FACIAL EMOTION DETECTION USING DEEP LEARNING

Ass Prof M. Venkata Rao[1], S. Harshini[2], G. Divya[3], T. Usha Gayathri Raimaji[4]
[1] Professor, [2, 3 & 4] Student
[1]venkatamarella670@gmail.com, [2] sureharshini2003@gmail.com, [3] garapatidivya22@gmail.com, [4] ushagayathri369@gmail.com
Department of Computer Science and Engineering,
Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India

Abstract— **In this world, one can infer a lot about someone's emotional state from their facial expressions. One thing is certain, though: people use all of their facial expressions to convey their emotions. AI technology enables the identification of emotions despite the distinct facial features and varied emotional expressions exhibited by individuals. Deep learning serves as a method for constructing convolutional neural networks (CNNs) capable of recognizing facial emotions, achieving notable prediction accuracy through thorough training of the dataset. Among the many expressions seen, the main goal is to recognize seven basic facial emotions. The main emotions that are commonly displayed on a human face include fear, disgust, surprise, rage, sadness, and neutrality. Various methods exist for achieving facial expression recognition, including deep face algorithms, machine learning approaches, artificial intelligence techniques, Tensorflow, and other methodologies. The suggested recognition model yields a highest accuracy of 73.9% on the enhanced FER2013 dataset, exhibiting an impressive rise above the state-of-the-art. This one appears to be the most direct and practical of all of them. Facial expression detection is becoming a major application in computer vision, with various useful uses in psychological computing, interaction between humans and computers, and mental well-being monitoring. Our approach is centered on creating a reliable and effective model that can precisely identify and categorize facial expressions.**

Keywords—Human Image Extraction, the Recognition of Emotions, Computer Vision, Real-time Detection, Facial Image Preprocessing, Unsupervised Learning, Human-Computer Interaction.

I. INTRODUCTION

Since moods are a neuro-physiological phenomenon, important information may be missed when evaluating state of mind without first monitoring alterations in the body [1]. Nevertheless, speech, face, and gesture recognition constituted the focus of several present investigations on using emotion as a component of HCI. In human contact, facial expressions play a crucial role in assisting us in understanding one another's intentions [2-3]. Typically, individuals rely on the tones of voice and facial expressions to deduce the emotional states of others, including happiness, irritation, or grief. According to several surveys, verbal parts express only Verbal communication constitutes one-third of human interaction, while nonverbal elements convey the remaining two-thirds [4]. Facial expressions serve as a primary channel for interpersonal communication transmission among many nonverbal cues because they conveyed delicate meaning [5-6]. Therefore, it makes sense that research on facial expression has drawn more curiosity nowadays as an outcome of its usage in emotional computing, computer animation, and the cognitive and perceptual sciences. Facial expression and facial emotion recognition are two related topics with growing interest in automated FER, which is referred to by different acronyms in each study [7]. The common components of face emotion detection were covered in this essay, which expanded on how quickly artificial intelligence techniques are developing.

The most crucial area of the man-machine interaction is the FER-based emotion detection domain. Various obstacles in the emotion-detecting domain include changes in posture, facial ornaments, uneven lighting, and more. The use of a traditional system for emotion recognition has the drawback of having a mutual optimizer for extracted features and classification. Researchers are giving DL techniques more attention to address that issue. Currently, the DL system is capable of leading the charge in classification jobs.

Fig. 1. Sample Images of seven facial emotions

In the provided Figure (Fig.1), sample images depicting seven facial emotions are displayed. This lists the emotions surprise, fear, disgust, contempt, angry, happy and sad. Three main processes comprise standard FER techniques: facial and face component detection, feature extraction, and expression classification [8]. First, a face picture may be recognized from an input image, and then from face regions, landmarks, or facial characteristics (such as the nose and eyes) can be recognized. Second, several temporal and spatial features were obtained from aspects of the face [9-10]. Thirdly, pre-trained facial expression (FE) methods producing the detection results use the extracted features to train AdaBoost, and SVM. The separation of feature extraction and classification stages in conventional systems is a drawback. Therefore, improving the system's performance is difficult [11]. In deep learning, size of the dataset is very important, and larger datasets yield better performance [12]. Previous research has shown that the Convolutional Neural Network (CNN) is an effective technique for tasks involving segmentation and classification [13–14]. Automated feature extraction stands out as a key advantage of employing Convolutional Neural Networks (CNNs).

