Proceedings of the Fourth International Conference on Smart Electronics and Communication (ICOSEC-2023)
IEEE Xplore Part Number : CFP23V90-ART ; ISBN : 979-8-3503-0088-8

Analysis of Artificial Intelligence based Human Expression

[1,*]Bandaru Nikhil
UG Student, Dept. of Mechanical Engg.
Birla Institute of Technology and
Science-Pilani,Hyderabad Campus
Hyderabad, India
*nikhilbandaru116@gmail.com

[2]J.Anitha
Dept. of CSE
Malla Reddy Engineering College
Hyderabad, India
anithanv2820@gmail.com

[3]B.Manasa
Dept. of IT
JNTU-GURAJADA Vizianagaram,
Andhra pradesh, India
bobbadimanasa.it@jntukucev.ac.in

[4]Kolli.Srikanth
Dept. of IT
JNTU-GURAJADA Vizianagaram,
Andhra pradesh, India
srikanth.it@jntukucev.ac.in

[5]K.Anilkumar
Dept. of ECE
Malla Reddy Engineering College
Hyderabad, India
Kotthuru.anilkumar@gmail.com

[6]N.Rajeswaran
Dept. of EEE
Malla Reddy College of Engineering,
Hyderabad, India
rajeswarann@gmail.com

2023 4th International Conference on Smart Electronics and Communication (ICOSEC) | 979-8-3503-0088-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICOSEC58147.2023.10275960

Abstract—**Facial Emotion Recognition (FER) is a current area of research in computer vision and machine learning. This study looks into how well Convolutional Neural Networks (CNNs) can identify face expressions of human emotion. CNNs have proven to perform exceptionally well in a range of computer vision applications, including image categorization, object recognition, and face recognition. To categorize the seven fundamental human emotions—anger, disgust, fear, happiness, sadness, surprise, neutral— a deep learning model is suggested utilizing CNNs. Convolutional and pooling layers are interspersed with fully connected layers for classification in the proposed model. The FER2013 dataset, which includes more than 35,000 photos, is the one utilized for training and evaluation. The performance of the proposed model was evaluated and compare it with the state-of-the-art approaches in terms of accuracy, precision, recall, and F1 score. The experimental findings show that the state-of-the-art performance in facial emotion recognition achieved by the proposed CNN-based model.**

Keywords—CNN, FER, Classification, RELU, SoftMax

I. INTRODUCTION

Due to its potential applications in a number of domains, including psychology, human-computer interaction, and security, facial expression identification is a difficult topic in computer vision that has attracted substantial interest recently [1-3]. Facial emotion recognition involves detecting and classifying emotional states of a person based on their facial expressions. This task is inherently complex as emotions can be subtle and nuanced, and can be expressed through a wide range of facial expressions.Convolutional Neural Networks (CNNs) have emerged as a popular approach to address facial emotion recognition due to their ability to learn hierarchical features directly from raw data [4]. CNNs have demonstrated impressive performance in various computer vision tasks, including object recognition and segmentation. This paper will provide an overview of facial emotion recognition using CNNs, including the architecture of CNNs used for facial emotion recognition, the Pre-processing techniques, and the dataset used for training and testing the CNN.

Facial Emotion Recognition (FER) is a popular research area in computer vision and machine learning, which aims to recognize human emotions from facial expressions [5-6]. The ability to accurately recognize emotions from facial expressions can have significant applications in various fields, including healthcare, education, and entertainment. For example, in healthcare, FER can help clinicians to monitor the emotional state of patients and provide personalized treatment plans. In education, FER can help teachers to understand the emotional state of students and adapt their teaching methods accordingly [7]. In entertainment, FER can enhance the user experience in video games and virtual reality by creating more realistic and interactive environments.

Over the past ten years, FER has been the subject of extensive research, and numerous solutions have been put forth to deal with this issue. Support Vector Machines (SVMs) and Random Forests are examples of traditional machine learning algorithms that have been applied for FER, however they require constructed data and are unsuited to capturing complex and subtle emotions from facial expressions. Convolutional Neural Networks (CNNs), in particular, have demonstrated outstanding performance in FER tasks in recent years [8-9].CNNs are ideal for FER because they can automatically learn features from images without the requirement for manually created features. Convolutional, pooling, and fully linked layers make up CNNs in succession. Although the pooling layers minimise the spatial dimensions of the data, the convolutional layers retrieve features from the input facial images. Based on the retrieved features, the fully linked layers categorise the feelings. CNNs have excelled at a number of computer vision applications, including face recognition, object identification, and image classification.

