Muliti-angled facial emotion detection

Nimantha Weerasooriya, Sutharshan Rajasegarar, and Selvarajah Thuseethan,

***Abstract*—Recognizing emotions from virtual faces is prominent Nowadays. In the near future, it will be crucial for multifarious virtual reality applications. A few researches have been carried out to recognize emotions from frontal virtual faces in recent times. However, recognizing emotions from non-frontal faces with partial occlusions and different emotion intensities is challenging. Addressing these challenges is the main objective of this paper. We have focused on recognizing facial emotions from non-frontal virtual faces by constructing a new image dataset with face images for five face angles (30, 45, 60, 90, 120) for both right and left sides. This dataset contains images with different emotion intensities. In this work, we have followed the transfer learning approach with four popular pertained models specifically inception v3, MobileNetV2, and EfficientNet-B0.**

https://www.overleaf.com/project/60d416f2aaf74424ebdd4f6a

***Index Terms*—Facial expression recognition, Transfer learning, Convolutional Neural Networks, Virtual face images.**

I. Introduction

In daily communication, facial expressions plays a significant role among other human expressions, voices, and body gestures. According to Mehrabian, the ability of emotion conveyance is higher in facial expressions than body gesture and vocal expressions. He has stated that facial emotions have 55% ability to convey emotions while verbal and vocal expressions having 7% and 38% ability to convey emotions respectively [1]. Facial emotion recognition is crucial for various areas such as healthcare, games, teaching, social media social researches, etc. For instance, facial recognition has been used to detect sexual assault victims from college students [2], for game-based learning [3], to detect patients suffering from Vestibular Schwannoma [4], and virtual scenario-based teacher-training [5]. According to an introduction of facial emotions by previous research, there are six basic facial emotions namely happiness, sadness, disgust, fear, anger, and surprise [6].

While most of the existing researches has focused on recognizing emotions from frontal and non-frontal real faces [7],[8], few researches have been conducted to recognize emotions from virtual faces [9],[10]. We have recognized that emotion detection from non-frontal virtual faces as a challenging and important area. Virtual face emotion recognition will be crucial for various virtual reality applications such as games, teaching software, and social media. Moreover, emotion recognition from non-frontal virtual faces can be considered the most challenging task in the emotion detection area.

It can be found some publicly available face emotion datasets to train deep learning models and to recognize emotions. Most of those datasets contain only frontal images. There are few datasets with non-frontal emotion face images. However, non-frontal face image dataset was most challenging. Therefore, we have non-frontal virtual face emotion dataset for our experiments.

The main contribution of this paper is the creation of a new dataset with multi angle virtual faces with different emotion intensities. The data collection was done by using UIBVFED virtual application [11]. The UIBVFED contains only frontal face emotions. We have expanded the dataset by extracting more emotion images with different angles and different emotion intensities. Fig 1 shows the images of different angles for same emotion. We have extracted images of six basic emotions introduced in [6].

The performance assessment of the state-of-the-art networks on our dataset is the second contribution of this paper. We have used three pre-trained deep learning models namely inception v3, MobileNetV2, and EfficientNet-B0 to train these emotion images and compared them.

Fig. 1: Same emotion in different face angles

The rest of this paper is organized as follows. Section 2 contains the details of reviewed related work. Section 3 contains detailed explanations for the data extraction and the description of the selected transfer learning approach. Section 4 describes the experiments and experimental results. Finally, drawn conclusions and future work has described in section 5.

