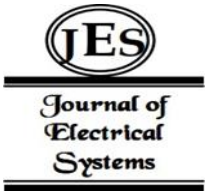
[1]Manisha Balkrishna Sutar

[2]Dr. Asha Ambhaikar

Innovative Facial Detection and Emotion Recognition System Utilizing Correlation Attention Module in Deep Convolutional Neural Networks



Abstract: - Facial detection and emotion recognition are pivotal in human-computer interaction, facing numerous challenges such as varying illuminations and occlusions. Despite efforts by multiple researchers, labeling diverse emotions remains a challenge, necessitating improvements in existing models. This study aims to introduce a deep learning-based facial emotion detection model capable of accurately discerning emotions from images. Leveraging a deep Convolutional Neural Network (CNN) classifier, facial expressions are identified with precision, thanks to effective feature learning from images. Additionally, a correlation attention module is devised to enhance the classifier's efficacy by establishing relationships between features extracted by Residual Network 101 (ResNet 101) and VGG 16. Evaluated on CK+48 and Japanese Female Facial Expression (JAFFE) datasets, the face detection model achieves efficiencies of 94.46%, 95.00%, and 92.19%, while the emotion detection model scores 98.73%, 98.10%, and 99.55%, and 98.73%, 98.10%, and 98.55% respectively, in terms of accuracy, sensitivity, and specificity.

Keywords: Emotion recognition, Face detection, deep Convolutional Neural Network (deep CNN) classifier, Human-computer interaction, Feature extraction.

1. Introduction

Facial expression detection is essential in computer vision-based applications such as human-computer interfaces, video interaction, cataloging, biometrics, including image recovery, security, etc. [1]. The ability to recognize human emotions is significantly influenced by facial expressions [2] [3]. Particularly, the application of machine and deep learning for enhanced performance in user authentication and general cyber security measures has increased dramatically [25]. Human emotional states are usually deciphered through facial expressions. The most common facial expressions, such as anger, contempt, fear, happiness, grief, and surprise, are used by many facial expression recognition systems even though other research suggests that emotional expressions on the face are culture-specific. Complex facial expressions—combinations of simple facial expressions—have also recently been examined and classified [4] [5]. For many interactive computing areas, including human-computer/machine interface, human-robot interaction, and human-AI engagement, the ability to recognize facial expressions is crucial. Prior research on facial expressions generally made use of a collection of human faces photographed in a controlled environment, resulting in various in-the-wild restrictions on the application. To realize the production of trust, understanding, and closeness between humans and robots in real-world contexts, a number of studies concentrating on effective behavior analysis in-the-wild have recently been introduced [22]. Important indicators of emotional states and intentions can be found in facial expressions. However, using the traditional methods for recognizing facial expressions might be challenging when dealing with complicated backgrounds and irrelevant face parts like the hat and the spectacles [6] [7] [8]. Although existing facial expression recognition has shown excellent results in controlled settings, in real-world dataset performance is still lacking. This is due to the wide range of factors that affect facial appearances, including skin tone, lighting, etc. The facial expression identification system should have high accuracy in both

[1] [1]PhD Scholar, Dept of CSE, Kalinga University,

Raipur, Chhattisgarh, India

[1]*more.manisha3030@gmail.com

[2] Professor, Dept of CSE, Dean Students Welfare, Kalinga University,

Raipur, Chhattisgarh, India

[2]asha.ambhaikar@kalingauniversity.ac.in

finding the exact image from the dataset that corresponds to the input image and identifying the expression. Additionally, facial image retrieval needs to be quick and reliable [2] [9-14].

The major objective is to create a deep learning-based facial expression identification model that accurately deciphers an individual's emotion from their face image. The input is collected and preprocessing is performed. Then the features using the correlation attention module and the hybrid textural pattern are extracted, where the features from the correlation attention module are used for face detection, and the concatenated features are used for emotion identification. The deep CNN classifier improved the performance by properly identifying the emotions expressed on the faces and the contributions are interpreted as follows:

- ***Deep CNN*** : Deep CNN effectively identifies emotions by detecting the face with high accuracy and reducing the complexities. The features are effectively learned and the training is provided with high efficiency, which provided better results.

