and results are recorded. Images are taken from FER2013 dataset which consists of a huge number of manually annotated facial images of people.

Keywords—Facial Expressions, Emotions, Machine Learning, Computer Vision, Face Masks

1. Introduction

Emotion recognition is a technology that aims to identify and interpret human emotions through various means, such as facial expressions, body language, and speech patterns. With the widespread use of face masks in the wake of the COVID-19 pandemic, traditional methods of emotion recognition have become more difficult. This has led to the development of new techniques specifically designed for individuals wearing face masks.

One approach is to use deep learning techniques to train a model on a dataset of images of individuals wearing masks and expressing different emotions. The model can then be used to predict the emotion of an individual in a new image by analyzing their facial features. This approach has shown promising results in laboratory settings, but there is still a need for more real-world testing.

2. Literature Survey

The covering of faces by masks, caps, and sunglasses are a key limitation to the face recognition methods in the real world. The term to describe the covering of the faces through these means can be termed as occlusion which is observed due to the objects that cover a part of the face.

Ref. paper [1] presents a tool for emotion recognition that compensates for the difficulty in identifying emotions when the subject is wearing a face mask. The tool applies an iterative training strategy based on convolutional neural network architectures. The architecture is trained to recognize five sub-classes of emotions. This implementation suffers from artificial setting and lack of practical applications. In Ref. paper [2] a machine learning project aimed at detecting facial emotions and face masks. The project uses deep learning neural networks on images to distinguish different emotions, both with and without masks. The goal of the model is to detect whether a person is wearing a mask and if so, to detect the mask, and if not, to detect the emotion in the face. However, this method lacks evaluation and has a limited dataset. Ref paper [3] presents a method for improving the effectiveness of facial expression recognition technology. The proposed method uses differently shaped masks for different facial orientations. The results show that the training of FER models on a simulated masked FER dataset is feasible. This method too suffers from problems similar to those of the previous method and also from a lack of generalizability. The study in Ref paper [4] aims at exploring the impact of wearing masks on social interaction and emotional moods. The system uses Haar feature-based cascade classifiers [11]-[12], to detect universal emotions on faces with and without masks. The drawbacks of this method include limited indexing, limited scope and limited international recognition.

3. Methodology

3.1. Proposed Method

The overall objective of the paper is to identify as accurately as possible, the emotion of the person in the facial image while the person is wearing a mask. The proposed method consists of training a machine learning model using images of people wearing face masks.

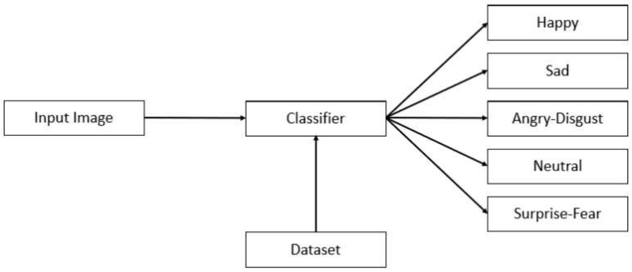


Fig. 1 Flow diagram of proposed method

3.2. Pre-Processing

Pre-processing included cleansing the dataset. Raw dataset had images which are blurred, and some images are in an irregular placement of the face. Dataset cleaning is achieved by removal of these images and manual rectification of incorrectly classified images. An additional step in pre-processing is the super-imposition of an image of a facemask onto each image in the dataset. This is achieved using a simple Python script.

3.3. Description of Classes in Dataset

Dataset consists of around 30000 images divided into seven classes namely: angry, sad, disgust, surprise, happy, fear, and neutral.

Anger	Anger is a response to strong feeling of non-co-operation and uneasiness.
Fear	It is a feeling of panic and terror.
Disgust	It is a feeling in response to aversion.
Surprise	It is a feeling in response to sudden unexpected and positive events.
Sad	It is emotion characterized by feeling of loss, helplessness, and despair.
Happy	It is an emotion in response to a state of well-being and elation.
Neutral	It is an emotion where there is no expression in the face. It is also known as “poker face.”