One of CNN's primary advantages is the automated feature extraction. Enhancements to the residual network: incorporating multi-scale convolution expands the perceptual field of the network model by utilizing convolution kernels of diverse sizes, enhancing its capacity to capture intricate features of expression changes. Additionally, integrating a maximum pooling layer optimizes the downsampling process, minimizing information loss. The intra-class distance is essentially reduced while the inter-class distance is simultaneously increased by using Arcface Loss in place of the conventional Softmax Loss. This adjustment aims to reduce the model's misclassification rate for comparable negative phrases. A deep convolutional neural network constructed with a focus on regions of interest specifically tailored for recognizing facial expressions was proposed by Mao et al. [15]. The suggested multi-region coordinate attention facial expression recognition based on the residual learning model demonstrates superior performance compared to numerous advanced conventional algorithms and deep learning models, yielding the model achieves superior outcomes on the FER2013 dataset.

II. RELEVANT WORK

Recent studies have focused on the recognition of facial expressions. A deep learning model intended for FER is presented in reference [16]. CNN is the model's foundation. Seven distinct emotions are classified by this system based on a person's facial image: sad, fearful, happy, angry, neutral, disgusted, and surprised. The FER2013 dataset [17], which contains 35,685 grayscale pictures, was utilised to train and test the model. Of the intended dataset, 70% was set aside for training and 30% for testing. The accuracy of this model is 73.9%. By using parameter optimisation strategies, Vulpe-Grigorasi & Grigore built yet another facial expression recognition system based on CNN [18].

A CNN-based attention FER model was presented by Minaee et al. [19]. The model concentrates attention on certain facial features, such as the lips and eyes, that are thought to have a greater influence on the classification. By employing the spatial transformer, the parameters are recovered together with the characteristics that are sent to the dense layer by the CNN layer [20]. Using the FER2013 dataset, the model was trained on 28,709, and 7178 pictures, respectively, for training and testing. For people with visual impairments, a DL facial emotion detection system is created [21]. With the CNN-based architecture, the face picture is categorised into seven emotional groups. The model demonstrated an accuracy of 73.9% on the FER2013 dataset after 100 training epochs. The projected class label is displayed to users by an Android application that incorporates the model, takes pictures, and classifies them.

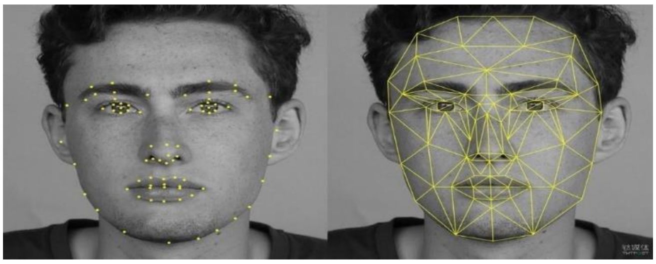


Fig. 2. Facial features with key points.

Here the above Figure (Fig.2) shows the facial features with key points. It likely refers to specific locations on a human face that are important for facial recognition in computer vision. Even while traditional face recognition algorithms that rely on hand-crafted features have made significant progress, over the past ten years, researchers have been more and more drawn to deep learning approaches because of their remarkable automatic detection capabilities. We will talk about some of the most recent FER research in this field, demonstrating deep learning strategies that have been proposed to enhance detection. On multiple sequential or static databases, conduct training and testing. Mollahosseini recommend deep neural networks for emotion detection across many data warehouses [22]. Subsequently, the augmentation data method was applied. In order to address overfitting issues and improve local performance, they utilise locally applied convolution layers and include the network-in-network technique as a choice.