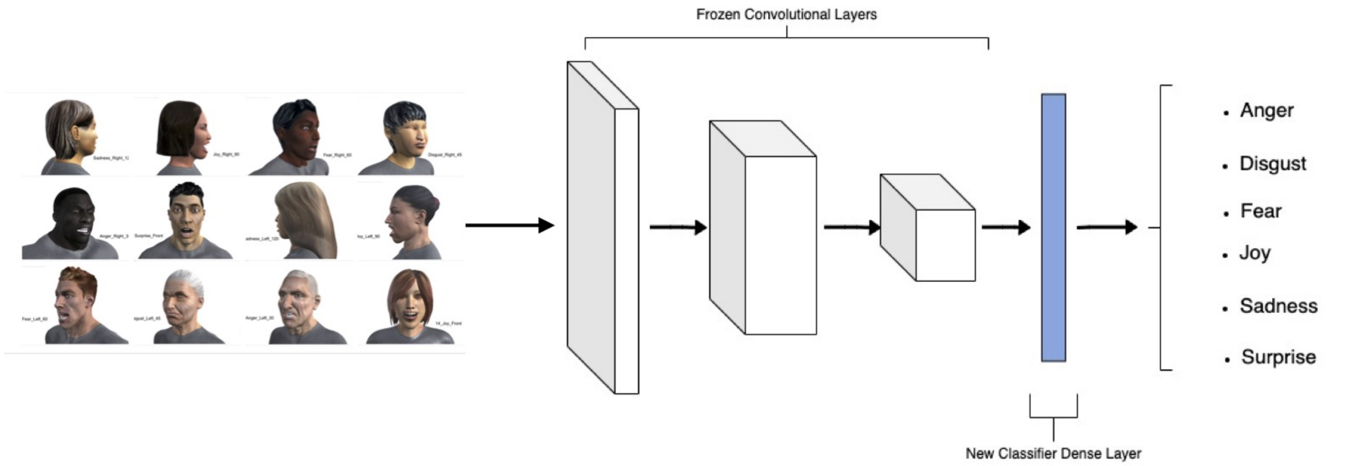

Fig. 2: Transfer learning approach

II. Related work

A. Existing datasets

An emotion face dataset with 213 images of different facial expressions was created by Michael J. Lyons et al. In this dataset, it has used 10 different Japanese female subjects. These 10 different female subjects have done 7 basic expressions (Sadness, Fear, Disgust, Joy, Anger, Surprise, and Neutral).

The extended Cohn-Kanade (CK+) dataset was created by Patrick Lucey et al. with 593 video sequences. Those video sequences were taken from 123 different subjects. Those subjects are in between 18 to 50 years of age. Each video is from the neutral expression to the targeted peak emotion. It has been recorded 30 frames per second from those videos. 327 videos from this datasets were labeled with seven emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise)

UIBVFED dataset contains 660 virtual face emotion images of 20 subjects. those images were labeled with seven emotions (anger, neutral, disgust, fear, joy, sadness, and surprise). Those each emotion has been devided to several emotion sub categories. Emotion anger, disgust, fear, joy, sadness, neutral, surprise have 4,3,3,14,6,1,1 emotion sub categories respectively.

The following problems can be seen in previous datasets.

- Lack of virtual facial expression image datasets. Most of the facial emotion datasets are created using real human face images. There are only a few datasets with virtual emotion face images.
- Inconsistent non-frontal face images. Most of those emotion datasets have only frontal face images. Emotion detection from angled faces can not be done from those datasets.
- Lack of images with different emotion intensities. Most of those datasets contain images with emotion peak. Those datasets do not contain images between neutral and highest emotion intensity.

To address these problems, we have created a multi-angled facial emotion image dataset. We have collected emotion images for 11 different face angles. The emotion image data were collected for 5 different face angles (30, 45, 60, 90, 120) for the left side and the same face angles for the right side with frontal face images.

The main importance of our dataset is having virtual face images with of 20 different subjects for 5 different face angles in both right and left sides with different emotion intensities.

B. Emotion recognition

In previous researches on emotion detection, it has mainly focused on emotion detection from real face images and frontal face images. Emotion recognition from virtual faces and non-frontal faces has been considered as a challenging task. It can be found various deep learning methods in previous researches that have been used to recognize emotions from frontal faces and few studies have been conducted to detect emotions from non-frontal and occluded face images.