The remaining section of the research is described as follows: Section 2 reviews the related efforts and examines their challenges. The architecture of deep CNN and the methodology for face detection and emotion classification are described in Section 3. The last phase of the work is described in Section 4.Motivation

The following part provides an analysis of the numerous research studies based on the emotion detection model that was carried out utilizing various methods.

2.1 Literature Review

2.1 The existing works carried out by the various researchers are interpreted as follows: to recognize human facial expressions, Li Yao et al. [1] presented active learning and Support Vector Machine (SVM) algorithms that categorized the facial action units. This model's inexpensive cost and enhanced training speed offset its slower classifier training speed. It also lowered the cost of labeling samples. H. Sikkandar and R. Thiyagarajan [15] developed a method of Improved Cat Swarm Optimization (ICSO) that identified the person’s emotional state through facial expression. This method improved the retrieval performance and reduced the computation time and achieved superior accuracy, but it consumed too much processing time and had slow convergence [16] [17]. Rekha Bhatia and Naveen Kumari [24] introduced a novel deep learning-based technique for facial emotion recognition that reduced the noise, and a joint trilateral filter was used that provided smooth edges. Using the face emotions dataset, the images are made noise-free. After that, 200 CLAHE was applied to the filtered images to increase their visibility. Jun Liu et al. [2] executed three structural methods for recognizing the face expression, which reduced the intra-class difference and improved the recognition performance. The disadvantage of the method is face recognition model in an unconstrained environment is not ideal and the misclassification of similar expressions is high. Adil Boughida [21] developed a novel face expression identification technique based on the Gabor generic algorithm. Gabor characteristics are taken from the areas of the human face that are of interest and are identified by facial landmarks. However, there are a number of restrictions on this research, including the sluggish convergence of genetic algorithms for datasets with a lot of features. This research presented a revolutionary face expression recognition system by Saiyed Umer et al. [23]. The method mainly concentrated on the identification of various expressions that can be seen on the human face. A facial region was identified during implementation from each collected image. The color facial image was then converted into a grayscale version to speed up the operations. This approach reduces noise artifacts and eliminates overfitting, however, data only performs so well.

2.2 Challenges

The challenges to be overcome in the research are,

Because of pose fluctuation and perspective modification, the accuracy of the recognition of the expressions gathered in a complicated context is lower [8].In most tasks with significant interference, hand-crafted features employed in conventional approaches cannot produce good outcomes [8].

Due to data noise and in-class similarity, certain traditional methods based on CNN experience misclassification of similar expressions [8]. The current CNNs framework can distinguish between many aspects of an image, such as eyes, nose, and mouth. When modeling complicated issues, however, end-to-end renders model learning unreliable, and it is only possible to obtain a good model after extensive training using data [18]. A technique based on deep learning utilizes the best features from the face and carries out classification. For facial expression recognition, however, it is difficult to gather a significant amount of training data under a variety of circumstances and thus necessitates more intensive processing. Therefore, it is vital to shorten the deep learning algorithm's computation time [2].

1. Proposed Methodology of Face Detection and Emotion Classification Model

Problem definition: The main aim is to be aware of the facial emotion classification based on the deep learning-based facial emotion identification model developed for the proper classification of an individual's emotion from the image. The data for the face detection model is initially collected from the CK+ dataset [20]. The equation below describes the data gathered from the standard repository.