Pre-processing data before network training can improve emotion classification, according to a study by Lopes et al. [23]. Before the CNN was put into practice, a number of preprocessing processes were carried out. Intensity normalisation, cropping, rotation correction, downsampling to 48x48 pixels, and data augmentation were some of these. During testing, the ideal weights that were obtained in training are used. Mohammadpour similarly employed these pre-processing techniques [24]. The identification of the face's essential components was suggested by Yolcu et al. [25]. They utilised three identically designed CNNs are used, each of which can recognise different face features like the mouth, eye, or eyebrow. The process involves cropping and keypoint facial detection stages prior to inputting the images into the CNNs.

For facial emotion detection, a secondary CNN was trained using the iconic face obtained alongside the raw image. Studies demonstrate that this approach provides more accuracy than using just iconized faces or raw photos. Using the FER2013 database, Agrawal et Mittal [26] conducted a study in 2019 to examine the impact of CNN parameter modification on recognition rate.

Every image is first defined as 48 x 48 pixels. These CNNs differ in the size and quantity of additional filters that are used. The optimizer (adam, SGD, or adadelta) that is selected as well as the use of a softmax function for classification are important considerations. All the scholars that were previously mentioned classify the basic emotions into groups: neutrality, fear, wrath, disgust, surprise, sadness, and happiness (Fig. 3).

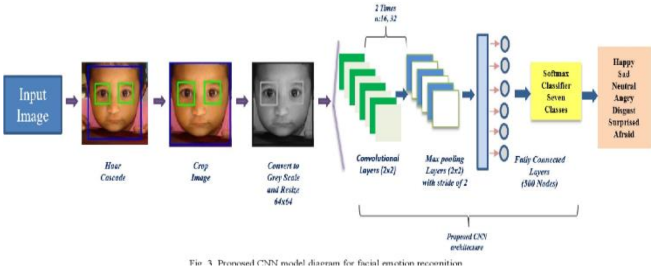


Fig. 3. A Deep Convolutional Neural Network Architecture model for recognizing human emotions.

Here, the human emotion model of the deep convolution neural network architecture is displayed in Figure 3 above. A softmax classifier is used to produce the output after the input image is transformed through a series of layers.

III. CNN Architecture

Owing to its great network depth, the CNN approach belongs to the Deep Learning algorithm family. Applied to visual data, it performs far better [27]. Availability of DL will mitigate the problem of backpropagation gradient loss, hence reducing the training time [28]. Hubel and Wiesel conducted visual brain investigations of the cat's [29] visual perception, which served as the basis for the early study that led to the CNN discovery. When it comes to current visual processing systems, the animal visual cortex is extremely potent. CNN functions in two dimensions, although it functions similarly to MLP. A single dimension is represented by each neuron in MLP, in comparison. Starting with a single feature, like changes in brightness or edges contribute to increased complexity in features, which aid in the individual characterization of objects based on layer thickness and image processing.

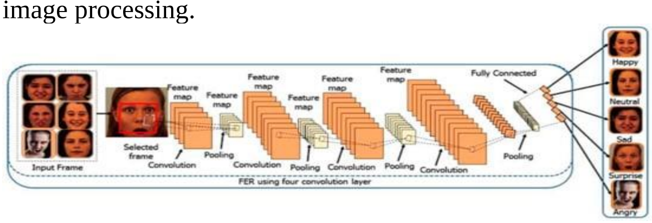


Fig. 4. CNN Architecture.

Feature extraction layers (also known as extraction layers) and classification layers (also known as classification layers) make up the majority of the layer types found in convolutional neural networks. The measurements of a layer are its breadth, height, and depth; the quantity of layers is its depth. Tens to hundreds of millions of layers make up a CNN, and each layer learns to identify different types of images. Image processing is applied, as shown in the picture. Each training image is exposed to different resolutions, and each image's output is treated before being used as an input for the layer that comes after it.

IV. DATASETS

The FER2013 dataset contains 35,887 48x48 pixel-resolution images of people's facial expressions that were generated via a Google image search. The FER2013 picture dataset is multifaceted, containing non-facial and textual images next to faces representing seven different emotions: happy, anger, sorrow, fear, surprise, disgust, and neutrality. Furthermore, several photos [30] have missing labels and noise input (sleepy faces).

Fig. 5. FER2013 Dataset

Here the Figure (Fig.5) shows the FER2013 dataset used in this project. 48x48 pixel grayscale portraits of faces make up this dataset. Here, the face is roughly in the middle and takes up the same amount of space in each picture.

