Date of publication DEC-31, 2024, date of current version APR-30, 2024.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.23.4.3760

Database Development and Recognition of Facial Expression using Deep Learning

MAYURI M. BAPAT, CHANDRASHEKHAR H. PATIL, SHANKAR M. MALI
School of Computer Science, Dr. Vishwanath Karad MIT World Peace University,
Pune, Maharashtra, India 411038.
Email: mmbapat@mitacsc.ac.in, chpatil.mca@gmail.com, shankar.mali@mitwpu.edu.in

Corresponding author: Shankar M. Mali (e-mail: shankar.mali@mitwpu.edu.in).

ABSTRACT Facial expressions serve as a means of conveying human emotions and individual intentions. The ability to perceive and interpret facial emotions is a relatively effortless job for humans, although it poses significant challenges when attempting to replicate this capability using a computer. Facial expressions can be detected from static photos, video, webcam data, or real-time photographs. The primary focus of this study is to create the SMM Facial Expression dataset and to develop a Convolutional Neural Network (CNN) model for accurately recognizing and classifying facial expressions. This model utilizes End-to-end (E2E) networks to analyze individual frames or clusters of frames received from the camera. The analysis is conducted through various layers of convolutional and pooling operations. The proposed model is evaluated on two benchmark datasets, add long form Cohn-Kanade (CK+) and Facial Expression Recognition 2013 Dataset (FER2013) for facial expression recognition (FER). The obtained accuracy rates are 85.27% and 82.18% for CK+ and FER2013, respectively. This study demonstrates that the SMM Facial dataset is comparable in quality to previously benchmarked datasets, and the proposed model holds potential for real-time facial expression recognition.

KEYWORDS Image classification; Convolutional Neural Network; Facial expression recognition; E2E Network.

I. INTRODUCTION

THE human face reflects human emotions. It is nonverbal and the easiest way of communication. Research shows that roughly 55% of human emotions are conveyed by using facial expressions [1]. Seven widespread facial expressions include Happiness, Anger, Fear, Sadness, Surprise, Contempt, and Disgust [2]. Facial expression recognition offers benefits to many phases of life. It is significant and beneficial for the safety and healthcare of human beings. It plays an irreplaceable role in the detection of human feelings at a particular instant without asking questions. This study aims to create a facial expression detection system by classifying input photos into distinct expression classes using a convolutional neural network. Facial expressions get recognized using videos, web cameras, images, and dialogues. Different parts of the face like eyes, eyebrows, forehead, nose, and eyelashes are used in facial expression detection [3, 4].

Recurrent neural networks (RNN) and Convolutional neural networks (CNN) have demonstrated efficacy in automated feature extraction and classification, rendering them valuable tools in contemporary research. Consequently, several scholars have advocated for the adoption of these technologies in the domain of human expression recognition. Various classifiers, such as Support Vector Machine (SVM) [2], K-Nearest Neighbor (KNN) [5], Haar Classifier, Bayesian Network [6], and Sparse Representation Method (SRM) [7], Deep Neural Network [9] are employed for facial expression recognition. The researchers made significant endeavors to design various deep neural network designs that yield satisfactory results inside this domain of research. Figure 1 illustrates the fundamental procedures employed in expression detection.

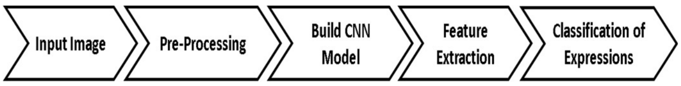

Figure 1. Steps in Expression Detection

Facial expression recognition using CNN accepts image data as input. Input images get pre-processed before training so that the model learns effectively leading to improved performance and better generalization on new, unseen data. Preprocessing helps to focus on necessary and particular areas of the image. CNN comprises different layers like the convolution layers, pooling layer, fully connected layer, and Soft-max layer. For feature learning, convolution, and pooling layers are responsible, whilst fully connected and Soft-max layers are used to classify expressions. The features from the input images are extracted using the convolution layer. Pooling is used to select the required information from given features.

