Kütahya Dumlupınar University
Institute of Graduate Studies

Journal of Scientific Reports-A
E-ISSN: 2687-6167
Number 49, June 2022

RESEARCH ARTICLE

FACIAL EXPRESSION RECOGNITION on PARTIAL FACE IMAGES USING DEEP TRANSFER LEARNING

İsmail ÖZTEL[1*], Gözde YOLCU ÖZTEL[2], Veysel Harun ŞAHİN[3]

[1*]Sakarya University, Faculty of Computer and Information Sciences, Department of Computer Engineering, ioztel@sakarya.edu.tr, ORCID: 0000-0001-5157-7035
[2]Sakarya University, Faculty of Computer and Information Sciences, Department of Software Engineering, gyolcu@sakarya.edu.tr, ORCID: 0000-0002-7841-2131
[3]Sakarya University, Faculty of Computer and Information Sciences, Department of Software Engineering, vsahin@sakarya.edu.tr, ORCID: 0000-0002-3381-1702

Receive Date:17.04.2022 *Accepted Date: 11.06.2022*

ABSTRACT

Facial expression recognition has a crucial role in communication. Computerized facial expression recognition systems have been developed for many purposes. People's faces can have occlusions because of scarves, facial masks, etc. in cases such as cold weather conditions or Covid-19 pandemic conditions. In this case, facial expression recognition can be challenging for automated systems. This study classifies facial images containing only the eyebrow and eye regions over six expressions with a deep learning-based approach. For this purpose, Radboud Face Database images have been used after cropping the area that includes eye and eyebrow regions. Some popular pre-trained networks have been trained and tested using the transfer learning approach. The Vgg19 pre-trained network achieved 91.33% accuracy over the six universal facial expressions. The experiments show that automated facial expression recognition can be applied with high performance by looking at the region containing eyes and eyebrows.

Keywords: *Convolutional Neural Networks, Deep Learning, Facial Expression Recognition, Transfer Learning*

1. INTRODUCTION

Facial expressions are produced by different contraction and relaxation of facial muscles. Human facial expressions are important for social communications. According to [1], in face-to-face communication, 7% of the message is transmitted by words, 38% by tone of voice, and 55% by body language. Facial expressions are also a type of body language.

Owing to technological developments, cameras became a part of our daily lives. Thus, facial expressions have been started to use in many fields. Facial expressions are essential for decision-making systems like criminal detection, patient follow-up, and driver attention monitoring. Although facial expressions are detected by the human brain effortlessly, it is pretty hard to detect for a machine. Particularly, a person’s image with the same facial expression can show differences in the analysis due to brightness, background, etc. [2]. Also, there may be occlusions on human faces because of some type of cloth like scarves, face masks, etc. Therefore, with the advances in computer vision, detecting facial expressions in a computerized manner has become a popular topic.

Detecting facial expressions using automated systems is one of the important topics of computer vision. Some areas that use automated facial expression recognition (FER) are avatar animation [3], medical applications [4], robotics [5], traffic [6][7], smart environments [8], human-computer interaction [9][10]. There are six universal facial expressions: anger, disgust, fear, happiness, sadness, and surprise. These expressions were identified by Ekman and Friesen [11]. The FER studies in the literature are generally based on these six expressions.

Recently, because of Covid-19 pandemic, people have used face masks. According to WHO reports, using a medical face mask can decrease the spread of disease. Currently, using medial face mask is mandatory in many countries. Face masks create facial occlusion on the human face. Thus, recognition of these people's facial expressions with mask becomes very hard.

The main motivation of the study is detecting the facial expressions for occlusion situations such as wearing a scarf or face mask. In this paper, a deep transfer learning-based FER system that works on partial face images is presented. The system detects facial expressions using just the top part of the face. In the context of the study, firstly, the face images have been prepared to include just the upper part of the face using the pre-processing methods. Then the system has been trained with different pretrained convolutional neural networks (CNN). During this stage, each network used the top part of the face as input and produced prediction label of the facial expression. The main contributions of this paper can be summarized as follows.

(1) The proposed approach uses just the top part of the face for FER. Therefore, it is robust to bottom face occlusions and motion.

(2) Using a large input image size in the input layer in the deep learning systems requires large memory and increases training and testing time. As the proposed approach uses the top part of the face image it requires a small input image size. Therefore, these disadvantages have been prevented.