Table I. FER Dataset Split-up Table

Num of classes	7
Num of training images	28821
Num of validation images	7066
Total Num of images	35887

Here the above Table 1 depicts the table for splitting up the FER dataset. The table has 7 facial expression classes and a total of 35887 photos, of which 28821 are used for training and 7066 are used for validation.

V.PROPOSED WORK

Utilizing neural networks based on convolution (CNNs), a type of deep learning, we want to build a reliable face expression detection model for our suggested system. The overarching objective is to discern the emotions of individuals, including anger, sadness, disgust, happiness, fear, surprise, and neutrality. Our approach involves a hybrid feature extraction method coupled with the utilization of CNNs for frame-based expression recognition within large-scale images. Our model is proposed based on the following criteria:

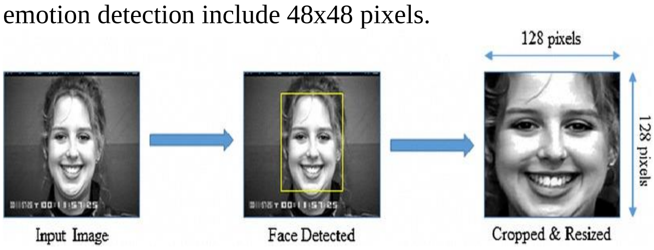
A. Preprocessing Techniques.
B. Split the data.
C. Build the model.
D. Accuracy.

A. PREPROCESSING TECHNIQUES:

Preprocessing in facial emotion detection using deep learning is a critical step that involves transforming raw facial images into a format suitable for training and testing deep neural networks. The model performs better overall and gains more capability to learn pertinent characteristics thanks to the preprocessing stage.

1) Resizing of Images:

Resizing images to a standardized size is a fundamental preprocessing step. This step ensures uniform dimensions for all input images. facilitating consistency during model training. Commonly used dimensions for facial emotion detection include 48x48 pixels.



2) Gray Scaling:

Converting color images to grayscale simplifies the input data, reducing computational complexity. Grayscale images contain intensity values ranging from black to white, capturing essential facial features without the computational overhead associated with RGB color channels.

3) Normalization:

Normalizing pixel values is crucial for ensuring this ensures that the input data maintains a consistent scale. Common normalization techniques are employed for this purpose involve scaling pixel values to fall within a specific range, such as [0, 1] or [-1, 1]. This helps prevent issues related to disparate pixel value ranges and accelerates convergence during training.

4) Face Alignment:

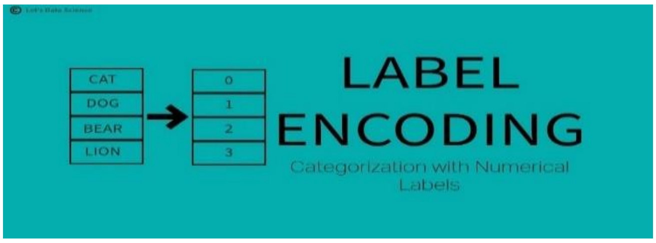
Face alignment corrects variations in head pose, ensuring that the face is consistently positioned within the image. This step is vital for capturing accurate facial features, as it reduces variability caused by different head orientations.

5) Data Enhancement:

Randomly transforming training images, such as rotating, flipping, and shifting them, is known as data augmentation. By doing so, the training dataset is effectively increased, reducing the likelihood of overfitting.

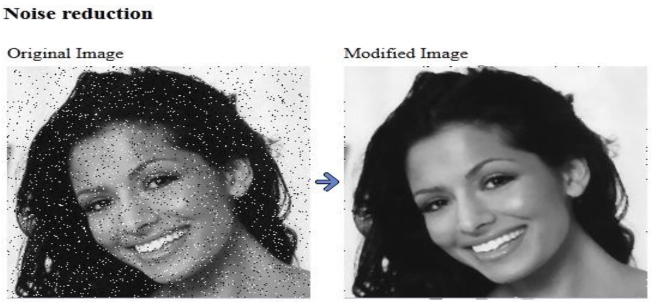
6) Label Encoding:

To train deep learning models, emotion labels must be encoded into a numerical format. Common encoding methods include one-hot encoding, where each emotion class is represented as a binary vector.



