*Article*

**A Step Toward Better Facial Emotion Recognition Models**

**Drumil Ved 1, Punith Nayak 1, Archana Kumar 1, Ashalatha Nayak 1, Shyam 1**

1. **Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India**

**Abstract:**  This model is a typical CNN model with convolution layer, pooling layer, activation layer, and finally classification layer. Images are loaded, normalized, and then encoded into category vectors using NumPy arrays and tensors as part of the data preprocessing step. Large models like these have a substantial risk of overfitting which was tackled using the most reliable method of dropout and regularization. Convolution layers are used for feature extraction in the neural network design, while dense linear layers—which have a greater dropout rate to minimize overfitting—are used for classification. The destination is to create a complex model with about 4 million parameters. This model performed satisfactorily on a standard dataset Facial emotion recognition - 2013 (FER-2103) with a scope of improvement being synthesis of input data and allowing an alternate methodology of testing results like K-fold cross validations method which also can reduce the risk of overfitting.

**Keywords:** Convolutional Neural Networks (CNN), Facial Expression Recognition (FER), Dropout.

1. **Introduction**

This model tries to make a deep convolutional network model to identify facial emotion recognition. Deep learning provides a great deal of accuracy and reliability when it comes to tasks like facial emotion recognition which can be employed to serve many purposes like in online education, car driver safety and business applications like advertisement. [1][2]

The world has move to a new era of online education, distant learning and teaching programs are popular today. In an online teaching leaning scenario, the instructor may not be able to understand the need and the sentiments of the learners. Which potentially leads to poor educational outcome for the learners. [1][2]

In such a scenario this model comes to rescue and helps the instructor to understand the class sentiments based on the facial expression of the learners. This model can be deployed in conjecture with the online or integrated to any custom learning portal. This model does not require heavy compute and can make predictions even on commodity hardware.

This model, which has over 4 million parameters including convolution layer, pooling layer, activation layer, effectively uses convolutional Neural networks to extract features and dense linear layers for classification. Due to this model having 4 million plus learnable parameters, and to minimize the risk of overfitting a high dropout was introduced. Dropout allows minimal overfitting by statistical methods of scaling up and down the parameter weights.

The rest of the paper is divided as follows., and section II is related to work. Section III elaborates on the objectives. Section IV describes the dataset in detail; section V is the methodology followed for this paper. Section VI describes the results. Section VII discusses the conclusion and future work; section VIII is for citations.

1. **Related Work**

A novel method combines Principal Component Analysis (PCA) for dimensionality reduction in facial emotion recognition with deep learning for optimal facial feature extraction. Because significant variation is preserved, using PCA's compression function on deep learning features increases accuracy and computing efficiency. The research shows that this method has potential to improve facial emotion recognition systems, as it performs better than direct feature extraction with the VGG-Face model. One disadvantage of this method is that, especially in complex emotional circumstances, PCA may be able to capture some intricacies of the complex human facial expressions which can lead to information loss.

[3][4][5].

This study focuses on face detection and recognition to meet the requirement for automated understanding and evaluation of picture and video datasets, especially in face identification, appearance recognition, and human-computer interaction applications. Highlighting the role facial recognition plays in biometrics. This paper's exclusive focus on biometrics-related facial recognition may be one of its limitations. This paper may have ignored the possible biases introduced during the training of the model which is a limitation [6][7].

The research uses Histogram of Oriented Gradient (HOG) characteristics to describe five algorithms to identify a few emotions from facial pictures. These algorithms cover conventional methods (SVM, MLP) and deep learning techniques. It evaluates these techniques based on four core feelings. It demonstrates the usefulness of the proposed FER-CNN paradigm and emphasizes how many applications it may be used for, such as e-learning, marketing, entertainment, and healthcare.

One of the possible limitations of this research is that it only analyzes a few simple emotions. Since there are many complex human emotions, detection of which is complex and can require high computation and time. Because of this, applying the findings to situations in the real world when a more comprehensive range of emotions is present may be more difficult [8][9][10].

Deep neural networks can aid in finding out different facial emotion features or support and provide new information about the complex relationship between stress, subjective experiences, facial expressions, the survey emphasizes the potential of facial emotion recognition systems for mental stress detection [8].