In the literature, automated FER studies are examined in three main methods [12], [13]. These are geometric-based, appearance-based, and action unit (AU)-based approaches.

In the geometric-based analysis, face shape is examined. For example, in [14] the facial expressions were classified using the distances and angles between facial components using facial landmarks.

Appearance-based approaches examine features of the facial texture. For example, Saurav et al. [15] introduced several techniques that make use of dynamic local ternary patterns. In their work they applied kernel extreme learning machine classifier. Iqbal et al. [16] proposed neighborhood-aware edge directional pattern which is a local descriptor. Dagher et al. [17] presented a three-stage support vector machine (SVM) for FER. Ali et al. [18] proposed a new method that uses Rectangular HOG feature extractor with Label-Consistent KSVD classifier.

In AU-based approach, the facial muscles movements are identified using a system. This system is called the Facial Action Coding System (FACS) [19]. Morales-Vargas et al. [20] introduced a model that recognizes facial expressions in terms of facial movements with the help of appearance features. They get help from the fuzzy models.

Also, Deep Learning approaches have shown outstanding performance in automated facial expression studies recently. For example, Jin et al. [21] introduced a small and effective deep network for FER named MiniExpNet. Their method is based on local facial regions. Fan et al. [22] introduced a new CNN namely hierarchical scale convolutional neural network (HSNet). Their approach enhances the information extracted from the network, kernel, and knowledge scale. Saurav et al. [23] presented CNNV3 approach. They used the approach in an intelligent embedded system that is developed to help visually impaired people. The system is a vision-based and deep-learning inspired system that performs haptic rendering of facial emotions. Happy et al. [24] introduced multi-face multi-part model. Their model makes use the information from part-based data augmentation. The facial parts which they used for training were obtained from different face samples of the same expression class. Jin et al. [25] proposed a discriminative deep association learning framework. In [26], Happy et al. proposed a weakly supervised learning approach. In their approach, they first train CNN with label smoothing. They do this in a supervised manner. Then, they adjust the weights simultaneously with labelled and unlabeled data. Luh et al. [27] utilized YOLOv3 to solve FER problem.

2. METHODOLOGY

The proposed system uses facial images as input, and it produces facial expression labels using just the eyes and eyebrows region of the face. For this purpose, first, just the eyes and eyebrows part of the facial images are cropped using the pre-processing methods. Then, using the deep transfer learning method, the facial expressions are classified. Different pre-trained networks have been used for this purpose and results have been compared. The system's pipeline is shown in Figure 1.

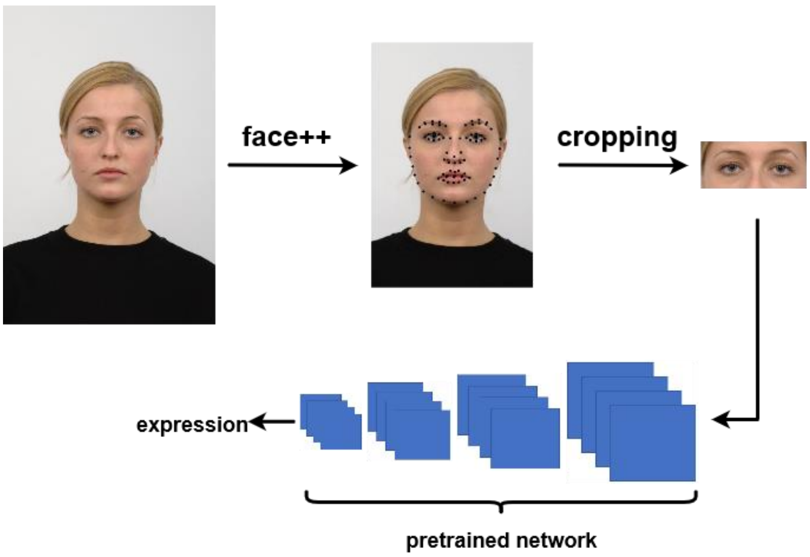

Figure 1. Proposed pipeline.

JSR A
Journal of Scientific Reports

Öztel, et all., Journal of Scientific Reports-A, Number 49, 118-129, June 2022.

2.1 CNN

The feature extraction phase required in classical machine learning methods is a challenging process. Because experts must determine the features that affect the solution of a problem. This a very time-consuming process. On the contrary, deep learning processes raw data without feature extraction phase. Therefore, deep learning has eliminated the problems in feature extraction.

