Facial Emotion detection using VGG19

Heppil Kheni
Computer Eng. Dept.
CHARUSAT University
Gujarat , India
21ce056@charusat.edu.in

Kirtan Matalia
Computer Eng. Dept.
CHARUSAT University
Gujarat , India
21ce070@charusat.edu.in

Trusha Patel
Computer Eng. Dept.
CHARUSAT University
Gujarat , India

Rikita Chokshi
Computer Eng. Dept.
CHARUSAT University
Gujarat , India

Abstract:

In this research, we're diving into the world of recognizing emotions on people's faces using a fancy tool called VGG19, a smart computer model. We're teaming up insights from computer science, neuroscience, and psychology to see how we can make this emotion detection thing even better. We're checking out what VGG19 can do to improve how accurately we can tell what someone's feeling. we've made this dataset with 758 images for training and 149 images for testing. We're also looking at what other researchers have tried before, like using fancy terms such as Histogram of Oriented Gradient (HOG) descriptors and deep learning stuff. Our goal? To make emotion detection systems that work really well and can be useful in everyday stuff like talking to computers or even helping out in therapy sessions. so we're ready to dive deep into this exploration!

1. Introduction:

Facial emotion detection, a crucial aspect of human communication, has garnered substantial interest in both academic and industrial domains. this area has witnessed significant growth, fueled by advancements in artificial intelligence and computer vision technologies. As highlighted by vario us surveys [1,2], nonverbal components, particularly facial expressions, convey a substantial portion of human communication, emphasizing the importance of facial emotion recognition (FER) in interpersonal communication.

With the exponential growth of data and computing power, computer vision has emerged as a prominent field of study. Researchers are increasingly drawn towards unraveling insights into human cognitive processes, prompting interdisciplinary studies encompassing computer science, neuroscience, and psychology. The analysis of facial expressions, along with face detection and recognition, constitutes a vibrant research area within computer vision [3].

Face detection, a fundamental component in FER systems, involves the identification of facial features within images or videos, facilitating subsequent tasks such as face recognition and emotion analysis. Notably, advancements in face recognition technology, initially introduced in the 1960s and continuously refined since then, have led to its widespread adoption in various domains, including security, forensics, and user authentication [4].

Emotion recognition technology has gained considerable traction in recent decades, owing to advancements in artificial intelligence techniques. It encompasses various modalities such as body posture, voice tone, and facial expressions, with a particular focus on leveraging facial cues for emotion inference. Recognizing facial emotions aids in enhancing human-computer interactions, affective computing, and educational processes, enabling a deeper understanding of individuals' internal states [5].

In this context, deep learning, a subfield of artificial intelligence, has emerged as a powerful tool for facial emotion detection. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in various computer vision tasks, including object detection, image classification, and facial recognition. Notably, models like VGG16 have gained prominence for their ability to extract rich hierarchical features from images, making them suitable candidates for FER tasks [6].

This paper aims to explore the efficacy of [12]VGG16, a pre-trained deep learning model, in facial emotion detection. By leveraging the hierarchical representations learned by VGG16, we seek to enhance the accuracy and robustness of emotion recognition systems. Our study builds upon existing research in FER.

2. Related Work:

A research topic is such that it deals with the development of an automated emotion identification system which is integrated with the input feature modalities and the self-supervised learning properties. The research applies SSL features to encode critical feedback gained through direct mental state checking. The approach is based on the use of the temporal convolution layers in the Wav2Vec architecture and training based on a self-supervised strategy with contrastive predictive coding concept [2].

Facial Emotion Recognition (FER) refers to the process of identifying emotions through facial expressions. It includes both primary, and also the combined emotions (CEs) which consist of the primary emotions. Micro expressions (MEs) are actually subtle facial movements which display emotions upfront in a momentarily. Facial Action Units (AUs) relate to the muscle movements that are concomitant with specific feelings. The study of FER with respect to the Facial Action Coding System (FACS) and Facial Landmarks (FLs) is crucial to understanding and categorizing facial expressions[3].