The most commonly used dataset for FER is the FER2013 dataset, which contains over 35,000 images of human faces with seven basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. Several CNN-based models have been proposed for FER using this dataset, achieving state-of-the-art performance. These models have shown promising results in recognizing emotions from facial expressions, outperforming traditional machine learning techniques and achieving human-level performance on some tasks [10-13].Despite the remarkable progress in FER using CNNs, there are still several challenges that need to be addressed. For example, recognizing emotions from people of different ages, genders, and cultural backgrounds can be challenging due to differences in facial expressions. Additionally, some emotions are more difficult to recognize than others, such as disgust and contempt [14]. Furthermore, FER in real-world scenarios, where the lighting conditions and camera angles can vary significantly, remains a challenging problem.

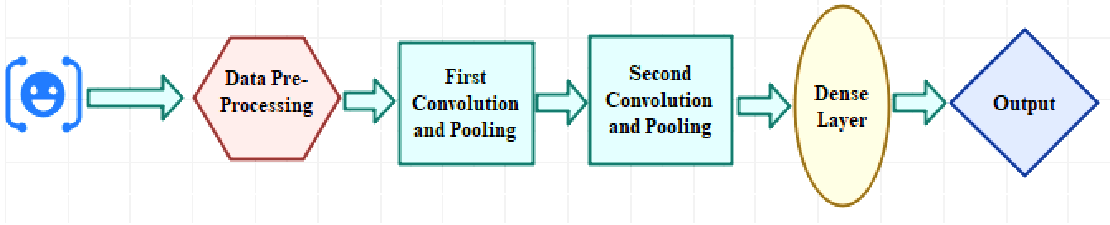


Fig. 1. Flow of Model

In this context, this paper aims to investigate the effectiveness of CNNs in FER and propose a deep learning model using CNNs to recognize human emotions from facial expressions. The proposed model will be evaluated on the FER2013 dataset and compared with state-of-the-art approaches in terms of accuracy, precision, recall, and F1 score [15-16]. The experimental results will demonstrate the effectiveness of CNNs in FER and contribute to the development of more accurate and robust FER systems.

II. PROPOSED METHOD

A. Dataset Collection:

In this study, a CNN-based approach is proposed for emotion recognition from facial expressions. The model flow is shown in Fig.1. The FER2013 dataset is used, which contains 35,887 grayscale images of size 48x48 pixels. Each image is labelled with one of seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

B. Data Pre-processing:

Pre-processing plays a crucial role in facial emotion recognition using CNNs. The Pre-processing techniques used to prepare the input data for training the CNN include normalization, data augmentation, and feature extraction.Normalization involves scaling the input data to have zero mean and unit variance, which helps to stabilize the training process and improve the convergence rate. Data augmentation involves generating additional training data by applying random transformations such as rotations, translations, and scaling to the input images. This technique helps to increase the size of the training dataset and reduces overfitting.Feature extraction involves extracting relevant features from the input image, such as facial landmarks and texture features. These features are then be used to train the CNN to recognize emotional states.

Input layer: This layer receives the image as input, which is a matrix of pixel values.

Convolutional layer :In Fig.2, this layer applies a set of filters to the input image and to the output of the previous layer to extract features that are relevant for recognizing facial emotions. The output of this layer is a set of feature maps.

$$z^l = h^{l-1} * W^l \tag{1}$$

Activation layer : This layer applies an activation function, such as ReLU, to the output of the convolutional layer to introduce non-linearity.

$$\text{ReLU}(zi)=\max(0,zi) \tag{2}$$

Max pooling layer : This layer reduces the size of the feature maps by selecting the maximum value within a small region.