CNN consists of a combination of layers. These layers are input, convolution, pooling, activation, dropout, fully connected, and classification layer. In the convolution layer, the filters move over the image and perform the convolution operation. Feature maps are created by mapping the features revealed by each filter. The mathematical model of the convolution is shown in Eq. 1 [28].

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \tag{1}$$

In the equation, S, I, and K represent the output, the input image, and the kernel respectively. Asterisk (*) is the convolution operation.

In the pooling layer, the input size is reduced in width and height. There is no change in depth. Pooling is not a mandatory layer. It may or may not be used according to the design.

In [18], it has been shown that CNNs using a nonlinear activation function like Relu are trained more quickly. Therefore, in this paper, Relu is used as the activation function. The mathematical model of Relu is shown in Eq. 2 [28].

$$f(x) = max(0, x) \tag{2}$$

In the dropout layer, some nodes of the network are removed to prevent the network from being dependent on a particular neuron. The fully connected layer is dependent on all fields of the previous layer. Softmax converts the predictions to positive values. Also, the softmax applies normalization to get a probability distribution of classes. The mathematical model of softmax is shown in Eq. 3 [28].

$$softmax(x)_i = \frac{exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \tag{3}$$

In the equation, x_i represents the input of softmax, and it is obtained from the fully connected layer.

The last layer is the classification layer. The number of the output of the classification layer is equal to the number of classes.

In the literature, the performance of classification problems is usually measured by accuracy. Since facial expressions are classified in this study, the performance has been measured with accuracy and comparisons have been made on this metric. The accuracy formula is given in Eq. 4.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)} \tag{4}$$

In the equation, TP represents true positives, TN represents true negatives, FN represents false negatives, and FP represents false positives.

2.2. Pre-trained Networks

From the network design and training viewpoint, two different approaches can be used in deep learning. The first one is to design and train a novel network from scratch. The second one is to get help from existing networks that are already trained. These networks are called pre-trained networks. The second approach is called transfer learning. Providing more accurate results with fewer data is one of the advantages of transfer learning. In addition, training a network using random weights is slower than transfer learning. Transfer learning allows the quick transfer of features. Therefore, new deep learning tasks can be achieved using fewer images in training phase.

In this study, nine different pre-trained networks have been used for FER on partial face images tasks. All of them were trained on the ImageNet database [29]. These networks were trained using over a million images. The networks can classify images into 1000 categories like the variety of animals, mouse, pencil, etc. Thus, the networks have robust feature representations for different images. An explanation of these networks is given below.

Nasnetmobile [30] has been introduced by Google. Its size is 20 MB, it has 5.3×10^6 parameters. It uses 224×224 sized input images. Shufflenet [31] is a CNN structure which is built especially for mobile devices with limited computing power. This network uses two new approaches to reduce computational costs. These approaches are pointwise group convolution and channel shuffle. Its size is 5.4 MB. It has 1.4×10^6 parameters and uses 224×224 sized input images. Googlenet [32] is a 22 layers deep learning network that is developed by Google researchers. Its size is 5.4 MB. It has 7.0×10^6 parameters. Its default input size is 224×224. Mobilenetv2 [33] is a CNN structure that was developed especially for mobile devices. Its architecture includes the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. Its size is 13 MB. It has 3.5×10^6 parameters. Resnet50 [34] is a convolutional neural network that has 50 layers. It uses residual blocks to improve the accuracy. There is "skip connections" concept at the core of the residual blocks. This is the strength of resnet50. Its size is 96 MB, and it has 25.6×10^6 parameters. Its input size is 224×224. The main approach for densenet201 [35] is the connections between the network layers. Its size is 77MB, it has 201 layers and 20×10^6 parameters. Alexnet [36] is a classic CNN architecture. In alexnet, grouped convolutions are designed for fitting the model across two GPUs. Its size is 227 MB, and it has 61×10^6 parameters. Vgg [37] has different versions. In this study both 16 depth and 19 depth version have been used. Vgg16 has 138×10^6 parameters and vgg19 has 144 $x10^6$ parameters. These models use 224×224 sized input images.

In this study, first, the last two layers of the pre-trained networks (fully connected layers and classification layers) have been modified for six classes. Then, they have been re-trained using the eye and eyebrow parts of the facial images.