3

$D \square \square D_d \, d \square 1$ (1)

Here, the database is denoted by D, and the number of images is represented as d. The emotions of the input data are detected using deep learning, for which the faces are detected from the input images acquired from the dataset. Let us suppose E be the emotions identified for the detected faces from the input image. The image is collected and the preprocessing step is performed that removes the noisy data from the input image. The features are extracted using a correlation attention module. The correlation attention module extracts the features using ResNet101 and VGG16, and the relationship between the features is also identified by this module. The features extracted from the correlation model are used for the recognition of the human face present in the image. On the other hand, the image's global and local features will be retrieved from the hybrid textural patterns. The features from the correlation attention module and the hybrid textural pattern are concatenated and fed forward to the deep CNN classifier for recognizing the emotion present in the image. Figure 1 shows the diagrammatic representation of the model.

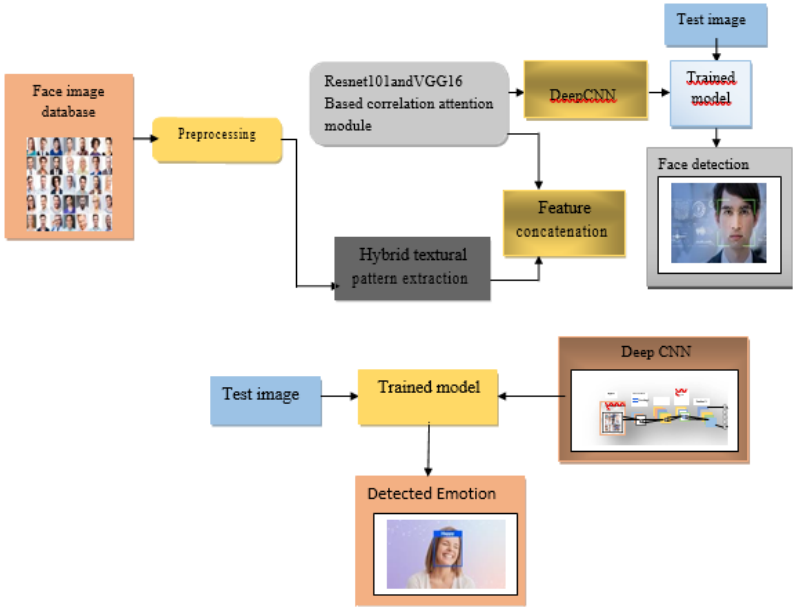


Fig. 1. Proposed Facial emotion recognition model

3.1 Preprocessing

3.2 Preprocessing is necessary due to the varying scales and orientations of the images. Because the scale and angle of the face might vary from image to image, the face detection procedure is extremely difficult. Due to the motions of the person being observed, faces may appear in images at varied sizes and angles if the shots were recorded with a stationary camera. Therefore, looking for a set pattern in the image is challenging. It is challenging to do this job when noise is present and face images are obscured. Complex backdrops and a wide range of lighting conditions in the input image might make it difficult to identify emotions. If the test image has a different lighting condition than the training images, facial expression recognition is likely to fail. A preprocessing phase is necessary because face points can be erroneously recognized if the illumination is not uniform.

3.3 Feature extraction : Correlation Attention Module and texture features :

Initially, the features from the images are recognized using the correlation attention module, where the features are extracted using ResNet101 and VGG16 and the relationship between the features is also identified by this correlation attention module. The correlation attention module finds similarities between the features of faces to accurately localize targets, for which the features using ResNet101 and VGG16 are used, and the relationship between the features is established to enhance the accuracy of the pixel-wise correlation encoding. The spatial information is exploited using the spatial-wise attention as well as the channel-wise attention to encode the channel-wise correlation. Furthermore, the extracted features from the correlation model are fed forward to the deep CNN classifier that accurately identifies the human face present in the image.