Providing an example solution achieved by employing facial detection, recognition, and emotion recognition along with NVIDIA's Jetson Nano. It demonstrates how OpenCVs DNN face detector uses deep Learning for accurate face detection and ResNet for further improvement of the performance. Besides, the work specializes on deep metric learning for face recognition purpose and highlights a technology for finding emotions with 96% accuracy as a unique feature. This is a novel standpoint compared to other single-level CNN approaches. It is realized by adjusting weights and updating exponents in an iterative way which makes the accuracy increase over iterations. FERC’s ingenious background removal procedure before feature extraction mitigates constrains such as distance from the camera thus promisingly opens new applications like student’s personal tutors and lie detector.

The model is expected to produce a sufficient amount of good performance with real-time recognition schema that provides a solid base for emotional detection. The model is trained on the FER2013 dataset and it is 96% accurate in training and 91.01% in validation. The model contains four layers of the convolution and two layers of the fully connected, having the best performance among all the available methods in this task. The authors would like to move ahead by giving more emphasis on exploring the new possibilities of facial for emotional detection [6].

Using deep learning involves three main steps: facial detection, features extraction, emotions classification. We have the DNN that provides information about the eyes and mouth regions, which determine the mood in a look that uses the computer vision approaches for the facial emotion recognition. It is shown on research that accuracy of 77.37% and 83% lead on two data sets. A lot of research comes up with data in support of the idea that not only saying words, but you can show your feelings as well (with gestures). The task is aimed on facial emotion recognition by applying Vision Transformer on top of the ResNet-18 architecture respectively. It is an evaluation where the results of the model and dataset per se are compared with the current top results. Therefore, the system empowers for successful use of the suggested approach in reality[8].

Like machine learning, these applications possess the capability to distinguish events, forecast results, and respond to real-world situations. for competing Algorithm, basis for this would be VGGNet architecture without any tuned data. The goal of this particular research is based on a hypothesis that CNN can be an appropriate tool for FER. Moreover, on other datasets, the proposed scheme was also able to identify seven classes with the help of very high precision rates [9]. The work describes the system of emotion detection which is done by face expressions captured from a live video stream or a picture in the past. Identified in python (2.7) with OpenCV and NumPy, the system comes to an attention of scanned images and then compares them with the training data-set for emotion recognition. The objective is development of a model working on a functional level and solving these problems efficiently by providing predictable output with the quality and level of interconnection of the modules ranged from very high to high.[10]

Other author introduces [11] two methods for facial emotion detection: one with Autoencoder to create unique emotional representations and another with a 8-layer Convolutional Neural Network as well. Two models were trained on the database of JAFFE, and, then, these models were tested on the images from the Labeled Faces in the Wild (LFW) dataset. Outcomes show that the CNN model and tweaking parameters lead to a performance far exceeding the existing schemes thereby paving way for refinement of autoencoders.

3. Methodology

With pretrained models that include VGG19, it is possible to get the model to perform accurate predictions and execute faster. VGG19, pre-trained CNN normally which is trained on huge dataset, is capable of extracting meaningful features from images.

The first step involves cleaning and formatting images to be inputted into the VGG 19 model, which is the model we will use to perform the analysis. The images are prepared after them and then they are fed into the VGG19 model. As the images are processed on the multiple convolutional and pooling layers the model is able to extract them.

The technique in which VGG19 is able to comprehend visual elements found within the original images, such as the detailed patterns of images, facial expressions, body language, hand gestures, and all kinds of other visual signals that can be useful in interpreting emotions. Consequently, come from the representations, the model is qualified to identify the exact feeling depicted on the input images.

Among the merits of resorting to pre-trained models such as VGG19 lies a tremendous cut in the labor-intensive process of training on large, labeled datasets, as the model already carries the general ability to detect many visual cues applicable for emotion prediction. First of all, this causes a decrease in time and computer resources and, at the same, the model will show a general good performance on its application to several datasets and scenarios. Emotion recognition with pre-trained models can be done from different perspectives, which meets the needs of human-computer interaction, sentiment analysis, and emotional computing.

3.1 Architecture:

VGG19 architecture is composed of several depthwise and fully connected layers organized in a sequential manner. The architecture can be divided into distinct blocks:The architecture can be divided into distinct blocks:

Convolutional Blocks:

The blocks are made up of convolutional layers each of them ending with an activation function (in most cases it is ReLU) and pooling layers which are often max pooling. The difference between two images is evaluated by the convolutional layer, working from the image input towards the hierarchical representations of visual patterns.