The SMM dataset, a comprehensive collection specifically designed for Facial Expression Recognition, is introduced in this paper. Additionally, a CNN-based deep learning model is developed from scratch and trained on the SMM dataset. The model's performance is then evaluated on two publicly available datasets: FER2013 and CK+. On the FER2013 dataset, an accuracy of approximately 82.18% is achieved by our model. Similarly, an impressive accuracy of about 85.27% is demonstrated on the CK+ dataset, indicating that the proposed model performs remarkably well on these benchmark datasets. Due to its high accuracy and robustness, our model has great potential for real-time facial expression recognition.

The format of the paper is as follows. The background of facial expression recognition is presented in Section I. Section II provides a literature review. Section III shows the SMM facial expression database development. Section IV describes the proposed method. The experiment's findings are presented in Section V. Section VI concludes the work done.

II. LITERATURE REVIEW

We have carried out a study of the existing literature in two sections. In the first section we have identified and studied the research work related to the datasets that are available and used for the experiments related to facial expression recognition. Table 1 represents the detailed analysis for the dataset papers studied. Followed by, in the second section we have gone through the literature that elaborates the machine learning methodologies and different architectures implemented for facial expression recognition.

Various studies were conducted on facial expression recognition using different datasets like Karolinska Directed Emotional Faces (KDEF) datasets and Kaggle's Facial Expression Recognition Challenge [8, 9], FER-2013 [10, 16, 17, 18], Japanese Female Facial Expressions (JAFFE) [11], Cohn Kanade [2, 12] that are frequently used. Some researchers also used datasets like Caltech faces, CMU database and NIST database [13], Taiwanese Facial Expression Image Database(TFEID) [14], KinFaceW-I and II [15], Yale [2], The MUG Facial Expression Database [19], MultiPie [20], MMI [21], GEMEP FERA [22], Static Facial Expressions in the Wild (SFEW) [23], AffectNet [24] and Radboud Faces Database (RAFD-DB) [25] in their research on facial expression recognition. Some investigators created their own datasets for their experiments. The comprehensive analysis for the dataset papers examined is shown in Table 1, where FE: Facial Expressions, NP: Number of Participants, NI: Number of Images.

Table 1. Comprehensive Analysis of Datasets

Sr. No.	Name of Dataset	FE	NP	NI	Gray/ Color	Type	Age	QU
1	JAFFE [11]	7	123	593	Gray	Posed	18-30	No
2	CMU-MOSEI [13]	6	1000	3228 Videos	Color	Natural		No
3	MUG [19]	6	86	821	Color	Posed		No
4	Multi PIE [20]	6	337	750,000	Color	Pose	-	No
5	MMI Dataset [21]	5	75	250+	Color	Posed and Spontaneous	19 - 62	No
6	SFEW [23]	7	-	1766	Color	Spontaneous	1-70	No
7	AffectNet [24]	8	-	420299	Color	Pose	All	No
8	RaFD [25]	8	67	29672	Gray	Posed	--	No
9	FERG-3D-3B [37]	7	4	39574	Color	Frontal	-	No
10	IMPA-FACE3D [57]	8	38	534	Color	Posed	20 and 50	No
11	The Bosporus Database [57]	6	105	4666	Color	3D and 2D human face	25-35	No
12	Oulu-CASIA [58]	6	80	900+	Color	-		No
13	UIBVFED database [59]	6	32	640	Color	Virtual	20-80	No
14	FEI Face Dataset [60]	2	200	2800	Color	Posed	19 to 40	No
15	10K US Adult Faces Database [61]	15	2222	10,168		POsed	-	No
16	EB+ [62]	9	240	1261 videos	Color	Planned	18–60	No
17	WSEFEP [63]	6	30	210	Color	Posed		No
18	Indian Movie Face database (IMFDB) [64]	6	100	34512	Color	Spontaneous	-	No
19	IEMOCAP [65]	9	10	151 Videos	gray	Pose	All	No
20	**Proposed Dataset**	**5**	**100**	4132	**Gray**	**Natural**	**18-22**	**Yes**

It is evident from the literature that researchers put a lot of efforts in creating datasets. A literature review has revealed that there is scope to create a dataset which contains the images from the interview video considering six types of stressors. Also, the strategies used recently mainly focus on facial expression analysis by using posed and spontaneous images and hence actual emotions of an individual are difficult to identify. Proposed research focuses on the development of a unique dataset which comprises natural images by applying six stressors on an individual using video interview technique. Natural images encompass a broader range of settings and scenarios, while spontaneous images specifically focus on capturing genuine facial expressions and reactions as they occur naturally in response to stimuli or situations. A questionnaire prepared while conducting an interview by a psychologist examines different types of stressors namely Relationship Stressor, Personal Stressor, Life Event Stressor, Work or Wealth related Stressor, Traumatic Stressor and Social Stressor which can be useful in capturing natural expressions. Hence, we propose a dataset called the SMM Facial Expression dataset. To validate the impact of stressors on facial expressions CNN model is developed and implemented.