This paper deals with identifying mental stress on face recognition utilizing a deep transfer network (DTN) and 3D morphable models (3DMMs). By utilizing 3DMMs to synthesize faces in different expressions, the study solves the problems of a need for annotated face images and dataset bias between artificial and natural photographs. This paper suggests synthesis of face images to provide a larger dataset for training deep neural networks and to reduce variance and/or bias of the given dataset. The results of the synthetic data are promising as it can potentially reduce the bias of the model. One potential limitation on the study could be the use of synthetic facial images generated by 3D morphable models (3DMMs) to augment the training data. Even though synthetic data can help minimize the shortage of labeled face photos, it may not fully capture the complexity and diversity inherent in real-world facial expressions and emotions. A performance gap could occur when the model is used in real-world scenarios where individuals exhibit different stress indicators and facial expressions that synthetic data might not sufficiently reflect.[12][13][14].

On the FER-2013 dataset, this article offers a Facial Emotion Recognition (FER) system that achieved 63.39% accuracy on a test set using Transfer Learning technique which was based on GoogLensNet architecture. It shows how widely FER is used in many different sectors and depicts how successful deep learning-based techniques are, particularly when classifying emotions in facial pictures [11].

To manage skewed FER datasets, this research paper presents a unique method for Facial Expression Recognition (FER) in educational contexts called HoE-CNN. This uses an ensemble of deep convolutional neural networks. Convolutional neural networks in an ensemble are used in this method. It addresses FER issues, particularly in multi-class labeling and imbalanced data, and demonstrates improved performance in comparison to individual deep learning techniques, with a focus on online learning applications.[15][16].

1. **Objective:**

This research paper aims to make a facial emotion recognition model that can help the teaching-learning process with low computing and the highest possible accuracy. The aim is to produce the best possible results with available datasets and known models to leverage and get the maximum possible outcome.

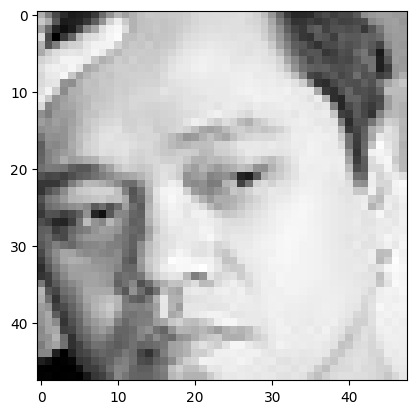
1. **Dataset**

The dataset is called the Facial Expression Dataset (FED-2013). The dataset contains seven emotions: Fear, Neutral, Angry, Happy, Surprise, Disgust, and Sad. Figures 1 and 2 show a sample set of datasets and one individual sample image respectively. Figure 3 shows the distribution of the image classes, the dataset had about 35900 images of all the classes combined. The dataset was divided into two parts: train and test. The train: test split was

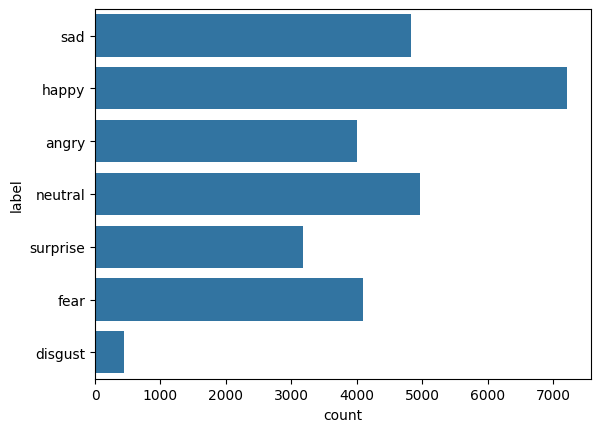
80: 20 % (28709: 7178).



**Fig 1.) A sample subset showing all the classes.**



**Fig 2.) A 48 x 48 grayscale image from dataset.**



**Fig 3.) This graph shows the distribution of various classes among train directory.**

1. **Methodology**

The methodology of this paper is to first preprocess the data by normalization and then to feed this normalized data to the model train it.

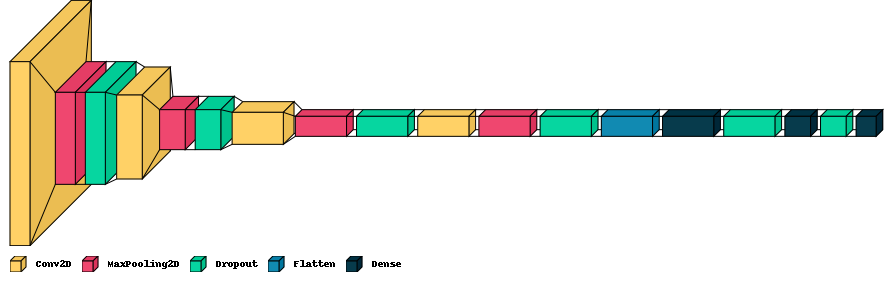
**Data Preprocessing:**

All the images were loaded to a NumPy array and then to a tensor. Each image was normalized. The images are then encoded into category vectors.