It can be found some previous researches that have done to recognize facial emotions using deep learning techniques. In [12], researchers have tried to recognize facial expressions from videos. They have used CNN and feature aggregation for the analysis. They have designed the temporal and spatial CNN network based on the AlexNet architecture and trained with image frames captured from emotion videos. Then they have tested the model with five datasets RML, BAUM-1 s, eNTERFACE05,MMI, and FER2013. The performance of spatial CNN is higher than temporal CNN for all datasets. Another research [13] has used an attentional convolutional neural network to recognize emotions from face images. To pay more attention on the specific region of the face to get the sense of the emotion, they have added an attention mechanism by spatial transformer network. The model has trained validated and tested with five datasets FER2013, CK+, JAFFE, and FERG. They were able to achieve 70.02%, 99.3%, 92.8%, and 98% for the datasets FER2013, FERG, JAFFE, and CK+ respectively. In [14], it has been proposed a convolutional neural network for emotion recognition. In their model, they have used convolutional and downsampling layers. They have done face detection from the JAFFE dataset before training the model. The emotion recognition rate is 100% for the proposed model and the model has performed well compared to the existing methods. The research [15] is about facial emotion recognition from a highly imbalanced dataset using transfer learning and weighted-cluster loss. In this research, they have used a pre-trained ResNet-50 model which was pre-trained on VGGFace2 data and the weighted cluster loss at the output layer. The model has been fine-tuned with AffectNet data. The proposed method achieved 60.7% accuracy and it is the best result compared to other experiments in this research.

Kim et al. [16] have proposed a convolutional neural network to recognize real-time patient's facial expressions during radiotherapy. They have used CK+ dataset to train the model and tested it with real patient images. By this proposed model, they have achieved 85.6% test accuracy. It can be found a proposal of the convolutional neural network to facial emotion recognition in [17]. They have combined the parallel siamese architecture and CNN in this experiment. CK+ and JAFFE datasets have been used to evaluate the proposed FER-CNN model. When the results of the proposed method get compared with existing methods, it can be seen that the proposed method has achieved a high accuracy of 92.06%. Personalized Movie Summarization has been done by facial expression recognition with CNN in [18]. In this research, they have done shot segmentation from a movie clip and they have produced a set of images by face detection. They have used the transfer learning approach for emotion recognition in this experiment. They have used MobileNet, SqueezNet, AlexNet, GoogleNet, and ResNet-50 with transfer learning and they have trained the ResNet-50 model from scratch. They have used VGG Face dataset to train ResNet-50 and KDEF for transfer learning. Their model has achieved 93.65% accuracy. It was the best method among other tested methods.

Another research [19] has proposed a convolutional neural network with attention mechanism for facial emotion recognition with partial occlusions. They have used three in-the-wild datasets, RAF-DB, AffectNet, SFEW, and three in-the-lab datasets, CK+, MMI, and Oulu-CASIA for this experiment. For evaluation, they have synthesized occluded images and used the FED-RO dataset with real occlusions. They have achieved 85.07%, 80.54% accuracy for RAF-DB clean and occluded images respectively, and 58.78%, 54.84% accuracy for AffectNet clean and occluded images respectively. Their proposed method has high accuracy than other tested and existing methods.

It can be seen that non-existence of emotion recognition from virtual faces as a limitation of previously conducted researches. In [9], they have done feature selection to improve facial expression recognition. They have used principal component analysis and genetic algorithm methods for future extraction. They have used Bosphorus and UIVBFED datasets for these experiments and achieved 86.62% and 93.92% median accuracy on these datasets respectively. In [10], a deep CNN has been used with ensemble learning to recognize emotions from virtual faces. They have proposed DCNN, DCNN-SVM, and DCNN-VC models in this research. For this experiment, they have used UIVBFED, FERG, JAFFE, CK+, and TFEID datasets. They have compared their results of the proposed method with some existing work and pre-trained models VGG19, ResNet-50 with voting classifies. The results of their method DCNN-VC achieved high accuracy on all datasets compared to other existing methods and used pre-trained models.

For facial emotion detection, a CNN-LSTM deep neural network has used in the research [20]. They have used CREMA-D dataset. The proposed model has CNN layers, LSTM layers and soft max layer. They have done several experiments by changing no of CNN layers, learning rate and number of epochs. The model has achieved highest test accuracy, 78.5% for 6 CNN layers, 50 epochs and 0.0001 learning rate. In the research [21], they have used a novel classifier called DR classifier with deep learning neural networks to recognise seven basic emotions from face images. They have used CK+ dataset and JAFFE dataset for this experiment. The proposed method has achieved 95.23% classification accuracy for JAFFE dataset and It has achieved 100% accuracy for CK+ dataset. Emotion recognition using transfer learning has done in [22]. They have used four pre trained models, Resnet50, VGG19, Inception V3 and Mobile Net for this experiment. Moreover, they have compared their results with existing work. ImageNet database has used to train those pre trained models and tested with CK+ database. Pre trained models VGG19, Resnet50, Inception V3 and MobileNet has achieved accuracies 96%, 97.7%, 98.5% and 94.2% respectively.