Kanade et al. [2] developed an Automatic Face Analysis model using permanent and transient facial features. Different facial action units like the inner, outer, and middle portions of the eyebrows, and lower and upper portions of eyelids, cheeks, and lips are used in expression recognition. They achieved 96.7% accuracy. Research suggested that a combination of the feature-based method and template-based method would give a more accurate result. Revina et al. [3] listed down different types of facial expression recognition techniques with major contributions and stated that the ROI method of segmentation was best suited for pre-processing giving 99% accuracy and the SVM classifier showed the most accurate result. Umarani and Srilakshmi [4] gave a brief introduction to different stages, techniques, and datasets used in the Facial expression recognition system. It was mentioned that expression recognition could be useful in criminal investigation, teaching and learning, healthcare, and Human-robot interfaces. It was also found that images could be sliced into two regions namely face region and non-face region. Nithya Roopa [8] used an inception net to identify facial expressions and achieved 35.6% accuracy and suggested that the same method could be applied to real-time expression recognition.

Zhang and Li [11] proposed a cross-connected Alex-Net enhanced convolutional neural network model in which the researcher added convolution and pooling layer in the original AlexNet. Mehendale [13] developed the FER model and used Expression Vector in the determination of facial expressions. It is also said that background removal helps in the identification of accurate expressions. An expression matrix is used in the proposed method to distinguish five types of expression. They suggested that the proposed technique can be beneficial in applications like a lie detector, and mood detectors.

Yue et al. [16] used the global average layer to create an extensible multi-channel deep convolutional neural network (CNN). The proposed system consists of four phases namely, Image Gaining, Face Locating, Face Recognition, and Expression Identification. The proposed method uses three different channels and improves the accuracy of Expression Net. There is a scope for improvement in network enactment and in reducing the size of the model. Christou and Kanojiya conducted the research [17] aimed to increase the accuracy of a FER2013 dataset. They classified input images into seven basic expressions by applying a convolutional neural network and confirmed the success of improving the accuracy of biometric applications. The overall accuracy of their proposed method is 91.12% and, 45% and 41% recognition rate for the classification of disgust and fear. Zhang et al. [18] proposed a method built on a convolution neural network as an image edge identification to detect facial expressions.

Minaee et al. [37] proposed an algorithm built on an attentional convolution network in which the author paid special attention to remarkable areas of the face like eyes, and mouth and found 70.02% accuracy for FER2013, 99.3% accuracy for FERG, 92.8% for JAFFE and 98.0% accuracy for CK+ dataset. The study conducted by Goren and Wilson [38] showed that one can identify expressions using the geometric shape of facial parts like mouth, eye, eyebrows, nose, etc. They suggested that one can improve the perception of facial expression by focusing on the collaboration between configuration and exterior appearance which helps to decide whether exterior perception is degraded due to configurable mechanism breakdown. Samadiani et al. [39] stated that illumination variation, subject dependence, and head pose are three main challenges in facial expression recognition (FER) and proposed a multimodal sensor framework for the automatic detection of FER. Song [40] proposed a novel idea of a fusion of dual-channel, i.e., ROI area division and Gabor feature to recognize facial expressions based on logical thinking and machine learning.

Lekdioui et al. [41] proposed a new facial decomposition method using regions of interest (ROI) to detect basic emotion state recognition. The second step is to resize and partition into blocks before feature extraction. Lastly, a multiclass SVM classifier is used to infer emotions. Kulkarni and Bagal [42] performed a comparative study of automatic Facial Expression Recognition by compensating for the influence of age on the recognition procedure. Log Gabor filter and Gabor filter were used in the proposed study for feature mining followed by the implementation of an SVM classifier for training and classification purposes. Matre and Shah [43] provided a brief description of different techniques used in facial expression detection with their advantages and disadvantages. The author mentioned that we can do facial expression recognition of black and white images as well as RGB images and concluded that the tensor perceptual color framework had the highest recognition rate and highest performance.