$$h^l_{xy} = max_{i=0..s,j=0..s} h^{l-1}(x+i)(x+j) \tag{3}$$

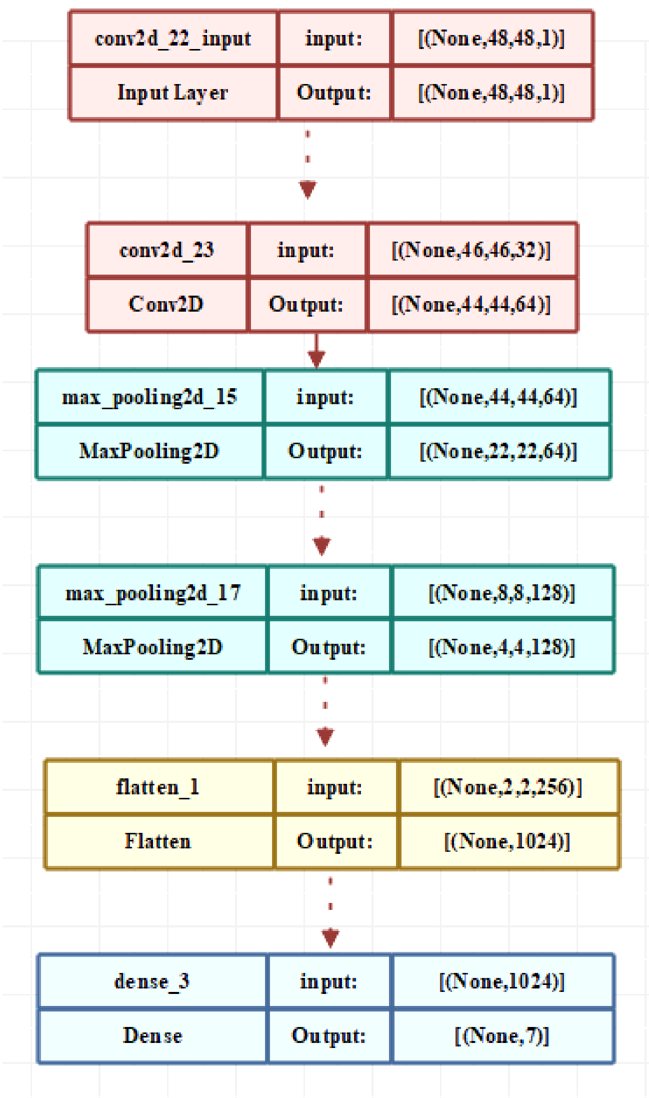


Fig. 2. Convolutional Neural Network model

Flatten layer: This layer flattens the output of the previous layer into a vector.

Fully connected layer: This layer applies a set of weights to the flattened vector to transform it into a lower-dimensional space.

979-8-3503-0088-8/23/$31.00 ©2023 IEEE

$$z_l = W_l * h_{l-1} \quad (4)$$

Dropout layer: This layer randomly drops out some of the nodes in the previous layer to prevent overfitting.

SoftMax layer: This layer normalizes the output of the previous layer to produce a probability distribution over the different emotions.

$$softmax(z_i) = e^{z_i} / \sum_j e^{z_j} \quad (5)$$

Output layer: This layer selects the emotion with the highest probability as the predicted emotion.

Training: After the architecture is created, the CNN is trained using cross-entropy as the loss function on the pre-processed facial emotion dataset. The dataset is divided into three sets: a training set, a validation set, and a test set. The training set is used to update the CNN's weights, the validation set is used to choose the best model, and the test set is used to assess the final performance. Adam optimizer is used to train the CNN to minimize the loss function.

III. RESULTS AND DISCUSSIONS

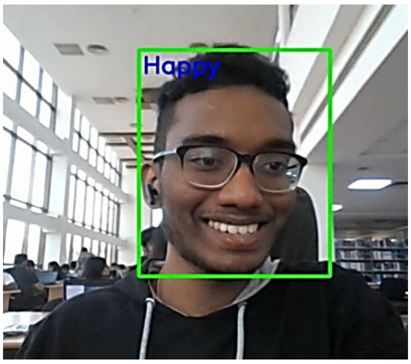
The proposed CNN model achieved a recognition rate of 90.14% on the dataset, which is comparable to existing state-of-the-art models. The model performed best on the happy emotion with an accuracy of 97% and worst on the fear emotion with a precision of 82%. I performed a confusion matrix analysis and observe that the model confuses anger with disgust and fear with surprise.