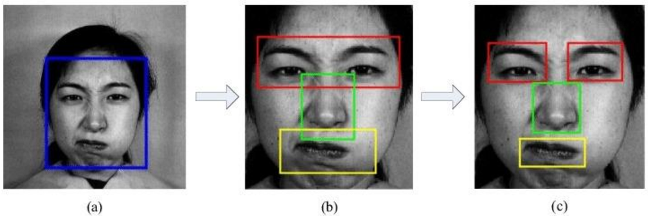
7) Noise Removal:

Removing noise from images helps eliminate irrelevant information that might hinder accurate emotion recognition.



8) Extraction of the Field of Interest:

The most pertinent elements are highlighted in the model by locating and extracting the region of interest, which is usually the facial area. Techniques like face detection algorithms or landmark detection can be employed for accurate ROI extraction. By implementing these preprocessing techniques, the input data is refined and standardized, allowing the deep learning model to effectively learn and recognize facial expressions.



B. SPLIT THE DATA:

This procedure entails partitioning the dataset into separate subsets, commonly comprising training, validation, and testing sets. Proper data splitting is essential for training the model, tuning its hyperparameters, and evaluating its performance on unseen data.

1) Training Set:

In this stage, the model undergoes training to identify patterns and characteristics linked to various face expressions. In order for the model to generalise successfully to new, unseen data, it needs a well-representative training set.

2) Validation Set:

As the model trains on the training set, it periodically evaluates its performance on the validation set. Adjustments to the model, such as learning rates or regularization parameters, can be made based on the validation performance.

C. BUILD MODEL:

A CNN is used to build the model. Python programming can be used to implement the model. additionally simulated in a Jupyter notebook, combining the layers of convolutional neural networks to build the model. Model fitting and compilation are handled using Keras, a deep learning toolkit that runs on top of TensorFlow. These confusion matrices, along with other graphs like accuracy and loss graphs, are plotted using Matplotlib and Seaborn. The activation functions are Relu and Softmax.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 46, 46, 128)       1280

 max_pooling2d (MaxPooling2  (None, 23, 23, 128)       0
 D)

 dropout (Dropout)           (None, 23, 23, 128)       0

 conv2d_1 (Conv2D)           (None, 21, 21, 256)       295168

 max_pooling2d_1 (MaxPoolin  (None, 10, 10, 256)       0
 g2D)

 dropout_1 (Dropout)         (None, 10, 10, 256)       0

 conv2d_2 (Conv2D)           (None, 8, 8, 512)         1180160

 max_pooling2d_2 (MaxPoolin  (None, 4, 4, 512)         0
 g2D)

 dropout_2 (Dropout)         (None, 4, 4, 512)         0

 conv2d_3 (Conv2D)           (None, 2, 2, 512)         2359808

 max_pooling2d_3 (MaxPoolin  (None, 1, 1, 512)         0
 g2D)

 dropout_3 (Dropout)         (None, 1, 1, 512)         0

 flatten (Flatten)           (None, 512)               0

 dense (Dense)               (None, 512)               262656

 dropout_4 (Dropout)         (None, 512)               0

 dense_1 (Dense)             (None, 256)               131328

 dropout_5 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 7)                 1799

=================================================================
Total params: 4232199 (16.14 MB)
Trainable params: 4232199 (16.14 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Fig.6. Depicts the analysis of proposed FER model built using keras.

Here the above Figure (Fig.6) shows an examination of the suggested FER model constructed with Keras. Here we can see the total params of 4232199 in which some of them are trainable params 4232199 and non-trainable params are 0.

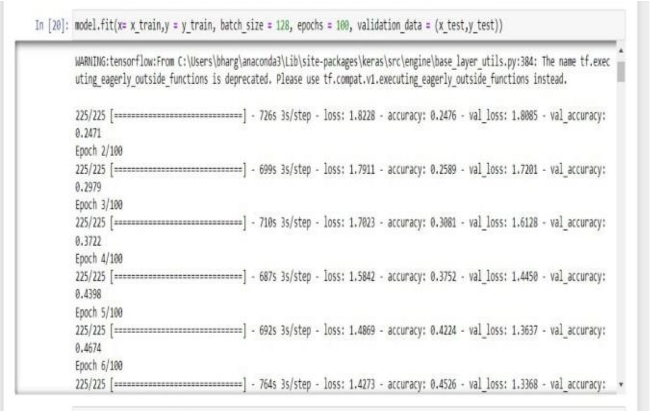


Fig.7. Loss and accuracy values of proposed system.