3. EXPERIMENTAL RESULTS

In this study, the learning rate has been determined as 0.0001 for densenet201, googlenet, mobilenetv2, nasnetmobile, resnet50, shufflenet, vgg16 and vgg19. Alexnet doesn't work with 0.0001. Thus 0.001 is used as the learning rate for alexnet. Also, the optimizer is sgdm, learn rate drop factor

JSR A
Journal of Scientific Reports

Öztel, et all., Journal of Scientific Reports-A, Number 49, 118-129, June 2022.

is 0.2, learn rate drop period is 5, the momentum is 0.9, the mini batch size is 8, validation frequency is 3 for all pre-trained networks. Each network has been trained using 50 epochs.

3.1. Dataset and Preprocessing

The Radboud Faces Database (RaFD) [38] was used in both training and testing phase in this study. The dataset contains whole face images of 67 people, including children and adults. The dataset includes three different gaze directions of people. These are looking left, looking frontal, looking right. In this study, the frontal faces are used for this task. There are eight expressions in RaFD. The facial expression recognition studies often investigate six main expressions. These are anger, disgust, fear, happiness, sadness and surprise. These expressions were defined in the early works of Ekman and Friesen [11]. The images used in this study have six universal facial expressions with different gaze directions (Figure 2). 1206 images were used in this study. 844 images were used for training and the rest of the images were used for testing.

As mentioned earlier, this study uses the upper part of the face for facial expression recognition. Therefore, the upper part of the face must be cropped from the raw image. For this purpose, firstly, the face keypoints were localized on the face images using Face++ toolkit [39]. This step is shown in Figure 1. In this phase, 83 facial keypoints were localized for each face image. Then the region between left and right contours of the face which have the same x-axis as the midpoint of the nose were cropped horizontally. Vertically, the bottom point is considered the nose midpoint. The vertical top point was determined using the equation of $horizontal\ area/vertical\ area = 0.2$. In cropped region 0.2 is determined experimentally. Since the wrinkles on the forehead are also a critical determinant in some facial expressions, these regions are especially included in the cropped image.

3.2. Comparative Results

In the proposed study, nine pre-trained networks have been restructured for image classification task and their performance have been compared in terms of accuracy. The comparison table is given in Table 2. As can be seen in the table, the best result obtained using vgg19 as 91.33%. Vgg16 is second behind it as 88.33%. It can be seen that vgg19 and vgg16 have the greater parameters compared to other networks. This success rate can be explained by the number of parameters. Although shufflenet's number of the parameter is less than nasnetmobile, its performance is better than nasnetmobile.

Figure 2. Sample images from the dataset for six expressions and three gaze directions ((a) anger, (b) disgust, (c) fear, (d) happy, (e) sad, (f) surprise, (g) looking frontal, (h) looking right, (i) looking left)

JSR A
Journal of Scientific Reports

Öztel, et all., Journal of Scientific Reports-A, Number 49, 118-129, June 2022.

Table 2. Recognition accuracies using different pre-trained networks.

Pretrained network	Accuracy (%)
nasnetmobile	67.67
shufflenet	81.67
googlenet	82.00
mobilenetv2	82.00
resnet50	84.00
densenet201	85.67
alexnet	87.33
vgg16	88.33
vgg19	91.33

The confusion matrix of the vgg19 network that shows the best accuracy is given in Figure 3. As can be seen in the figure, sad expression is predicted as 100% success rate by the system. The system sometimes confuses surprise and fear expressions. When the sample images given in Figure 2 are examined, it can be seen that the eyes are more open than normal in both expressions. This may cause confusion.

The proposed approach has been compared to the other studies that use the same RaFD database. The comparison results are shown in Table 3. In the literature, the studies mostly used the whole face. Although the proposed study produced lower results than studies using whole faces, its performance is quite close to them using a small part of the face. In [40], five facial expressions were classified using partial faces. Our study produced a result close to that, despite the classification of six expressions in our study.

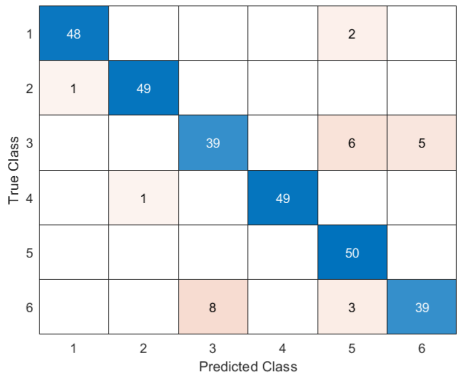

Figure 3. Confusion matrix of vgg19 (1: angry, 2: disgust, 3: fear, 4: happy, 5: sad, 6: surprise).