The texture of the image is described using a range of texture descriptors and the single texture descriptor is often provided by a texture-based descriptor. It is enhanced by a hybrid texture pattern description in this research, which helps in deep feature extraction. In emotion detection, the hybrid textual pattern plays a significant role, referring to the description of the image texture using more than one texture descriptor. To extract the texture features of an image, the hybrid patterns LBP, LDP, and LTP are used. The Local Binary Pattern (LBP) is used for the classification of texture in the center pixel of the image. This value can be evaluated by comparing the grey level of its neighbor levels to their code as three values named local ternary pattern (LTP). The LTP-attained pattern is further coded from the upper and lower binary pattern. The Local Directional Pattern (LDP) captures the texture of each position of the pixel to stabilize the threshold to generate the features.

The correlation attention module and the hybrid textual patterns both are concatenated, and these concatenated features are fed forward to the deep CNN classifier to develop a trained model which detects the emotion effectively without any complexity..

3.4 Face detection and emotion classification using deep CNN classifier

The features extracted from the correlation attention module are fed forward to the deep CNN classifier that will identify the face present in the image with high efficacy. Then the concatenated features are used for recognizing the facial emotion. A deep CNN classifier is used for the detection of the emotion present in the face. It is utilized to automatically recognize the key features and categorize the data. The architecture of the deep CNN classifier is depicted in Figure 2.

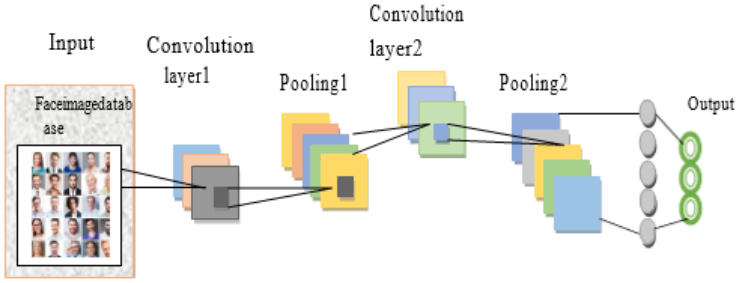


Fig.2. Architecture of deep CNN

4. Result and Discussion

4.1 The model's accomplishments are evaluated against those of alternative methods and discussed below.

4.2 Experimental Setup

The emotion detection is run in Python using the 8GB internal memory and Windows 10 OS, and the detailed interpretation of the results achieved is shown in the below sections.

4.3 Experimental Result

The various emotions recognized by the deep CNN classifier on the preprocessed images are provided in Figure 3**.**

	Contempt	**Disgust**	**Fear**
Input image			
Pre processed image			
Outputimage			

Fig.3. Experimental result using the proposed model

4.4 Dataset Description

4.4.1 Extended Cohn-Kanade (CK+48)

4.5 A small dataset called CK+48 [20], Facial Expression Recognition 2013 [19], contains 981 images divided into seven types. Images are 48 x 48 in size and have a color scheme of grey scale. To acquire a decent distribution and normalize the degree of data variation, other classes might benefit from the variations and feature distributions of the classes during the merging phase. Images extracted from video frames typically exhibited minimal variance, and the overall number of elements is insignificant when compared to other datasets. Images are in frontal view and have a clear pattern for facial expression as opposed to the FER-2013.

4.6 ComparativeAnalysisBasedOnCK+48For Face Detection

The performance of the deep CNN is comparatively analyzed with the evaluation metrics, including accuracy, sensitivity, and specificity, which are shown in Figure 4. The improvement of face detection based on training percentage using the CK+48 dataset as compared with the other existing methods is shown below figures.

Figure 4a) illustrates the outperformance of deep CNN with a greater improvement in face detection of 0.01% when compared with the ANN approach, and the accuracy rate is 94.46% achieved with the training percentage at 90.

Figure 4b) illustrates the achievements of face detection in deep CNN, as 0.09% is greatly improved when compared with ANN, and the sensitivity rate is 95.00% achieved with the training percentage at 90.

Figure 4c) illustrates the face detection improved in deep CNN, as 7.57% is effectively performed with ANN, which has a specificity rate of 92.19% achieved with the training percentage of 90.