In the illustrated graph (Fig.7), the depicted model exhibits a notable trend: as time progresses, there is a discernible decrease in loss while simultaneously witnessing an increase in accuracy across both training and validation phases.

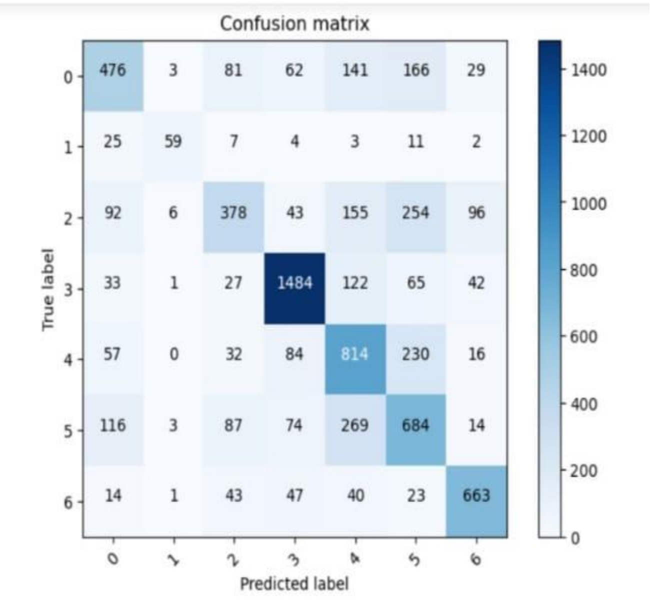


Fig.8. Depicts Confusion Matrix.

Here the above Figure(Fig.8) shows the confusion matrix.

D. ACCURACY:

Accuracy serves as a prevalent metric for assessing the effectiveness of a deep learning algorithm, quantifying the percentage of accurately classified instances relative to the total instances in the test dataset. Here we achieved an accuracy of 73.9%.

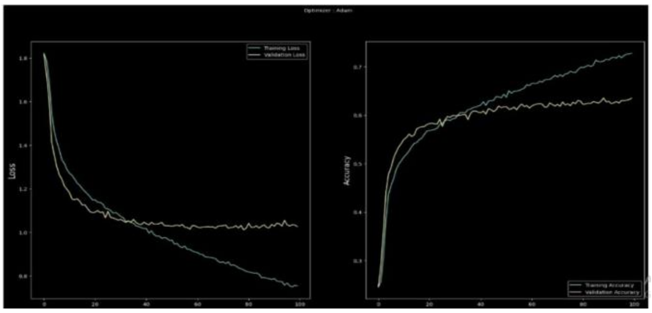


Fig.9. illustrates the loss and accuracy of training and validation (testing) data.

In this instance, the model training process is shown in the above Figure (Fig.9) as loss, accuracy, val loss, and val_accuracy. By using the Adam Optimizer the training loss is recorded as 0.72%, while the validation loss is obtained as of 1.03% for 100 epochs. Similarly, we achieved the Training accuracy is of 73.9% and a validation accuracy is of 0.62% for 100 epoches.

Table II. The above Table 2 represents the accuracy and loss percentage of Existing and Proposed System.

FER2013 Dataset	Accuracy
Existing System [31]	73.4
Proposed System	73.9

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, deep learning-based facial emotion recognition has made tremendous strides, completely changing how computers perceive and decipher human emotions from facial expressions. We have examined the theoretical underpinnings, approaches, difficulties, and uses of deep learning-based facial emotion Identification in this study. Deep learning techniques can be used to automatically extract discriminative characteristics from raw face pictures using Convolutional Neural Networks (CNNs), in particular, as useful tools. CNNs have shown impressive performance in recognizing and understanding a wide the spectrum of emotional states encompasses fundamental emotions such as happiness and sadness, extending to more intricate expressions like scorn and humor., by utilizing large-scale datasets and hierarchical learning. Even so, there are still several obstacles to overcome and room for development. To properly train deep learning models, one of the main obstacles is the requirement for big annotated datasets containing a variety of demographic and facial expression representations.