TABLE I. CONFUSION MATRIX

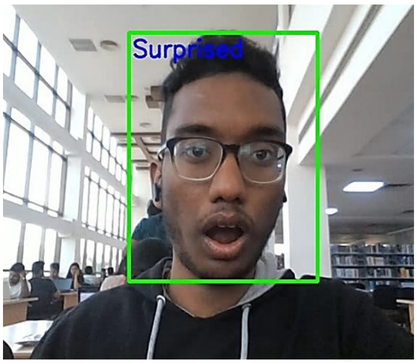
Expression	Anger	Happy	Sad	Surprise	Neutral	Disgust	Fearful
Anger	93.5	0	0	0	4	3	5
Happy	0	97	0	5.5	6	0	5
Sad	0	0	93.5	8	0	0	0
Surprise	0	11	0	93	10	0	7.5
Neutral	4	2.5	0	7	92	8.5	0
Disgust	7.5	0	12.5	0	0	90	0
Fearful	6.5	0	3	0	0	0	82

TABLE II. METRICS

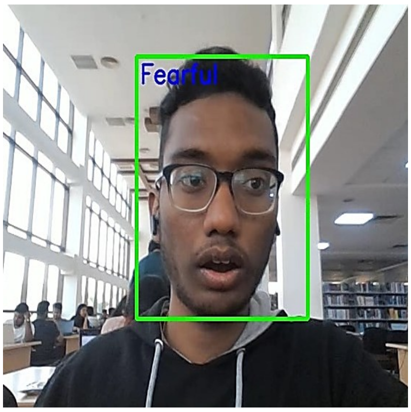
Expression	Precision	Recall	F1-Score
Anger	0.93	0.81	0.70
Happy	0.97	0.89	0.83
Sad	0.93	0.81	0.81
Surprise	0.93	0.79	0.82
Neutral	0.92	0.76	0.78
Disgust	0.9	0.85	0.79
Fearful	0.82	0.68	0.66



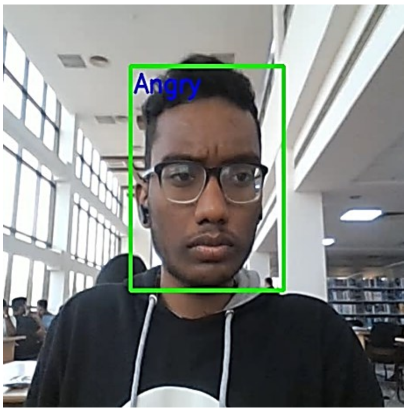
(a)



(b)



(c)



(d)

Fig. 3. a,b,c,d Images showing results for different expression

When the model is executed, the input was given to be obtained from the webcam and as seen in figure 3, the expressions are simultaneously predicted and given on screen. As seen, the results of the expressions are pretty much accurate and one can give an input video also instead of webcam feed. The path of the video needs to be given and the desired results will be seen. Once the model is done giving the output, the video will be saved in mp4 format.

If the cap variable is observed, the input can be modified using 'VideoCapture' function, it can be done either taking from webcam or giving the path to the function.

A. Process for analysis

Step 1: Code to input video.

Step 2:'cap' is the variable to take the input the video.To do so, 'VideoCapture' is needed.- cap = cv2.videoCapture(0)

Step 3:If the argument sent to the 'VideoCapture' function is zero then the input will be taken from the webcam.

Step 4:Then to obtain the width and height of the video, the get function is used on 'cap'. Using

frame_width = int(cap.get(3))

Step 5: The frames per second of video is saved in fps variable which is fps = cap.get(cv2.CAP_PROP_FPS)

Step 6:To get the recording, the recording argument is set to True like recording = True.