Fully Connected Blocks:

Fully connected layers are the last components in the network to accomplish with high classification rate. The layers represent the features, that were extracted, and map them to the classes that correspond to their values.

Bottleneck:

The VGG19 Network can contain structure of bottlenecks; these are subnetworks of layers creating shortcuts within the network. Such shortcuts are often customized with a parameter of shortcut = true or shortcut = false affecting the information so its spread in the network.

By using all of these blocks, the VGG19 architecture can be divided into three main parts:

Backbone:

Within the VGG19 model, the convolutional layers are used; thus, the network learns how to extract high level visualization from the input image features.

Neck:

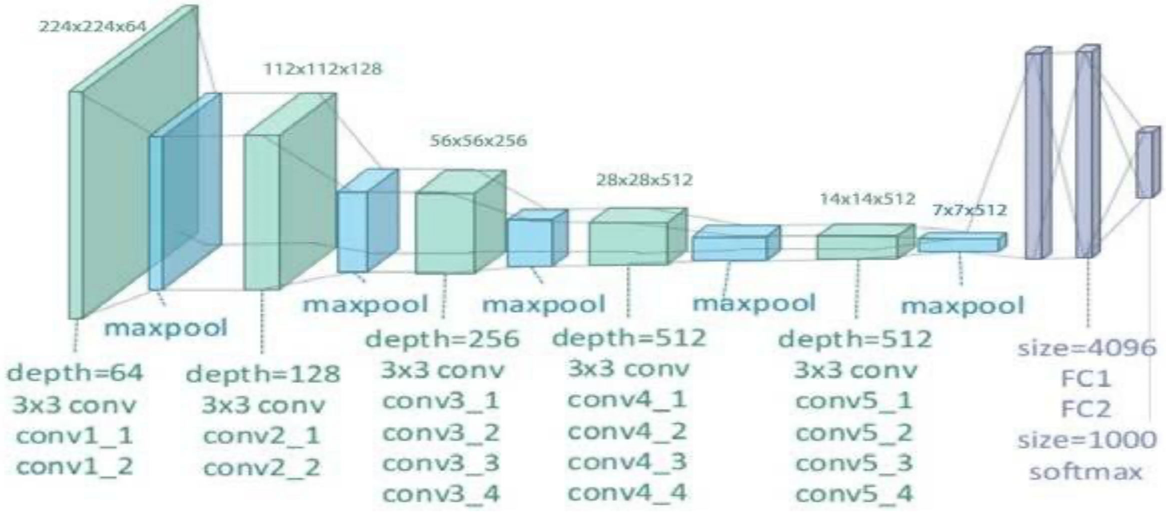
Last layer of VGG19 assigns the outputs of different layers of the deep model creating a neck to the framework. This consolidation scheme provides the model a high level of capability to capture both complicated patterns and semantic information exist in the given data.

Head:

The first layer of VGG19 performs a full involved to predict the output variables (i.e., classes or categories) from the extracted features. Besides the class labels, which are the predictions, in some cases bounding box coordinates or other information could be extracted as well.

How combinedly they work:

Input images are mapped to features extracted by the VGG19 backbone which is made of the convolutional layers that distinguish complex characteristics progressively. These functionalities are the ones that travel to the neck, where they beget and get reinforced. Lastly, the head of the model classifies the class or category of the given image using the concatenated features, therefore, the output of the entire model can be obtained.



3.1.1 VGG19 architecture

3.2 Pre-Processing:

It will be helpful to pre-process the data before giving them to the VGG19 algorithm as well as for obtaining the appropriate features and training the model.

Resizing:

Resizing to a target size of (112, 112) pixels is meant to bring uniformity and consistent processing. Resizing provides a common size to input images and implies, that they will comply with the VGG19 architecture.

Color Mode:

The images as input are converted into RGB color mode to make sure that color representation across the images are the same. RGB color mode is generally found in the computer vision tasks when three color channels are employed for the color capturing purposes.

Class Mode:

The correct mode for input images, as specified by class, is 'sparse'. In these types of images, labels associated with classes are represented by integer indices corresponding to the classes. This mode should be used in several cases when the target variables are discrete and mutually exclusive.