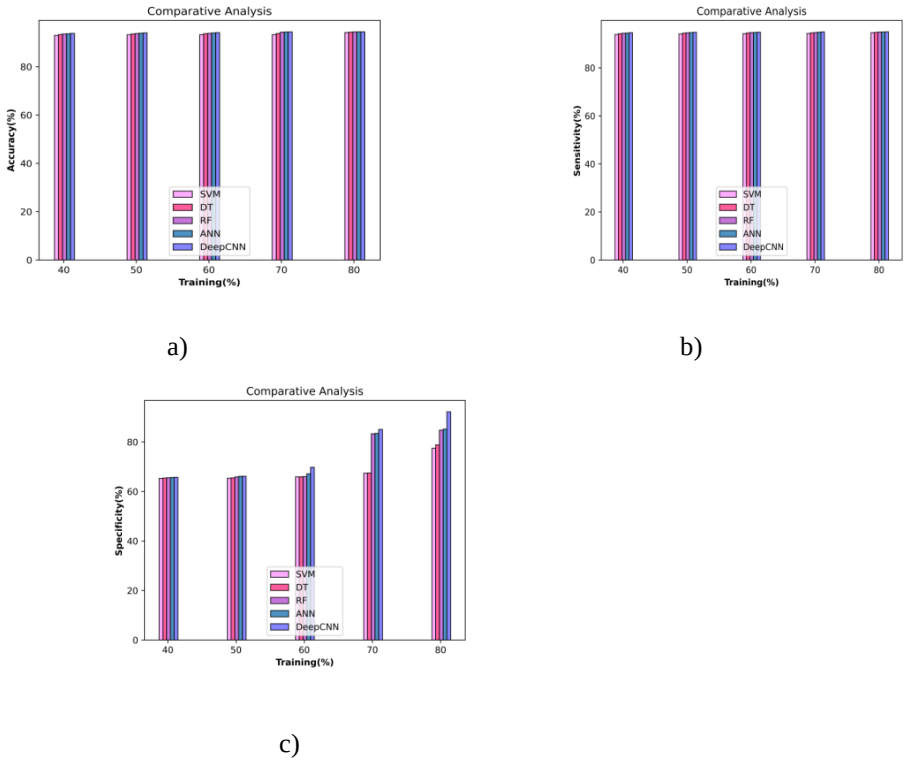


Fig.4. Comparative analysis for face detection model based on CK+48 database a) Accuracy b) Sensitivity and c) Specificity

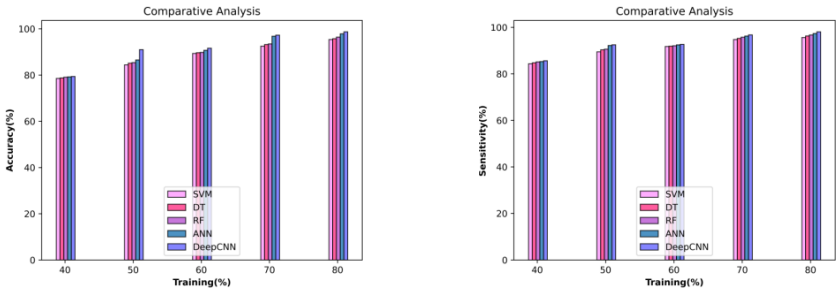
4.7 Comparative Analysis Based On CK+48 For Emotion Detection

The performance of the deep CNN is comparatively analyzed with the evaluation metrics, including accuracy, sensitivity, and specificity, which are shown in Figure 5. The improvement of emotion detection based on the training percentage using the CK+48 dataset as compared with the other existing methods is shown below figures.