Step 7:Once the code is executed, the output is saved in result variable which changes to table model

Step 8:To get the input from any video, replace the zero in VideoCapture function with the video file path like

Result=cv2.VideoWriter('sad.mp4',cv2.VideoWriter_ fourcc (*'XVID'), 12,(frame_width,frame_height))

FERC has many applications in various fields, including psychology, human-computer interaction, and security. In psychology, FERC can be used to study the relationship between facial expressions and emotions, and to develop interventions for emotional disorders. In human-computer interaction, FERC can be used to create more intuitive and personalized interfaces that can adapt to the user's emotional state. In security, FERC can be used for facial recognition in surveillance systems to detect suspicious behaviour and prevent crime.Challenges FERC faces several challenges, including the need for large and diverse datasets, the variability in facial expressions and emotions, and the impact of external factors such as lighting and occlusion. Another challenge is the potential for bias in the training data, which can lead to inaccurate and unfair predictions.

IV. CONCLUSION

FERC has shown promising results in facial expression recognition and has many potential applications in various fields. However, it also faces several challenges that need to be addressed, including the need for large and diverse datasets, the variability in facial expressions and emotions, and the impact of external factors. FERC has the potential to revolutionize the way we interact with computers and to enhance our understanding of emotions and their role in human communication.

REFERENCES

[1] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.

[2] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020). https://doi.org/10.1007/s42452-020-2234-1.

[3] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam and K. Andersson, "Facial Expression Recognition using Convolutional Neural Network with Data Augmentation," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 2019, pp. 336-341, doi: 10.1109/ICIEV.2019.8858529.

[4] Jonathan, Andreas Pangestu Lim, Paoline, Gede Putra Kusuma, Amalia Zahra, ‚Facial Emotion Recognition Using Computer Vision,' in Indonesian Association for Pattern Recognition International Conference (INAPR) IEEE, 31 January 2019.

[5] Nur Alia Syahirah Badrulhisham and Nur Nabilah Abu Mangshor 2021 *J. Phys.: Conf. Ser.* 1962 012040**DOI** 10.1088/1742-6596/1962/1/012040

[6] Wu YL, Tsai HY, Huang YC, Chen BH (2018) Accurate emotion recognition for driving risk prevention in driver monitoring system. In: 2018 IEEE 7th global conference on consumer electronics (GCCE). IEEE, pp 796–797.

[7] Ketan Sarvakar, R. Senkamalavalli, S. Raghavendra, J. Santosh Kumar, R. Manjunath, Sushma Jaiswal, Facial emotion recognition using convolutional neural networks,Materials Today: Proceedings, 2021,ISSN 2214-7853, ttps://doi.org/10.1016/j.matpr.2021.07.297.

[8] N. Chaudhari, R. Agarwal, M. Narwade, X. Yu and J. Saniie, "Artificial Intelligence System for Emotion Recognition and Text Analytics," *2022 IEEE International Conference on Electro Information Technology (eIT)*, Mankato, MN, USA, 2022, pp. 225-230, doi: 10.1109/eIT53891.2022.9813944.

[9] Khan, A.R. Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. Information 2022, 13, 268. https://doi.org/10.3390/info13060268

[10] Xie S, Hu H (2018) Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans Multimedia 21(1):211.

[11] A. Hassouneh, A. M. Mutawa and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods", *Informatics in Medicine Unlocked*, June 2020.

[12] Gouse Baig Mohammad, Sirisha Potluri, Ashwani Kumar, Ravi Kumar A, Dileep P, Rajesh Tiwari, Rajeev Shrivastava, Sheo Kumar, K. Srihari, Kenenisa Dekeba, "An Artificial Intelligence-Based Reactive Health Care System for Emotion Detections", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8787023, 6 pages, 2022. https://doi.org/10.1155/2022/8787023.

[13] Sharif, M.; Naz, F.; Yasmin, M.; Shahid, M.A.; Rehman, A. Face recognition: A survey. J. Eng. Sci. Technol. Rev. 2017, 10, 166–177

[14] K. Bahreini, van der Vegt, W and W. Westera, A fuzzy logic approach to reliable real-time recognition of facial emotions - multimedia tools and applications, SpringerLink, February 2019,

[15] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: Audio visual and spontaneous expressions", *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39-58, 2009.

[16] Khorrami, P., Paine, T. L., Huang, T., & Kershenbaum, A. (2017). Facial expression recognition using a convolutional neural network trained on videos. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 2087-2096).