Batch Size:

Input data are randomly processed in batches during lent training. Each batch will have a predefined number of images. The default set batchsize is 32, which is a balance between better computation and model precision when training.

By using these pre-processing factors, VGG19 prepares the input images so that the model can successfully isolate and predict emotions. The union of down sampling, color model change, model definition and batch processing are considered to be useful in the preprocessing of images for better training and inference with the VGG19 model.

3.3 Dataset:

Collected dataset for Angry, Confused, Happy and Sad for emotion prediction.

3.5.1 Angry

3.5.2 Angry

3.5.3 Confused

3.5.4 Confused

3.5.5 Happy

3.5.6 Happy

3.5.7 Sad

3.5.8 Sad

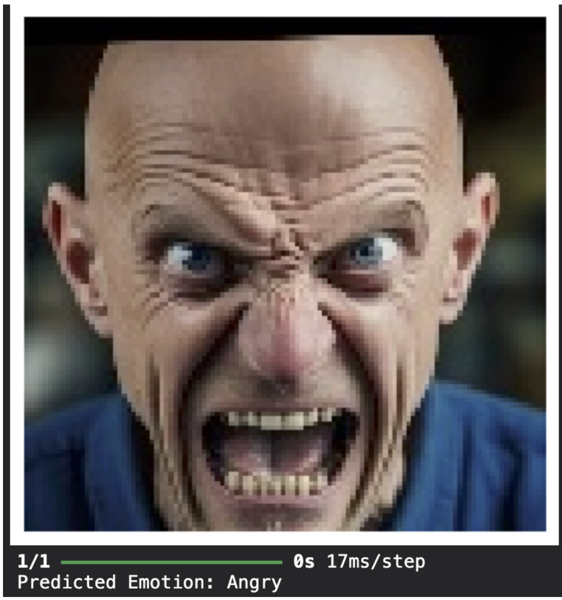
In Dataset, there are a total of 758 images for training and 149 images for testing. Dataset Link

3.4 Simulation Result:

These all results are calculated based on training accuracy, train loss, testing accuracy and test loss.

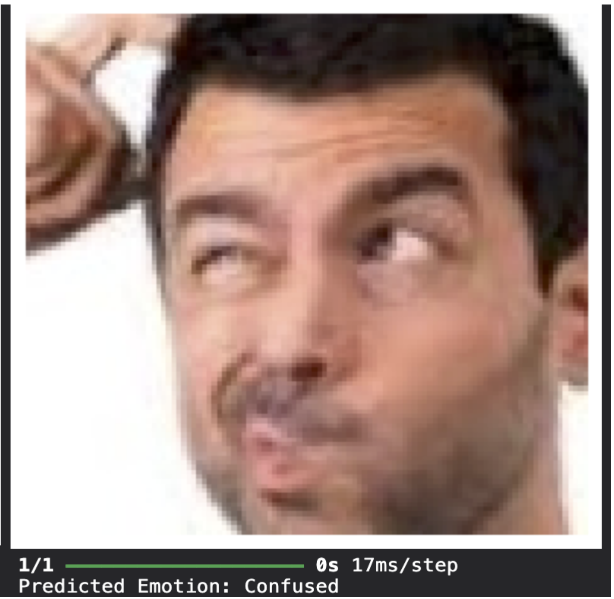
Result:

Testing Accuracy	Train Loss	Testing Accuracy	Test Loss
99.76%	0.0075	94.66%	0.56



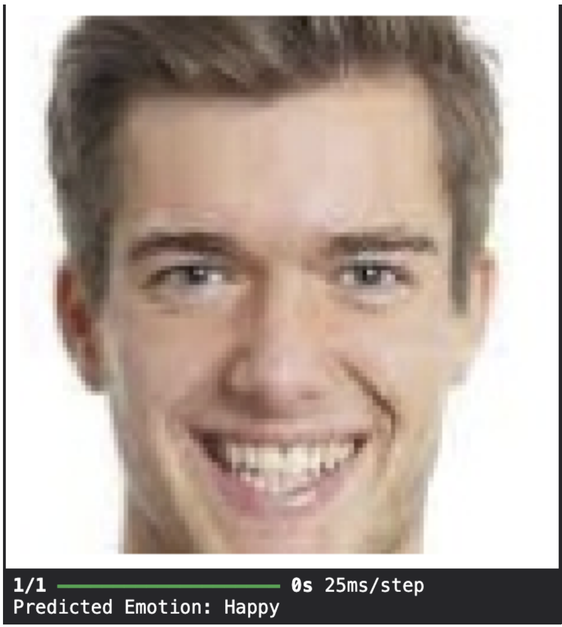
```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 17ms/step
Predicted Emotion: Angry
```