Table 3. Comparison with the studies used RaFD.

Study	#Expression	Whole/partial face	Accuracy
Oztel at al. [40]	5	whole	94.07
Nayak at al. [41]	6	whole	95.60
Ali at al. [18]	6	whole	98.80
Oztel at al. [40]	5	partial	92.59

Proposed approach with vgg19	6	partial	91.33

4. CONCLUSION

WHO recommends wearing a face mask to reduce the spread of Covid-19. In this case, an important part of the face for facial expression recognition is closed. Also, lower part of the face can be occluded some reasons such as wearing a scarf. In these case, facial expressions must be detected using only the eyes and eyebrows area. This is a challenging problem for machine learning. With this motivation, our study proposes a deep transfer learning-based approach for partial face. The system just uses upper part of the face for predicting facial expression. For this purpose, database images have been cropped to include just upper part of the face. Then the proposed system has been trained using nine different pre-trained networks and the results were compared. The vgg19 network produced the best result as 91.33%. The best score obtained for sad expression as 100%. Happy and disgust expression detected as 98%. The system sometimes confused surprise and fear expressions. This situation has been interpreted with images in the study. The study was also compared with other studies that use the same database in the literature and produced comparable results.

The proposed system can be useful for using facial expressions in situations where facial movement is active such as speaking, and in cases of occlusions due to scarf, mask, etc. It is planned to expand the same work for real time in the near future.

ACKNOWLEDGEMENT

This work was supported by the Sakarya University Scientific Research Projects Unit (Project Number: 2020-9-33-43).

REFERENCES

[1] Mehrabian, A., (1968), Some referents and measures of nonverbal behavior, Behavior Research Methods & Instrumentation, 1, 203–207.

[2] Lopes, A.T., de Aguiar E., De Souza A.F., and Oliveira-Santos T., (2016), Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order, Pattern Recognition, 67, 610-628.

[3] Zalewski, L. and Gong, S., (2005), 2D Statistical Models of Facial Expressions for Realistic 3D Avatar Animation, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2, 217–222.

[4] Yolcu, G., Oztel, I., Kazan, S., Oz, C., Palaniappan, K., Lever, T.E. and Bunyak, F., (2017), Deep learning-based facial expression recognition for monitoring neurological disorders, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1652–1657.

[5] Bartlett, M.S., Littlewort, G., Fasel, I., Chenu, J., Kanda, T., Ishiguro, H., Movellan, J.R., (2003), Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification, Advances in Neural Information Processing Systems, 16.

[6] Zhang, Y. and Hua, C., (2015), Driver fatigue recognition based on facial expression analysis using local binary patterns, Optik - International Journal for Light and Electron Optics, 126(23), 4501–4505.

[7] Shaykha, I., Menkara, A., Nahas, M. and Ghantous, M., (2015), FEER: Non-intrusive facial expression and emotional recognition for driver's vigilance monitoring, 57th International Symposium ELMAR (ELMAR), 233–237.

[8] Kim, J.-B., Hwang, Y., Bang, W.-C., Lee, H., Kim, J.D.K. and Kim, C.-Y., (2013), Real-time realistic 3D facial expression cloning for smart TV, IEEE International Conference on Consumer Electronics (ICCE), 240–241.

[9] Niforatos, E. and Karapanos, E., (2015), EmoSnaps: a mobile application for emotion recall from facial expressions, Personal and Ubiquitous Computing, 19(2), 425–444.

[10] Terzis, V., Moridis, C.N. and Economides, A.A., (2013), Measuring instant emotions based on facial expressions during computer-based assessment, Personal and Ubiquitous Computing, 17(1), 43–52.

[11] Ekman P. and Friesen W.V., (1971), Constants across cultures in the face and emotion., Journal of Personality and Social Psychology, 17(2), 124–129.

[12] Hsu, C.-T., Hsu S.-C. and Huang, C.-L., Facial expression recognition using Hough forest, (2013), 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 1–9.

[13] Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J. and Metaxas, D.N., (2012), Learning active facial patches for expression analysis, 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2562–2569.

[14] Khan, F., (2020), Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks, arXiv:1812.04510.