Debnath et al. [45] offered a convNet CNN classification model using feature extraction techniques like region-based oriented FAST, local binary pattern (LBP), and rotated BRIEF (BRB) techniques. Yao and Qiu [47] proposed hybrid features combining convolutional neural networks and dense SIFT features for low-pixel facial images. This model uses local features like eyes, eyebrows, and mouth to detect facial expressions.

From the literature, it is found that the retrieval of facial features, like geometry and texture features, is done using a variety of traditional approaches. Appearance-based feature extraction and feature-based feature extraction are the two categories of feature extraction methodologies. Local Binary Patterns (LBP) [26], Local Directional Patterns (LDA) [27], and the Fisher Face Method are considered feature-based extraction methods. Principle Component Analysis (PCA) [3], Gabor wavelet [28], Line Edge Map (LEM) [29], and Facial Action Units (FAC) [30] are vision-based feature extraction methods. On the other hand, different feature extraction methods are developed and used by researchers in facial expression such as the Angle and Distance method [31], DSAE [32], Wavelet Entropy [33], PCA + Fisher Face + HOG [34], Facial landmarks point [35], Face Alignment with Regression Tree [36] which produce different results.

According to the aforementioned literature analysis, there is still scope for advancement in the field of facial expression recognition. As an enhancement, the use of IoT and Human-Robot Interface (HRI) makes it possible to detect real-time expressions. The research focus is required on the recognition of facial landmarks. Design of automatic face alignment component in occlusion and dark light is possible. When recognizing facial expressions, consideration of multiple forms like age, gender, and personality attributes of the subject may improve accuracy. Genetic Property Evolution Framework designing is beneficial and may help in criminal detection, confidentiality breaches, etc. Subject dependence is a complicated issue addressed by researchers that involves a considerable range of data to acquire expressions of a subject at a particular age, but the classifier has a tendency to overfit and needs more attention in the future. Due to the extensive research, Large Pose Variation can be managed by 3-D models, and acquiring polarization-state imagery of faces is an efficient way of harnessing more textual and geometric data. There is a need to design a reliable standard FER system by adding several layers along with a different number of filters at each in the CNN model.

This research mainly focuses on the creation of the SMM Facial expression dataset and the development of the CNN model for FER via E2E (End-to-end) networks to process individual frames or collections of frames coming from the camera through various convolutional and pooling layers. In the context of machine learning, an end-to-end network refers to a model architecture, where the input data is directly transformed into the desired output, without the need for intermediate representations or separate processing stages. The model is capable of processing both spatial information (from individual frames) and temporal information (from sequences of frames) in its analysis. This model is then used for real-time Facial Expression Recognition and is predicted to be simple to handle. For this, we have created a dataset named SMM Facial expression dataset which captures the emotions of adults of 18-22 ages only using a questionnaire. The SMM Facial expression dataset is used to train the CNN model, which shows good accuracy in detecting facial expressions.

III. DATABASE DEVELOPMENT

In the proposed research, as mentioned above the questionnaire is prepared to study the facial expressions of students. To create a dataset, a questionnaire is prepared under the guidance of a psychological counselor. The questionnaire consists of questions related to facial expression identification. The age group considered in the research is from 18 to 22. Students aged 18 to 22 encounter a myriad of challenges as they navigate the transition from adolescence to adulthood. One of the primary issues they face is academic pressure, often stemming from the demands of higher education institutions and the competitive job market. Balancing coursework, extracurricular activities, and part-time jobs can lead to overwhelming stress and anxiety. Additionally, many students struggle with financial burdens, including tuition fees, housing costs, and student loan debt, which can hinder their ability to focus on their studies. Social pressures also play a significant role, as young adults strive to establish their identities while navigating peer relationships, romantic entanglements, and societal expectations. Mental health concerns, such as depression, anxiety, and loneliness, are increasingly prevalent among this age group, exacerbated by the challenges of adjusting to newfound independence and adulthood responsibilities. Moreover, substance abuse and unhealthy coping mechanisms often arise as misguided attempts to cope with these stressors. In essence, the journey from 18 to 22 is fraught with obstacles that require support, guidance, and resources to ensure the well-being and success of young students. The questionnaire is specially designed with the help of a psychologist.