**Model Architecture:**

The methodology to make the neural network is reasonably intuitive. The first few Convolution layers of the neural network are for feature extraction, and the following dense linear layers are dedicated to classification. In implementing this model, to curb overfitting and optimize training, a dropout of 40 % was chosen and a relatively large neural network with over 4,232,199 trainable parameters. A convolution kernel size of 3 x 3 was chosen as it is a common practice to choose a smaller size and then try with a larger kernel size of 5 x 5. The standard Rectified Linear Unit (ReLU) was chosen to introduce nonlinearity. The convolution and max pooling layers kept decreasing the image from 48 x 48 to 1 x 1 while doubling the feature maps in every convolution layer and starting from 128 feature maps in the first convolution layer to 512 in the third convolution layer, doubling every convolution layer. The later linear layers are for classification, and the successive linear layer size decreases until the size equals the number of classes. Then SoftMax activation function at the final output layer to get a definitive answer.

The standard T4 GPU of Google Colab was used for this project. Figure 5 represents the model architecture.



**Fig 4.) Visual representation of the model architecture.**

1. **Results**

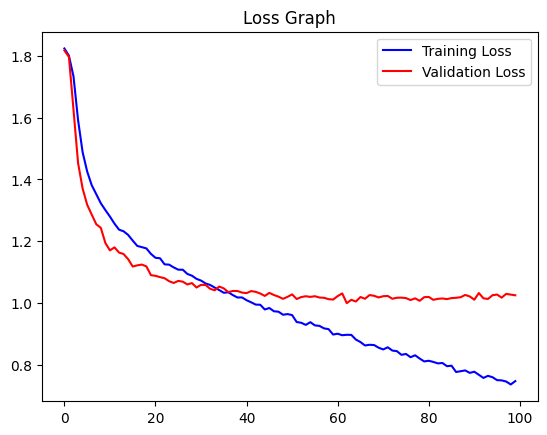
The model showed an accuracy of about 63.50 % while testing; the training was 100 epochs. Figure 6 and 7 shows the results in the form of a graph and output prediction of the model.

epochs

epochs

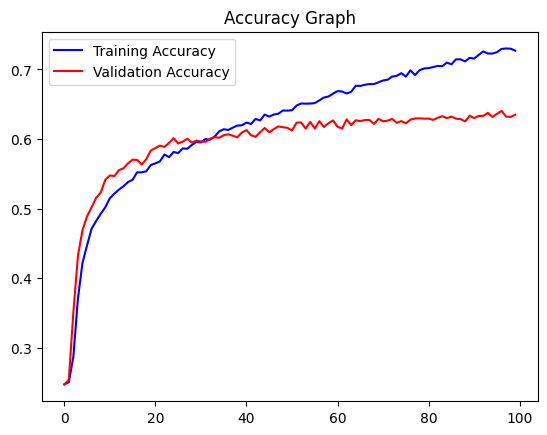
Loss

Accuracy

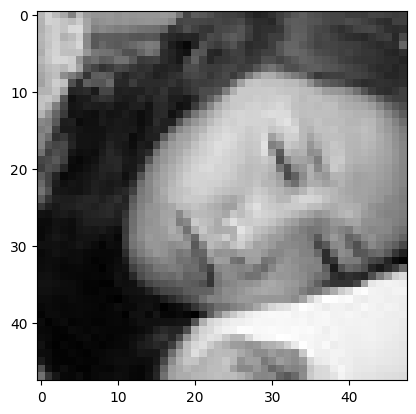
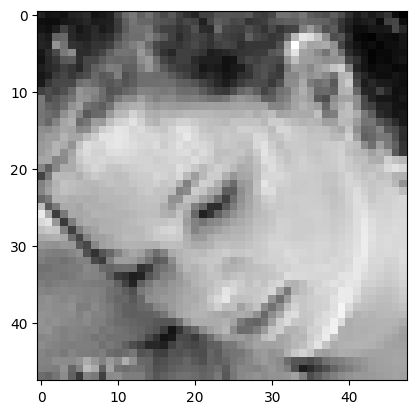


Epochs

Epochs



**Fig 5.) Showing the loss and accuracy matrices of the model**



**Original Output: sad**

**Predicted Output: neutral**

**Original Output: sad**

**Predicted Output: sad**

**Fig 6.) Showing the exact model predictions**

The Figure 7 compares the model with its contemporary models.