[15] Saurav, S., Saini, R. and Singh, S., (2021), Facial Expression Recognition Using Dynamic Local Ternary Patterns with Kernel Extreme Learning Machine Classifier, IEEE Access, 9, 120844–120868.

[16] Iqbal, M.T.B., Abdullah-Al-Wadud, M., Ryu, B., Makhmudkhujaev, F. and Chae, O., (2020), Facial Expression Recognition with Neighborhood-Aware Edge Directional Pattern (NEDP), IEEE Transactions on Affective Computing, 11(1), 125–137.

[17] Dagher, I., Dahdah, E. and Al Shakik, M., (2019), Facial expression recognition using three-stage support vector machines, Visual Computing for Industry, Biomedicine, and Art, 2(1).

[18] Nair, V., Hinton, GE., (2010) Rectified Linear Units Improve Restricted Boltzmann Machines, Proceedings of the 27th international conference on international conference on machine learning. Omnipress, USA, 807–814

[19] Ekman, P. and Friesen, W.V., (1978), Facial action coding system: a technique for the measurement of facial movement.

[20] Morales-Vargas, E., Reyes-García, C. A. and Peregrina-Barreto, H., (2019), On the use of action units and fuzzy explanatory models for facial expression recognition, PLoS One, 14(10).

[21] Jin, X., and Jin, Z., (2021), MiniExpNet: A small and effective facial expression recognition network based on facial local regions, Neurocomputing, 462, 353–364.

[22] Fan, X., Jiang, M., Shahid, A.R. and Yan, H., (2022), Hierarchical scale convolutional neural network for facial expression recognition, Cognitive Neurodynamics.

[23] Saurav, S., Saini, A.K., Saini, R. and Singh, S., (2022), Deep learning inspired intelligent embedded system for haptic rendering of facial emotions to the blind, Neural Computing and Applications, 34, 4595-4623.

[24] Happy, S.L., Dantcheva, A. and Bremond, F., (2021), Expression recognition with deep features extracted from holistic and part-based models, Image and Vision Computing, 105, 104038.

[25] Jin, X., Sun, W. and Jin, Z., (2020), A discriminative deep association learning for facial expression recognition, International Journal of Machine Learning and Cybernetics, 11(4), 779–793.

[26] Happy, S. L., Dantcheva, A. and Bremond, F., (2019), A Weakly Supervised learning technique for classifying facial expressions, Pattern Recognition Letters, 128, 162–168.

[27] Luh, G.-C., Wu, H.-B., Yong, Y.-T., Lai, Y.-J. and Chen, Y.-H., (2019), Facial Expression Based Emotion Recognition Employing YOLOv3 Deep Neural Networks, 2019 International Conference on Machine Learning and Cybernetics (ICMLC), 1-7.

[28] Goodfellow, I., Bengio, Y. and Courville, A., 2016, Deep learning, MIT Press.

[29] Krizhevsky, A., Sutskever, I. and Hinton, G.E., (2007), ImageNet Classification with Deep Convolutional Neural Networks, Handbook of Approximation Algorithms and Metaheuristics, 1–1432.

[30] Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V., (2018), Learning Transferable Architectures for Scalable Image Recognition, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8697–8710.

[31] Zhang, X., Zhou, X., Lin, M. and Sun, J., (2018), ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6848–6856.

[32] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., (2015), Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9.

[33] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C., (2018), MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4510–4520.

[34] He, K., Zhang, X., Ren, S. and Sun, J., (2016), Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[35] Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K.Q., (2017), Densely Connected Convolutional Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[36] Krizhevsky, A., Sutskever, I. and Hinton, G.E., (2012), ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, 25.

[37] Simonyan K. and Zisserman A., (2015), Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations.

[38] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T. and van Knippenberg, A., (2010), Presentation and validation of the Radboud Faces Database, Cognition & Emotion, 24(8), 1377–1388.

[39] Face++ Cognitive Services, (2022), Face++ Web Page, https://www.faceplusplus.com, Access Date: May 23, 2022.

[40] Oztel, I., Yolcu, G., Oz, C., Kazan, S. and Bunyak, F., (2018), iFER: facial expression recognition using automatically selected geometric eye and eyebrow features, Jorunal of Electronic Imaging, 27(2), 1.

[41] Nayak, S., Happy, S.L., Routray, A. and Sarma, M., (2019), A Versatile Online System for Person-specific Facial Expression Recognition, TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2513–2518.