Table 1. Description of classes

The size of each image is 48*48.Since most of the prominent features of the face are getting covered due to masks, it is extremely difficult to predict all the seven emotions accurately. To resolve this issue, we decided to club some of the emotions as a single class. Fear and surprise are similar emotions, as both the emotions show similar changes in face while expressing them. Eyes pop out in both cases. Similarly, anger and disgust are similar in movement of face while expressing. Mouth looks very similar in both the emotions and eyebrows are squeezed together. Based on these facts, anger and disgust are clubbed together and fear and surprise are clubbed together. This makes a total of five classes which are anger-disgust, fear-surprise, happy, sad, neutral. As there are no datasets for emotion detection of people wearing masks readily available, we convert this dataset into a dataset of people wearing masks by using a simple Python program that overlays a facemask image onto each image in the FER-2013 dataset.

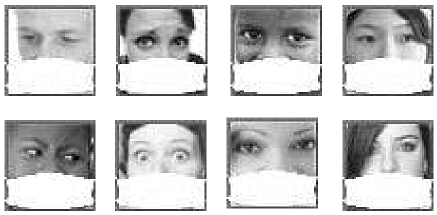


Fig. 2 Sample images in dataset

4. Model

The Convolutional Neural Network (CNN) is a powerful machine learning technique designed to process structured data arrangements. With its ability to identify patterns, shapes, and gradients in input images, CNN has become a popular choice for computer vision tasks. This makes it ideal for recognizing faces and other visual features in images. The accuracy of CNN in recognizing and classifying objects makes it a highly useful tool in the field of computer vision.

The Convolutional Neural Network (CNN) is an advanced machine learning model that can handle raw image data without any pre-processing. This is made possible by the convolutional layers that form the core of the network. These layers are designed to identify patterns and shapes in the image data, and by stacking multiple convolutional layers on top of each other, CNNs are able to analyze more and more complex features. This ability to handle raw data and identify complex features gives the CNN its robustness, making it a valuable tool for a variety of computer vision tasks.

By having 3 to 4 convolutional layers the model can identify even the digits that are handwritten and by having the other twenty-five layers it will be able to classify the facial images of human beings. The docket of the sphere will be activating the machines to get a view on the world as we human are able to view and can be used in a wide range of applications. Its ability to process raw image data and recognize complex features makes it an effective tool for tasks such as object detection, edge recognition, image inspection, image classification, consumer recommendation systems, and even natural language processing (NLP). This versatility allows CNNs to be used in multiple areas and provides a flexible solution for various computational challenges.

4.1. Convolutional Neural Network (CNN) Design:

The Convolutional Neural Network (CNN) is a type of forward-feeding neural network that is composed of multiple layers. These layers are organized in a specific sequence and placed one on top of the other to form the architecture of the CNN.

This is a sequential design which will grant access to CNN(Convolutional Neural Network) for analysing the hierarchical attributes. In CNN, few of these layers are then formed by assembling the layers and the hidden layers will be convolutional layers which are then supervened by the activation layers.

The Convolutional Neural Network (CovNet) relies on pre-processing in order to function effectively. This pre-processing step is similar in nature to the organization of neurons in the human brain, particularly the visual cortex. By drawing inspiration from the structure of the human brain, the CovNet is able to process image data and make predictions in a similar manner. The pre-processing step is a crucial component of the CovNet and helps to ensure its accuracy and effectiveness.

4.2. Layers in CNN:

4.2.1. Conv2D layer:

A convolutional layer is a fundamental aspect of the architecture of a Convolutional Neural Network (CNN) that extracts features from input data. This layer involves a combination of linear and non-linear operations, such as the convolution operation and an activation function. The layer is defined by several parameters, including the kernel size, padding, strides, input shape, and activation function. For instance, a common configuration sets the kernel size to 2, the strides to either (5,5) or (3,3), the input shape to (48,48,1) for all layers, and the activation function to the Rectified Linear Unit (ReLU). The convolutional layer plays a crucial role in the feature extraction process, as it processes the input data and extracts meaningful features that are used in subsequent layers for prediction.