3.4.1 Angry



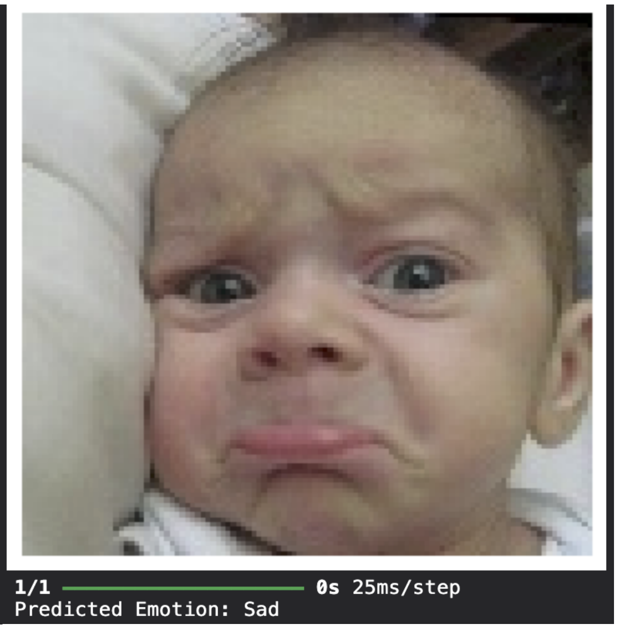
```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 17ms/step
Predicted Emotion: Confused
```