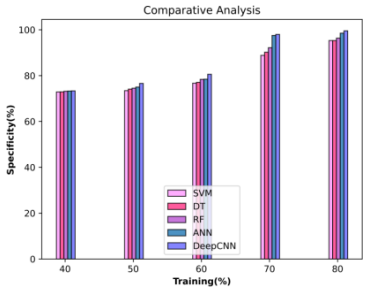
Figure 5a) illustrates the outperformance of deep CNN with a greater improvement in emotion detection of 0.90% when compared with the ANN approach, and the accuracy rate is 98.73% achieved with the training percentage at 90.

Figure 5b) illustrates the achievements of the emotion detection in deep CNN as 0.76% is greatly improved when compared with ANN, and the sensitivity rate is 98.10% achieved with the training percentage at 90.

Figure 5c) illustrates that emotion detection is improved in deep CNN as 1.04% is effectively performed with ANN, which has a specificity rate of 99.55% achieved with a training percentage of 90.



a) b)



c)

Fig. 5. Comparative analysis for emotion detection model based on CK+48 database a) Accuracy b)sensitivityandc)specificity

4.8 Comparative Discussion

In this part, the outcomes of the various emotion recognition models are discussed extensively. The approaches used for comparative evaluation include SVM, DT, RF, and ANN. The comparison explanation in the table demonstrates that deep CNN performs best in terms of specificity, accuracy, and sensitivity.

Table1. Comparative discussion

Database								
	Face detection				**Emotion detection**			
Model	Accuracy	Recognition error	Sensitivity	Specificity	Accuracy	Recognition Error	Sensitivity	Specificity
SVM	94.16	5.84	94.61	77.49	95.39	4.61	95.66	95.30
DT	94.30	-94.3	94.74	78.80	95.72	-95.72	96.32	95.31
RF	94.40	-94.4	94.88	84.79	96.45	-96.45	96.73	96.36
ANN	94.46	-94.46	94.91	85.21	97.83	-97.83	97.35	98.51
DeepCNN	94.46	-94.46	95.00	92.19	98.73	-98.73	98.10	99.55

2. Conclusion

The recognition of emotional expressions using deep learning models has been widely discussed in recent years. Although it's a difficult task, recognizing emotions still seems like a vast space for improvement. Our methodology emphasizes aspects and incorporates emotion recognition from facial expressions to recognize user feelings and give suitable solutions. In this work, the deep CNN classifier accurately identified emotions from facial expressions and gave high accuracy in the analysis and detection of emotions from facial expressions. Based on the performance in accuracy, sensitivity, and specificity using CK+48, the efficiency of the face detection model is 94.46%, 95.00%, and 92.19% respectively, and the emotion detection model attains the values of 98.73%, 98.10%, and 99.55% also using JAFEE, the efficiency of the face detection model is 94.46%, 95.00%, 92.13% also for the emotion detection model attains the values of 98.73%, 98.10%, and 98.55% respectively.

References

[1] Avanthi, M. and Chandra Sekhar Reddy, P., "Human Facial Expression Recognition Using Fusion of DRLDP and DCT Features", Smart Computing Techniques and Applications, pp. 193-199, 2021.

[2] Sikkandar, H. and Thiyagarajan, R., "Deep learning based facial expression recognition using improved Cat Swarm Optimization", Journal of Ambient Intelligence and Humanized Computing, vol. 12,no. 2, pp.3037-3053, 2021.

[3] Song, I., Kim, H. J. and Jeon, P. B., "Deep learning for real-time robust facial expression recognition on a smart phone", In proceedings of IEEE International Conference on Consumer Electronics (ICCE), pp. 564-567, 2014.

[4] Masi, I., Wu, Y., Hassner, T. and Natarajan, P., "Deep face recognition : A survey", In proceedings of 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), pp. 471-478, 2018.

[5] Kim, J.C., Kim, M.H., Suh, H.E., Naseem, M.T. and Lee, C.S., "Hybrid Approach for Facial Expression Recognition Using Convolutional Neural Networks and SVM", Applied Sciences,vol.12,no.11, pp.5493, 2022.