III. METHOD

A. Data collection

Due to the difficulty of finding a non-frontal virtual emotion face dataset, we have constructed a new dataset with non-frontal emotion face images with different emotion intensities. To construct this dataset, we have used UIBVFED virtual application [11]. This virtual application has 20 different faces and 32 different subcategories of 6 basic emotions. Table I shows the number of emotion subcategories of the dataset.

TABLE I: Emotion sub categories

Emotion	Number of emotion sub categories
Anger	4
Sadness	6
Disgust	3
Joy	14
Fear	4
Surprise	1

Due to the importance of the non-frontal virtual image dataset, we have collected emotion images for 11 different face angles. The emotion image data were collected for 5 different face angles (30, 45, 60, 90, 120) for the left side and the same face angles for the right side with frontal face images. First, we have created small video clips of each emotion for 20 different faces by screencasting. The video is from low emotion intensity to high emotion intensity. Then we have extracted image frames from those created videos. The image frame was captured every 0.05 second of the video. Virtual face emotion image dataset was created using these image frames. Fig 3 shows the image frames of surprise emotion from low intensity to high intensity.

Fig. 3: Surprise emotion from low intensity to high intensity

The dataset was created using image frames of different emotion intensities for 32 emotion subcategories and 20 virtual faces for each angle. Final dataset contains 60582 virtual face images. Table II and III show the datasets that we have created for different angles.

We have combined right and left images of each angle and created 6 sub datasets based on the angle. Before training pre-trained models with this dataset, we have splitted the dataset into 3 sets. We have created a training set with 60% of images and validation set and test set was created using 20% of images for each set. Due to the imbalance of the subcategories of emotions, the dataset has become highly imbalanced. Joy has 14 subcategories while surprise having only one subcategory. Fig 4 shows the composition of emotion images.

The majority images of the dataset is from joy emotion. it is 45%. Surprise has the lowest percentage of images in the dataset. Surprise emotion has only 3% of images. As this data imbalance leads to a biased results when model training, we had to re-sample the dataset to construct a balanced dataset.

TABLE II: Number of images

Angle	Number of images
Front	5605
Left 30	6267
Left 45	5342
Left 60	5224
Left 90	5256
Left 120	5275
Right 30	5751
Right 45	5381
Right 60	5335
Right 90	5531
Right 120	5615

TABLE III: images of combined dataset

Angle	Number of images
Front	5605
30	12018
45	10723
60	10559
90	10787
120	10890

We have used image augmentation for expand the images of classes with low number of images. For image augmentation, we have used rotation, width shift, height shift, shear and zoom methods. We have re-sampled the image classes in training set and created a balanced dataset to avoid building a biased model. Fig 5 shows the augmented images of surprise emotion.

B. Transfer learning

We propose transfer learning approach for recognizing facial emotions from our virtual image dataset. We have used three pre-trained deep learning models, MobileNetV2, InceptionV3, and EfficientNetB0 to find out the best model for emotion recognition from virtual faces. 60% of images were used for model training, 20% of images were used for validating and the rest 20% of images were used for model testing. We have re-sized images to [224,224] before train models. 'Imagenet' weights were used to train the model and we have used the 'softmax' activation function with six output nodes for the prediction layer. We have used 'Adam' as the optimizer with a 0.0001 learning rate. The early stopping callback was used by monitoring validation loss with a 0.001 minimum delta value.

We have trained six datasets with different face angles (front, 30, 45, 60, 90, 120) on these selected pre-trained models. We have used 30 epochs for training.