Furthermore, improving the generalization and dependability of face emotion recognition systems requires tackling problems like data imbalance, domain shift, and model interpretability. Furthermore, investigating few-shot learning, domain adaptation, and transfer learning strategies can help with emotion recognition in a variety of dynamic real-world settings.

Furthermore, the potential to improve the temporal modeling and contextual comprehension of facial expressions is presented by advances in deep learning architectures, including transformer models, recurrent neural networks, and attention processes. These advancements may pave the way for more sophisticated and context-aware emotion identification algorithms that can pick up on minute differences in emotional states.

Moreover, the use of facial expression recognition is expanding into new fields like affective computing, virtual reality, and social robots in addition to more established ones like healthcare and human-computer interaction. Researchers can provide more opportunities for emotion-aware technologies that improve user experiences, lead to better mental health outcomes, and promote more sympathetic human-machine interactions by utilizing deep learning techniques.

The field is well-positioned to maintain its current development and innovation trajectory by tackling current issues and venturing into uncharted research and application territories. This will pave the way for a future in which machines will be able to comprehend and react to human emotions with an unparalleled level of precision and empathy.

VII. REFERENCES:

[1] Ergen and Burhan, "Facial Landmark Based Region of Interest Localization for Deep Facial Expression Recog-nition", Technical Gazette, vol. 29, no. 1, pp. 38–44, 2022

[2] Arora, Malika and Munish Kumar, "AutoFER: PCA and PSO based automatic facial emotion recog-nition", Multimedia Tools and Applications, vol. 80, no. 2, pp. 3039-3049, Jan. 2021.

[3] Chhikara, Prateek, P. Singh, R. Tekchandani, N. Kumar and M. Guizani, "Federated learning meets human emotions: A decentralized framework for human–computer interaction for IoT applications", IEEE Internet of Things J, vol. 8, no. 8, pp. 6949-6962, 2021.

[4] T. Vo, G. Lee, and H. Yang, "Pyramid with super resolution for in-the-wild facial expression recogn-ition", IEEE Access, vol. 8, pp. 131988-132001, 2020.

[5] Houssein, Essam H., Asmaa Hammad and Abdelmgeid A. Ali, "Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review", Neural Computing and Applications, vol. 34, pp. 1-31, May 2022.

[6] Mellouk, Wafa and Wahida Handouzi, "Facial emotion recognition using deep learning: review and insights", Procedia Computer Science 175 pp. 689-694, 2020.

[7] H. Liu, T. Liu, Z. Zhang, A. K. Sangaiah, B. Yang and Y. Li, "Arhpe: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction", IEEE Transactions on Industrial Informatics, vol.18, no.10 pp. 7107-7117, 2022.

[8] Farzin and Saeed, "Prediction of groundwater table and drought analysis; a new hybridization strategy based on bi-directional long short-term model and the Harris hawk optimization algorithm", Journal of Water and Climate Change, vol. 13, no. 5, pp. 2233-2254, May 2022.

[9] Stoean and Catalin, "Metaheuristic-based hyperpar-ameter tuning for recurrent deep learning: application to the prediction of solar energy generation", Axioms vol. 12, no. 3, pp. 266, Mar. 2023.

[10] Z. He, Z. Li, F. Yang, L. Wang, J. Li, and C. Zhou, "Advances in multimodal emotion recognition based on brain–computer interfaces", Brain sciences, vol. 10, no. 10, pp. 687, Sep. 2020.

[11] Veni, R. Anand, D. Mohan, and E. Paul, "Feature fusion in multimodal emotion recognition system for enhancement of human-machine interaction", IOP conference series: materials science and engineering, vol. 1084. no. 1. IOP Publishing, 2021.

[12] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recog-nition", Alexandria Engineering Journal, vol. 61, no. 6, pp. 4435–4444, June 2022.