[6] Liu, T., Wang, J., Yang, B. and Wang, X., "Facial expression recognition method with multi-label distribution learning for non-verbal behavior understanding in the classroom", Infrared Physics&Technology,vol.112, p.103594, 2021.

[7] Deng, W., Hu, J. and Guo, J., "Compressive binary patterns: Designing a robust binary face descriptor with random-field eigen filters", IEEE transactions on pattern analysis and machine intelligence, vol.41, no.3, pp.758-767, 2018.

[8] Liu, J., Feng, Y. and Wang, H., "Facial expression recognition using pose-guided face alignment and discriminative

features based on deep learning", IEEE Access,vol.9,pp.69267-69277,2021.
[9] Goodfellow, I., Bengio, Y. and Courville, A., "Deeplearning",MITpress,2016.
[10] Parkhi, O. M., Vedaldi, A.and Zisserman,A.,"Deepfacerecognition",2015.
[11] Mollahosseini, A., Chan, D. and Mahoor, M.H., "Going deeper in facial expression recognition using deep neural networks", In proceedings of IEEE Winter conference on applications of computer vision (WACV),pp. 1-10, 2016.
[12] Zhao, K., Chu, W.S. and Zhang, H., "Deep region and multi-label learning for facial action unitdetection", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.3391-3399, 2016.
[13] Sharif Razavian, A., Azizpour, H., Sullivan, J. and Carlsson, S., "CNN features off-the-shelf : an astounding baseline for recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 806-813, 2014.
[14] Mitra, S.K., "Modular FER: A Modular Facial Expression Recognition from Image Sequence Based on Two Dimensional (2D) Taylor Expansion", SN Computer Science, vol.2, no.3, pp.1-15,2021.
[15] Yao,L.,Wan,Y., Ni, H. and Xu, B., "Action unit classification for facial expression recognition using active learning and SVM", Multimedia Tools and Applications, vol.80, no.16,pp.24287-24301, 2021.
[16] Poli, R., Kennedy, J. and Blackwell, T., "Particle swarm optimization", Swarm intelligence,vol.1,no.1, pp.33-57, 2007.
[17] Dorigo, M., Birattari, M. and Stutzle, T., "Ant colony optimization", IEEE computational intelligence magazine, vol.1, no.4, pp.28-39, 2006.
[18] Sun, X., Zheng, S. and Fu, H., "ROI-attention vectorized CNN model for static facial expression recognition", IEEE Access,vol.8,pp.7183-7194,2020.
[19] FER-2013,https://www.kaggle.com/datasets/msambare/fer2013.
[20] CK+,https://paperswithcode.com/dataset/ck.
[21] Boughida, Adil, Mohamed Nadjib Kouahla, and Yacine Lafifi, “A novel approach for facial expression recognition based on Gabor filters and genetic algorithm,” Evolving Systems, vol. 13, no. 2, pp. 331-345, 2022.
[22] Jeong, Jae-Yeop, Yeong-Gi Hong, Daun Kim, Yuchul Jung, and Jin-Woo Jeong, “Facial expression recognition based on multi-head cross attention network,” arXiv preprint arXiv : 2203.13235, 2022.
[23] Umer, Saiyed, Ranjeet Kumar Rout, Chiara Pero, and Michele Nappi, “Facial expression recognition with trade-offs between data augmentation and deep learning features,” Journal of Ambient Intelligence and Humanized Computing, pp. 1-15, 2022.
[24] Kumari, Naveen, and Rekha Bhatia, “Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter,” Soft Computing, vol. 26, no. 16, pp. 7817-7830, 2022.
[25] Siddiqui, Nyle, Thomas Reither, Rushit Dave, Dylan Black, Tyler Bauer, and Mitchell Hanson, “A Robust Framework for Deep Learning Approaches to Facial Emotion Recognition and Evaluation,” In 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML), IEEE, pp. 68-73 2022.