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No | Dataset | Model | Accuracy. |
| 1. | JAFFE1 | DenseNet161 | 96.51% |
| 2. | FER20132 | CNN | 71.97% |
| 3. | FER21032 | Improved ResNet 18 | 83% |
| 4. | FER30132 | This Model | 64.49% |

1. Japanese Female Facial Expression (JAFFE)
2. Facial Expression Recognition 2013 (FER2013)

**Figure 7 comparison with other models [17][18]**

1. **Conclusion and Future Objective**

Even though this model for identifying emotions shows promise, a few issues must be resolved in further research. Two datasets, CK+ and FER2013, were used in this work; however, more dataset exploration is necessary for robustness, significantly increasing dataset size with larger subsets of CK+ and FER2013 to improve emotion identification results. Furthermore, future research could improve validation by creating task-specific CNN models for emotion recognition systems and combining various assessment approaches like k-fold cross-validation and statistical analysis. Many directions could be explored in the future. Research on ablation datasets like EMOTIC and EMO-DB could be done to increase generalizability. Moreover, continuous optimization efforts could enhance this model architecture by adding new blocks for increased precision and efficiency.

Moreover, adding speech and body language to the list of emotions that can be recognized provides creative ideas for valuable applications. Conclusion: Although this model shows encouraging results, it is essential to address the constraints that have been found and incorporate them into future developments. With continued study and cooperation, this model has great potential to advance the field of emotion recognition and eventually serve society by becoming more functional and integrating into real-world situations.