IV. Results

In this section we are presenting the results obtained by training Front, 30, 45, 60, 90, 120 face angle datasets on selected pre-trained models MobileNetV2, InceptionV3, and

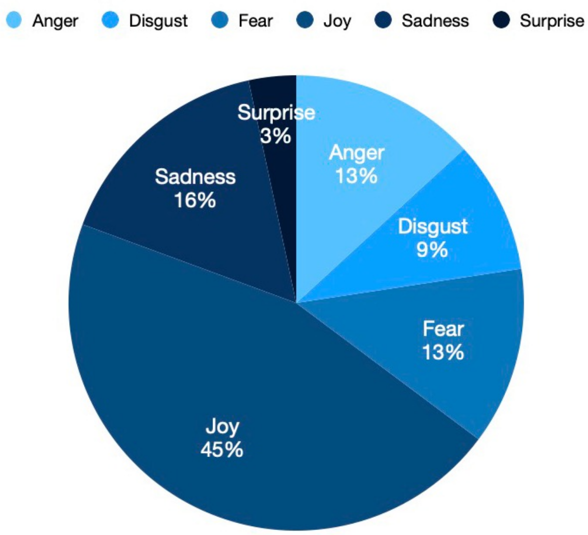

Fig. 4: Composition of emotion images

Fig. 5: Augmented images of surprise emotion

EfficientNetB0. It can be seen in each angle MobileNetV2 pre-trained model has performed well than the other two networks InceptionV3 and EfficientNetB0. In Fig. 6, we can see the test accuracy of MobileNetV2 achieved the highest accuracy in each face angle. InceptionV3 has achieved the second-highest test accuracy. EfficientNetB0 has achieved the lowest test accuracy compared to two other models.

Table IV - Table IX shows the confusion matrices of MobileNetV2 in each dataset with different face angles. It

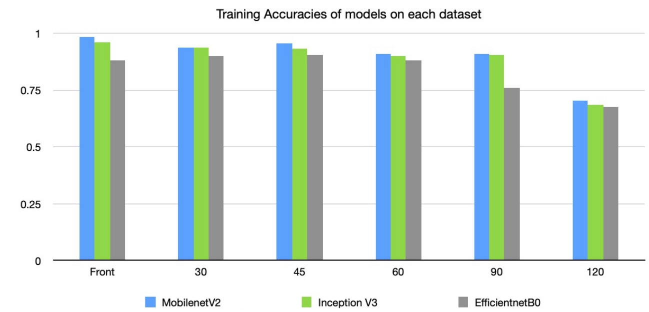

Fig. 6: Accuracies of selected models on six datasets

shows the classification accuracy of MobileNetV2 for all six emotions in percentages.

TABLE IV: Confision matrix of MobileNet for front images

	AN	DI	FE	JO	SA	SU’
AN	100%	0	0	0	0	0
DI	0	91%	0	0	9%	0
FE	0	0	98%	0	2%	0
JO	0	0	0	100%	0	0
SA	0	0	0.4%	0	99.6%	0
SU	0	0	0	0	0	100%

TABLE V: Confision matrix of MobileNet for angle 30 images

	AN	DI	FE	JO	SA	SU
AN	95%	0	0	1%%	0.4%	3.6%
DI	0	90%	1%	8.5%	0	0.5%
FE	4%	1.1%	83.5	4.6%	0.4%	6.4%
JO	0.8%	0	0	99%	0	0.2%
SA	6.7%	0	0	1.6%	91.3%	0.4%
SU	10.3%	0	6.5%	1.3%	0	81.9%

TABLE VI: Confision matrix of MobileNet for 45 images

	AN	DI	FE	JO	SA	SU
AN	100%	0	0	0	0	0
DI	0	98%	2%	0	10	0
FE	0	1.4%	98%	0	0.6	0
JO	3.2%	0.4%	2.9%	92.8%	0.7%	0
SA	0	0.5%	1.5%	0	98%	0
SU	0	0	0	0	15%	85%

According to the confusion matrices, it can be seen that the model MobileNetV2 has classified images for front angle better than other angles. For images in angles 60, 90, 120 the model has not given the best emotion classification. However, the emotion Joy has been classified very well in every angle for MobileNetV3. Fig. 7 shows the emotion classification accuracy by angles.