[13] A. M. Mutawa, M. Murugappan and A. Hassouneh, "Designing a framework to detect emotions in real-time by combining facial expressions and EEG data, utilizing machine learning algorithms and deep neural networks", Informat. Med. Unlocked, vol. 20, 2020.

[14] K. Mohan, A. Seal, O. Krejcar and A. Yazidi, "FER-net: facial expression recognition using deep neural net", Neural Computing and Applications, vol.33, no. 15 , pp. 9125-9136, 2021.

[15] Mao, Rui, Runxin Meng and Ruijing Sun, "Facial expression recognition based on deep convolutional neural network", Third International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI 2022), vol. 12509. SPIE, 2023.

[16] Lalitha and Sneha K, "A deep learning model for face expression detection", 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). IEEE, pp. 647-650, Aug. 2021.

[17] Vulpe-Grigorași, Adrian and Ovidiu Grigore. "Convolutional neural network hyperparameters optimization for facial emotion recognition", 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE). IEEE, pp. 1-5, Mar. 2021.

[18] Minaee, Shervin, Mehdi Minaei and Amirali Abdolrashidi, "Deep-emotion: Facial expression recog-nition using attentional convolutional network", Sensors, vol. 21, no. 9, pp. 3046, April 2021.

[19] J. L. Joseph and S. P. Mathew, "Deep learning-based facial expression recognition for visually impaired people", in Proceedings of the IEEE 4th International Conference on Computer-Assisted Speech, Power and Communication Technology (GUCON), pp. 1–5, September 2021.

[20] Mollahosseini, Ali, David Chan, and Mohammad H. Mahoor, "Going deeper in facial expression recognition using deep neural networks", 2016 IEEE Winter conference on applications of computer vision (WACV), pp. 1–10, March 2016.
doi: 10.1109/WACV.2016.7477450.

[21Lopes and Andre Teixeira, "Facial expression recognition with convolutional neural networks: coping with few data and

the training sample order", Pattern recognition, vol. 61, p. 610–628, January 2017.
doi: 10.1016/j.patcog.2016.07.026 .

[22] Mohammadpou and Mostafa, "Facial emotion recognition using deep convolutional networks", *2017* IEEE 4th international conference on knowledge-based engineering and innovation (KBEI). IEEE, 2017. doi: 10.1109/KBEI.2017.8324974.

[23] Yolcu and Gozde, "Facial expression recognition for monitoring neurological disorders based on convolutional neural network", Multimedia Tools and Applications, vol. 78, no. 22, pp. 31581–31603, Nov. 2019. doi: 10.1007/s11042-019-07959-6.

[24] Agrawal, Abhinav and Namita Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy", The Visual Computer, vol.36, no. 2, pp. 405-412, 2020. doi: 10.1007/s00371-019-01630-9.

[25] Mazen, Fatma Mazen Ali, Ahmed Aly Nashat, and Rania Ahmed Abdel Azeem Abul Seoud. "Real time face expression recognition along with balanced FER2013 dataset using CycleGAN." International Journal of Advanced Computer Science and Applications vol.12, no.6, (2021).

[26] G. C. Porusniuc, F. Leon, R. Timofte, and C. Miron, "Convolutional neural networks architectures for facial expression recognition", Proc. E-Health Bioeng. Conf. (EHB), pp. 1-6, Nov. 2019.

[27] Arriaga, Octavio, Matias Valdenegro-Toro, and Paul Ploger, "Real-time convolutional neural networks for emotion and gender classification", arXiv preprint arXiv: pp. 1710.07557, 2017.

[28] A. Ahmad & Yayasan Cahaya Islam, "Mengenal Artificial Intelligence, Machine Learning, Neural Network, and Deep Learning", Journal of Indonesian Technology, 2017.

[29] Simonyan, Karen, and Andrew Zisserman, "Two-stream convolutional networks for action recognition in videos", Advances in neural information processing systems, vol. 27, 2014.

[30] Dataset Link
https://www.kaggle.com/datasets/msambare/fer2013

[31] Gupta, Swadha, Parteek Kumar, and Raj Kumar Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models", Multimedia Tools and Applications, vol.82, no. 8, pp. 11365-11394, 2023.