1. **References**
   * 1. A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, Serbia, 2018, pp. 1-4, doi: 10.1109/NEUREL.2018.8587011.
     2. Rit Lawpanom, Wararat Songpan \* and Jakkrit Kaewyotha, “Advancing Facial Expression Recognition in Online Learning Education Using a Homogeneous Ensemble Convolutional Neural Network Approach”.
     3. B. Islam, F. Mahmud and A. Hossain, "High Performance Facial Expression Recognition System Using Facial Region Segmentation, Fusion of HOG & LBP Features and Multiclass SVM," 2018 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2018, pp. 42-45, doi: 10.1109/ICECE.2018
     4. S. Alfattama, P. Kanungo and S. K. Bisoy, "Face Recognition from Partial Face Data," 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT), Bhubaneswar, India, 2021, pp. 1-5, doi: 10.1109/APSIT52773.2021.9641286.
     5. Xingfu Zhang and Xiangmin Ren, "Two Dimensional Principal Component Analysis based Independent Component Analysis for face recognition," 2011 International Conference on Multimedia Technology, Hangzhou, China, 2011, pp. 934-936, doi: 10.1109/ICMT.2011.6002199.
     6. A. Kumari Sirivarshitha, K. Sravani, K. S. Priya and V. Bhavani, "An approach for Face Detection and Face Recognition using OpenCV and Face Recognition Libraries in Python," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1274-1278, doi: 10.1109/ICACCS57279.2023.10113066.
     7. P. Podder, T. Z. Khan, M. H. Khan, M. M. Rahman, R. Ahmed and M. S. Rahman, "An efficient iris segmentation model based on eyelids and eyelashes detection in iris recognition system," 2015 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2015, pp. 1-7, doi: 10.1109/ICCCI.2015.7218078.
     8. F. J. Ming, S. Shabana Anhum, S. Islam and K. H. Keoy, "Facial Emotion Recognition System for Mental Stress Detection among University Students," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10252617.
     9. H. -c. Yang and X. A. Wang, "Cascade Face Detection Based on Histograms of Oriented Gradients and Support Vector Machine," 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, Poland, 2015, pp. 766-770, doi: 10.1109/3PGCIC.2015.14.
     10. R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, Indonesia, 2016, pp. 163-168, doi: 10.1109/ISEMANTIC.2016.7873831.
     11. F. J. Ming, S. Shabana Anhum, S. Islam and K. H. Keoy, "Facial Emotion Recognition System for Mental Stress Detection among University Students," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10252617.
     12. Y. Zhang, H. Wang, F. Xu and K. Jia, "A Deep Face Recognition Method Based on Model Fine-tuning and Principal Component Analysis," 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Honolulu, HI, USA, 2017, pp. 141-146, doi: 10.1109/CYBER.2017.8446331.
     13. Z. Ding, N. M. Nasrabadi and Y. Fu, "Task-driven deep transfer learning for image classification," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 2414-2418, doi: 10.1109/ICASSP.2016.7472110.
     14. D. Sopiak et al., "Generating face images based on 3D morphable model," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 2016, pp. 58-62, doi: 10.1109/FSKD.2016.7603151.
     15. Imane Bouslihim(B) and Walid Cherif, “Facial Emotion Recognition Using a GoogLeNet Architecture”
     16. Z. Lihong and G. Zikui, "Face Recognition Method Based on Adaptively Weighted Block-Two Dimensional Principal Component Analysis," 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, Bali, Indonesia, 2011, pp. 22-25, doi: 10.1109/CICSyN.2011.18.
     17. Zhang, S.; Zhang, Y.; Zhang, Y.; Wang, Y.; Song, Z. A Dual-Direction Attention Mixed Feature Network for Facial Expression Recognition. Electronics 2023, 12, 3595. https://doi.org/10.3390/electronics12173595
     18. L. Pham, T. H. Vu and T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 4513-4519, doi: 10.1109/ICPR48806.2021.9411919.
     19. P. Dinkova, P. Georgieva, A. Manolova and M. Milanova, "Face recognition based on subject dependent Hidden Markov Models," 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Varna, Bulgaria, 2016, pp. 1-5, doi: 10.1109/BlackSeaCom.2016.7901570.
     20. C. -W. Tan and A. Kumar, "Integrating ocular and iris descriptors for fake iris image detection," 2nd International Workshop on Biometrics and Forensics, Valletta, Malta, 2014, pp. 1-4, doi: 10.1109/IWBF.2014.6914251.
     21. P. C. Neto, J. R. Pinto, F. Boutros, N. Damer, A. F. Sequeira and J. S. Cardoso, "Beyond Masks: On the Generalization of Masked Face Recognition Models to Occluded Face Recognition," in IEEE Access, vol. 10, pp. 86222-86233, 2022, doi: 10.1109/ACCESS.2022.3199014.
     22. Z. An, W. Deng, T. Yuan and J. Hu, "Deep Transfer Network with 3D Morphable Models for Face Recognition," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 2018, pp. 416-422, doi: 10.1109/FG.2018.00067.
     23. Rabie Helaly, Seifeddine Messaoud, Soulef Bouaafia,  Mohamed Ali Hajjaji, Abdellatif Mtibaa, “DTL-I-ResNet18: facial emotion recognition based on deep transfer learning and improved ResNet18”
     24. Saining Zhang, Yuhang Zhang, Ye Zhang , Yufei Wang and Zhigang Song 4, “A Dual-Direction Attention Mixed Feature Network for Facial Expression Recognition”
     25. Fei Yang, Dan Xia and Fanbao Meng, "An improved multi-view face synthesis method based on point distribution models," 2012 24th Chinese Control and Decision Conference (CCDC), Taiyuan, 2012, pp. 208-211, doi: 10.1109/CCDC.2012.6244030.
     26. C. Li, J. Feng, L. Hu, J. Li and H. Ma, "Review of Image Classification Method Based on Deep Transfer Learning," 2020 16th International Conference on Computational Intelligence and Security (CIS), Guangxi, China, 2020, pp. 104-108, doi: 10.1109/CIS52066.2020.00031
     27. L. Wang, B. Liu, S. Su, Y. Cheng and S. Li, "An improved 3D Bilinear Multidimensional Morphable Models used in face recognition," 2014 International Conference on Information Science, Electronics and Electrical Engineering, Sapporo, Japan, 2014, pp. 2052-2056, doi: 10.1109/InfoSEEE.2014.6946284.
     28. Z. Chang, X. Zhang, S. Wang, S. Ma, Y. Ye and W. Gao, "STAE: A Spatiotemporal Auto-Encoder for High-Resolution Video Prediction," 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/ICME51207.2021.9428231.
     29. Y. Luo, J. Wu, Z. Zhang, H. Zhao and Z. Shu, "Design of Facial Expression Recognition Algorithm Based on CNN Model," 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 580-583, doi: 10.1109/ICPECA56706.2023.10075779.
     30. L. Greche, N. Es-Sbai and E. Lavendelis, "Histogram of oriented gradient and multi layer feed forward neural network for facial expression identification," 2017 International Conference on Control, Automation and Diagnosis (ICCAD), Hammamet, Tunisia, 2017, pp. 333-337, doi: 10.1109/CADIAG.2017.8075680.
     31. P. N. Maraskolhe and A. S. Bhalchandra, "Analysis of Facial Expression Recognition using Histogram of Oriented Gradient (HOG)," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 1007-1011, doi: 10.1109/ICECA.2019.8821814
     32. M. S. Kaushik and A. B. Kandali, "Recognition of facial expressions extracting salient features using local binary patterns and histogram of oriented gradients," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 1201-1205, doi: 10.1109/ICECDS.2017.8389632.
     33. J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah and D. Dunaway, "A 3D Morphable Model Learnt from 10,000 Faces," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 5543-5552, doi: 10.1109/CVPR.2016.598.