TABLE VII: Confision matrix of MobileNet for 60 images

	AN	DI	FE	JO	SA	SU
AN	70.2%	0	0	9.2%	19.3%	1.3%
DI	0.55%	55.74%	0.55%	15.3%	27.86%	0
FE	0.3%	0	89%	1.8%	4.6%	4.3%
JO	0.1%	0.1%	0.1%	98.96%	0.42%	0.32%
SA	0.5%	0	0.96%	1.44%	96.6%	0.5%
SU	0	0	12.1%	1.7%	0	86.2%

TABLE VIII: Confision matrix of MobileNet for 90 images

	AN	DI	FE	JO	SA	SU
AN	86.5%	1.3%	3.4%	0	8.8%	0
DI	8.5%	80.9%	2.1%	0	8.5%	0
FE	8.3%	0.75%	81.15%	0	9.8%	0
JO	0.2%	0.4%	0.7%	98.3%	0.4%	0
SA	7%	2%	2%	0	89%	0
SU	0	3.5%	0	0	0	96.5%

TABLE IX: Confision matrix of MobileNet for 120 images

	AN	DI	FE	JO	SA	SU
AN	37.8%	0.8%	27.4%	25.5	7	1.5
DI	8.9%	37.1%	14.7%	28.8%	9.5%	1%
FE	3.7%	0	75.3%	16.6%	2.6%	1.8%
JO	0.2%	0	3%	96%	0.5%	0.3%
SA	8.7%	1.1%	22.2%	20.4%	47.1%	0.5%
SU	4.6%	0	30.8%	13.8%	0	50.8%

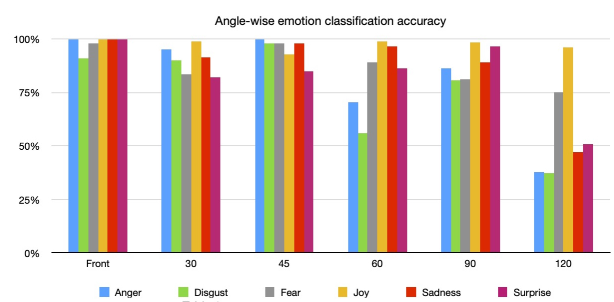

Fig. 7: Angle-wise emotion classification

According to Fig. 7, it can be seen that the images of Joy emotion have classified well in every angle. However, the images of angle 120 have not classified well compared to the other five angles. When we consider the emotion classification of the front angle, it has the best classification compared to other angles.

V. Conclusion

Overall, according to our experimental results, we can conclude that the MobileNetV2 is the best model to recognize facial emotions from virtual faces for different angles. Moreover, it can be concluded that the emotion Joy has the best classification accuracy for every angle and angle 120 has some difficulties with emotion classification. The inter-class similarity of image frames and inconsistency of emotion in some images in angle 120 has affected the classification problem in angle 120. In the future, it can be considered those limitations and by fixing those problems, we will be able to achieve a good classification in images of angle 120 as well.

VI. Abbreviations

The following abbreviations are used in this manuscript.
AN - Anger
DI - Disgust
FE - Fear
JO - Joy
SA - Sadness
SU - Surprise

References

[1] Mehrabian G (2007) Nonverbal communication. Aldine, New Brunswick, NJ, USA.

[2] A. Melkonian, L. Ham, A. Bridges and J. Fugitt, "Facial emotion identification and sexual assault risk detection among college student sexual assault victims and nonvictims", Journal of American College Health, vol. 65, no. 7, pp. 466-473, 2017. Available: 10.1080/07448481.2017.1341897.

[3] M. Ninaus et al., "Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data", Computers Education, vol. 142, p. 103641, 2019. Available: 10.1016/j.compedu.2019.103641.

[4] S. Blom, H. Aarts, H. Kunst, C. Wever and G. Semin, "Facial emotion detection in Vestibular Schwannoma patients with and without facial paresis", Social Neuroscience, pp. 1-10, 2021. Available: 10.1080/17470919.2021.1909127.

[5] S. Park and J. Ryu, "Exploring Preservice Teachers' Emotional Experiences in an Immersive Virtual Teaching Simulation through Facial Expression Recognition", International Journal of Human–Computer Interaction, vol. 35, no. 6, pp. 521-533, 2018. Available: 10.1080/10447318.2018.1469710.