4.2.2. MaxPooling2D layer :

Max pooling is a widely used pooling operation in Convolutional Neural Networks (CNNs) that helps to extract features from the feature map. This operation involves selecting a sub-grid from the feature map and returning the maximum value in each sub-grid while discarding all other values. The commonly used filter size for max pooling is 2x2 with a stride of 2, which leads to a decrease in the height and width dimensions of the image by half, while the depth dimension remains unchanged. The parameters used for this layer include pool_size and strides, with different pool sizes such as (5,5) and (3,3) for various layers, and a standard stride value of (2,2).

4.1.3. Dense layer:

A dense layer in neural network architecture consists of multiple neurons, where each neuron is connected to all the neurons in the preceding layer and receives input from all of them. This layer is simple and provides a dense connection of neurons.Dense layer is used to classify data from the convolutional layer. Arguments include units and activation. Dense layer is part of the fully connected part of the model. Values used for units are 1024 for inner layer and seven for output layer, to get seven different classes/labels. All the dense layers except output layer have an activation ‘relu’ whereas the output layer has the activation ‘softmax.’

4.2.4. Flatten layer:

A flatter layer converts data of multiple dimensions to data of a single dimensional vector. The output from a flatten layer can be used as input for a dense layer. The last layer connects all the pixels from previous layer to seven different classes, each corresponding to an emotion. Flatten layer is not provided with any arguments. Flatten layer is part of the fully connected part of the model.

4.2.5. Dropout layer:

The Dropout layer randomly sets input units to zero during each step of training. The number of input units made zero depends on the dropout rate of the dropout layer. This prevents overfitting of the model as large amount of training data may be used in training the model. Dropout layer only has one argument called the dropout rate. This value specifies the percentage of nodes to be marked null out of all the weights from the previous layer. This value ranges from 0 to 1, zero meaning that no weight is dropped and one meaning all the weights are dropped. Dropout rate is 0.2 for most applications.

5. Experimental Results and Discussion

For the purpose of creating, training, and testing the emotion detection model, we have used a system with a Windows operating system, powered by AMD Ryzen 5 5500u processor. According to the workflow of the proposed method, a CNN model is designed to classify images into 5 different classes. The available 7 classes in the dataset are clubbed together in various combinations such as a in a trial-and-error method. The combinations of clubbing together angry and disgust, and surprise and fear provided us with the highest accuracy among all the combinations we have tried. This CNN architecture model helped us achieve an accuracy of 94.31%. The CNN model itself was arrived at after a lot of experimenting. We have tried different methods such as using ResNet, LBP, VGG-16 and VGG-19. However, a simple CNN architecture without the use of any of the above has given us the maximum accuracy.

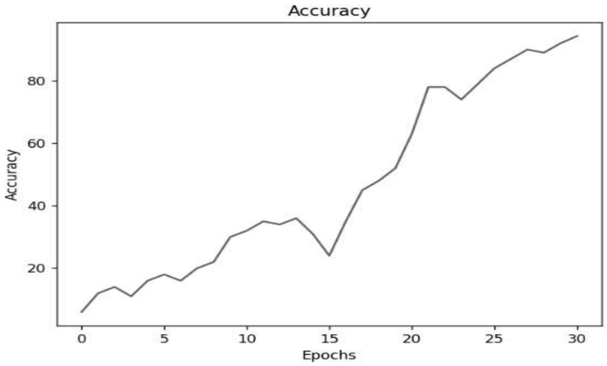
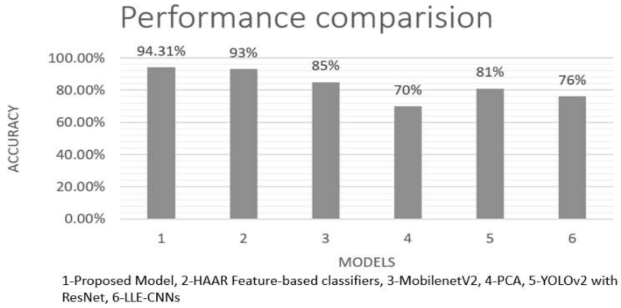