3.4.2 Confused



```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 25ms/step
Predicted Emotion: Happy
```

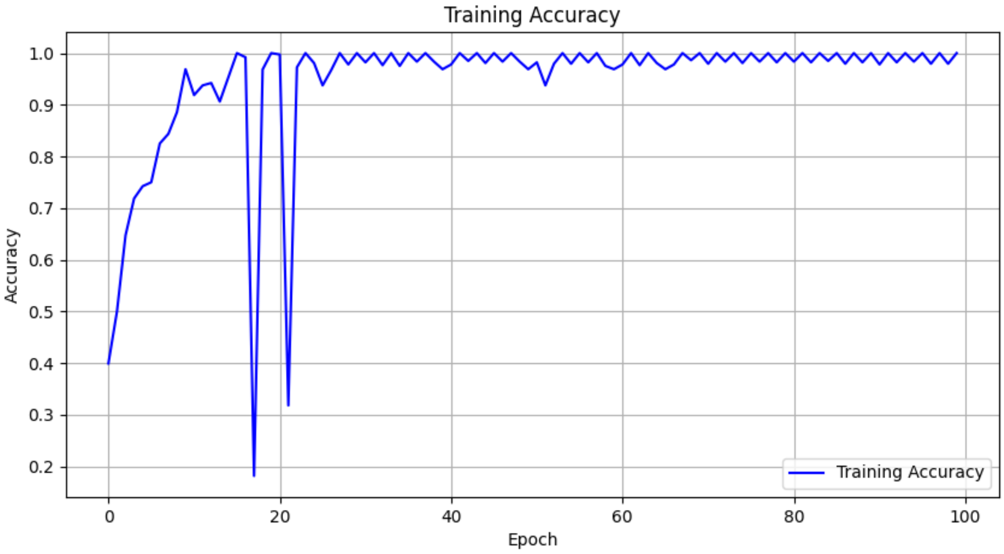
3.4.3 Happy



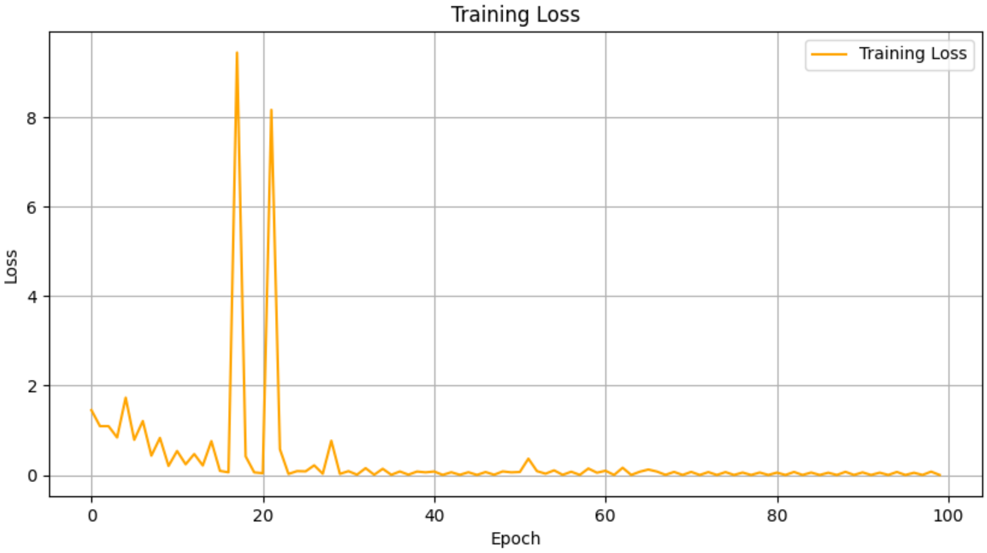
```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 25ms/step
Predicted Emotion: Sad
```

3.4.4 Sad

Result Graphs:



3.4.5 Training vs Accuracy



3.4.6 Training vs Loss

Conclusion :

In conclusion, this paper delves into the potential of VGG19 for facial emotion detection, addressing the significance of accurate emotion recognition in human communication. Through interdisciplinary research insights and advanced methodologies, including CNNs and Vision Transformer models, we showcase the promise of enhancing emotion detection algorithms. Our exploration underscores the importance of continued research in this area, with implications for various domains such as human-computer interaction and clinical practice. By leveraging pre-trained deep learning models like VGG19, we pave the way for more reliable and efficient emotion recognition systems, ultimately contributing to a deeper understanding of human emotions and intentions in diverse contexts.

References:

1. Orrite, C., Ganán, A., & Rogez, G. (2009). Hog-based decision tree for facial expression classification. In Pattern Recognition and Image Analysis: 4th Iberian Conference, IbPRIA 2009 Póvoa de Varzim, Portugal, June 10-12, 2009 Proceedings 4 (pp. 176-183). Springer Berlin Heidelberg.
2. Chaudhari, A., Bhatt, C., Krishna, A., & Travieso-González, C. M. (2023). Facial emotion recognition with inter-modality-attention-transformer-based self-supervised learning. Electronics, 12(2), 288.
3. Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. sensors, 18(2), 401.
4. Sati, V., Sánchez, S. M., Shoeibi, N., Arora, A., & Corchado, J. M. (2021). Face detection and recognition, face emotion recognition through NVIDIA Jetson Nano. In Ambient Intelligence–Software and Applications: 11th International Symposium on Ambient Intelligence (pp. 177-185). Springer International Publishing.
5. Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2(3), 446.
6. Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Facial emotion detection using neural network. the international journal of scientific and engineering research.
7. Jaiswal, A., Raju, A. K., & Deb, S. (2020, June). Facial emotion detection using deep learning. In 2020 international conference for emerging technology (INCET) (pp. 1-5). IEEE.
8. Chaudhari, A., Bhatt, C., Krishna, A., & Mazzeo, P. L. (2022). ViTFER: facial emotion recognition with vision transformers. Applied System Innovation, 5(4), 80.
9. Khaireddin, Y., & Chen, Z. Facial emotion recognition: State of the art performance on FER2013. arXiv 2021. arXiv preprint arXiv:2105.03588.
10. Puri, R., Gupta, A., Sikri, M., Tiwari, M., Pathak, N., & Goel, S. (2020). Emotion detection using image processing in python. arXiv preprint arXiv:2012.00659.
11. Dachapally, P. R. (2017). Facial emotion detection using convolutional neural networks and representational autoencoder units. arXiv preprint arXiv:1706.01509.
12. Dubey, A. K., & Jain, V. (2020). Automatic facial recognition using VGG16 based transfer learning model. *Journal of Information and Optimization Sciences*, *41*(7), 1589-1596.