[6] Ekman P, Friesen WV, O'Sullivan M, Chan AYC, Diacoyanni-Tarlatzis I, Heider KG, Krause R, LeCompte WA, Pitcairn T, Bitti PER (1972) Universals and cultural differences in facial expressions of emotion. J Pers Soc Psychol 53(4):712–717.

[7] M. Akhand, S. Roy, N. Siddique, M. Kamal and T. Shimamura, "Facial Emotion Recognition Using Transfer Learning in the Deep CNN", Electronics, vol. 10, no. 9, p. 1036, 2021. Available: 10.3390/electronics10091036.

[8] V. Tumen, O. Soylemez and B. Ergen, "Facial emotion recognition on a dataset using convolutional neural network", 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017. Available: 10.1109/idap.2017.8090281 [Accessed 2 June 2021].

[9] V. Perez-Gomez, H. Rios-Figueroa, E. Rechy-Ramirez, E. Mezura-Montes and A. Marin-Hernandez, "Feature Selection on 2D and 3D Geometric Features to Improve Facial Expression Recognition", Sensors, vol. 20, no. 17, p. 4847, 2020. Available: 10.3390/s20174847.

[10] V. Chirra, S. Uyyala and V. Kolli, "Virtual facial expression recognition using deep CNN with ensemble learning", Journal of Ambient Intelligence and Humanized Computing, 2021. Available: 10.1007/s12652-020-02866-3.

[11] M. Oliver and E. Amengual Alcover, "UIBVFED: Virtual facial expression dataset", PLOS ONE, vol. 15, no. 4, p. e0231266, 2020. Available: 10.1371/journal.pone.0231266.

[12] R. Gupta and L. Vishwamitra, "Facial expression recognition from videos using CNN and feature aggregation", Materials Today: Proceedings, 2021. Available: 10.1016/j.matpr.2020.11.795.

[13] S. Minaee, M. Minaei and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", Sensors, vol. 21, no. 9, p. 3046, 2021. Available: 10.3390/s21093046.

[14] E. Taee and Q. Jasim, "Blurred Facial Expression Recognition System by Using Convolution Neural Network", Webology, vol. 17, no. 2, pp. 804-816, 2020. Available: 10.14704/web/v17i2/web17068.
[15] Q. Ngo and S. Yoon, "Facial Expression Recognition Based on Weighted-Cluster Loss and Deep Transfer Learning Using a Highly Imbalanced Dataset", Sensors, vol. 20, no. 9, p. 2639, 2020. Available: 10.3390/s20092639.
[16] K. Kim et al., "Facial expression monitoring system for predicting patient's sudden movement during radiotherapy using deep learning", Journal of Applied Clinical Medical Physics, vol. 21, no. 8, pp. 191-199, 2020. Available: 10.1002/acm2.12945.
[17] S. Xie and H. Hu, "Facial expression recognition with FRR-CNN", Electronics Letters, vol. 53, no. 4, pp. 235-237, 2017. Available: 10.1049/el.2016.4328.
[18] I. Ul Haq, A. Ullah, K. Muhammad, M. Lee and S. Baik, "Personalized Movie Summarization Using Deep CNN-Assisted Facial Expression Recognition", Complexity, vol. 2019, pp. 1-10, 2019. Available: 10.1155/2019/3581419.
[19] Y. Li, J. Zeng, S. Shan and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism", IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439-2450, 2019. Available: 10.1109/tip.2018.2886767.
[20] A. Hans and S. Rao, "A CNN-LSTM BASED DEEP NEURAL NETWORKS FOR FACIAL EMOTION DETECTION IN VIDEOS", INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES, vol. 7, no. 1, pp. 11-20, 2021. Available: 10.29284/ijasis.7.1.2021.11-20.
[21] Anjani Suputri Devi D and Satyanarayana Ch, "An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier", Multimedia Tools and Applications, vol. 80, no. 12, pp. 17543-17568, 2021. Available: 10.1007/s11042-021-10547-2.
[22] M. Chowdary, T. Nguyen and D. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications", Neural Computing and Applications, 2021. Available: 10.1007/s00521-021-06012-8.