Fig. 3 Accuracy Graph



6. Proposed method accuracy with other existing method0s

Sl. No.	References	Dataset	Accuracy
1.	Proposed method	FER	94.31%
2.	Harr-Feature based clasifiers[3]	FER	93%
3.	MobileNe V2[2]	Real time data	85%
4.	PCA[8]	ImageNet	70%
5.	YOLO V2 with ResNet[9]	MMD	81%
6.	LLE- CNNs[10]	Real time data	76%

Table 2. Performance Comparison

7. Conclusion

In conclusion, the ability to detect facial emotions of individuals wearing masks is a challenging task due to the occlusion of key facial features such as the mouth and nose. But, in times where wearing a mask has become the new normal, it is imperative that machines learn to collaborate with subjects who wear masks. The field is still in early stages of development and more research is needed to improve the robustness of the models and account for variations in mask type and individual face shape. The problem is further compounded by the fact that there are no datasets for emotion detection of subjects wearing masks.

The main issue addressedisthe similarity in the images of different emotions. Upon close inspection, the study concludes that emotions very similar to each other need to be put in a singular label. After restructuring the dataset to accommodate these similarities by reducing the number of emotions from seven to five, a significant increase in accuracy is observed and accuracy of 94.31% is recorded. This proves that deep learning techniques have potential for high accuracy in detecting emotions even when faces are partially obscured. Convolutional Neural Networks trained on large datasets of masked facial images have the potential to make models better.

Further study on modifying images into different versions of themselves before being passed as input to the model may achieve higher accuracy. Similar methods can also be developed in situations where the face is occluded by other objects such as sunglasses, hats, etc. The application that was built using the model could identify the five emotions in real time along with the higher accuracy further emphasizes the effectiveness of the model over the other methods.

8. References

[1] Magherini, R., Mussi, E., Servi, M. and Volpe, Y., 2022. Emotion recognition in the times of COVID19: Coping with face masks. Intelligent Systems with Applications, 15, p.200094.

[2] Vishnu Dinesh, Arun Prakash, & Amal Dasan (2021). FACIAL EMOTION AND FACE MASK DETECTION. International Journal of Innovations in Engineering Research and Technology, 8(07), 190–195.

[3] Yang, B., Wu, J., & Hattori, G. (2020). Facial Expression Recognition with the advent of face masks. Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia.

[4] Nawal Y. Abdullah Ahmed M Alkababji Masked face with facial expression recognition based on deep learning, Indonesian Journal of Electrical Engineering and Computer Science 27(1):149-155.

[5] Castellano, G., De Carolis, B. & Macchiarulo, N. Automatic facial emotion recognition at the COVID-19 pandemic time. Multimed Tools Appl (2022).

[6] Wang, Bingshu, Zheng, Jiangbin, Chen, C.(2021) A Survey on Masked Facial Detection Methods and Datasets for Fighting Against COVID-19, Journal of IEEE Transactions On Artificial Intelligence, Vol. 00, No. 0, December 2021

[7] Alturki R, Alharbi M, AlAnzi F, Albahli S. Deep learning techniques for detecting and recognizing face masks: A survey. Front Public Health. 2022 Sep 26.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009, IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2010, pp. 248–255.

[9] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection,” Sustainable Cities and Society, vol. 65, p. 102600, Feb. 2021.

[10] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, “Face recognition based on convolutional neural network,” in Proceedings of the International Conference on Modern Electrical and Energy Systems, Nov. 2017, vol. 2018-January, pp. 376–379.

[11] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, vol. 1, pp. I-511-I–518.

[12] B. Shetty, Bhoomika, Deeksha, J. Rebeiro, and Ramyashree, “Facial recognition using Haar cascade and LBP classifiers,” Global Transitions Proceedings, vol. 2, no. 2, pp. 330–335, Nov. 2021.