The interview was conducted by a psychological counselor before the 12th std. and university exams because students were stressed at that time. Video of different subjects is captured using mobile and selfie sticks in a one-to-one and controlled manner. The controlled environment of an interview setting often ensures that both the camera and subjects remain relatively stationary throughout the recording process. As a result, the images captured during interviews exhibit higher clarity and sharpness, with fewer instances of distortion or blurring. This stability in the interview setting contributes to a lower prevalence of blurred images within the dataset, allowing for a higher proportion of clear and visually informative frames. We have considered the following questionnaire during database development:

i. Kindly introduce yourself. - This question may make students express a neutral emotion since it is a straightforward and common question that they have likely encountered before. There might be a sense of comfort in talking about themselves and their families, leading to a neutral or relaxed facial expression.

ii. What is one thing you are proud of, but would never tell anyone? - This question could evoke surprise or intrigue because it asks students to share a secret accomplishment. The idea of keeping something hidden might make them slightly curious or even amused, leading to expressions of surprise.

iii. Five years down the line where do you expect yourself to be? - This question might bring about a mix of emotions depending on each student's personal goals and plans. Some students may feel excited and happy when imagining a positive future, while others might feel uncertain or anxious about their plans, resulting in a variety of emotional expressions.

iv. Are you doing what you believe in or compromising? - This question may prompt thoughtful contemplation and potentially lead to expressions of concern or seriousness. It addresses personal values and decision-making, which can be significant and impactful for students, making them reflect on their choices and their alignment with their beliefs.

v. How do you react to any situation? (Positive or negative) - This question is broad and open-ended, leading to various emotional expressions depending on the specific situations students can recall. Positive experiences might result in expressions of happiness, while negative ones could evoke sadness, anger, or frustration.

vi. What is stress according to you? - This question delves into a complex and potentially sensitive topic, causing students to think about their own experiences with stress and its impact. They might express emotions like concern, seriousness, or even sadness as they consider the challenges they face.

vii. How do you hide your stress by speaking up or by being quiet? Why? - This question addresses coping mechanisms, and students may show expressions of introspection or thoughtfulness as they reflect on their ways of dealing with stress. It could lead to expressions of contemplation or vulnerability.

viii. What do you need to let go of and what are you thankful for? - This question has a reflective tone, leading students to think about things they want to release from their lives and the positive aspects they are grateful for. Expressions of gratitude, introspection, or even a mix of emotions might be seen in response to this question.

In the next step, using Python, the frames from the video are extracted. During conversion, no data is lost as 2 frames are selected for every second. Setting the frame rate to 2 frames per second (fps) can be advantageous for both the detection and removal of blurred images in certain contexts. While this frame rate may seem low compared to standard video frame rates, it offers specific benefits in scenarios where motion is minimal or where subtle changes in the scene occur over longer intervals. With a frame rate of 2 fps, each captured frame represents a significant sequential gap, allowing for clearer differentiation between static and moving elements within the scene. For the detection of faces in frames, the Haar cascade classifier [4] is employed, which is a machine learning-based approach widely used for object detection tasks, including face detection. This classifier works by using Haar-like features, which are simple rectangular filters that help identify patterns of intensity changes in the image. It has been trained on a large dataset of positive examples (faces) and negative examples (non-faces) to learn the patterns associated with a face. When applied to an input image, the classifier searches for these patterns to identify potential face regions.

Once the face regions are detected in a frame, they are cropped to create face-centric sub-images. This cropping process isolates the face, making it easier for subsequent analysis. In this case, the images are cropped to dimensions of 128x128 pixels, providing a standard size for further processing. To make the processing of the model faster and more efficient, the cropped face images are resized to a smaller dimension of 48x48 pixels. This downsizing reduces computational complexity and memory requirements, making it beneficial for real-time applications or when working with large datasets. The proposed model for expression recognition requires supervised data, meaning each image is manually annotated with the corresponding expression label, such as happy, sad, angry, etc. This mapping of images with their respective expressions is done manually by humans to provide ground truth labels for training. Before feeding the images into the model, a series of preprocessing steps are performed to ensure data quality and suitability for the expression recognition task. Firstly, blurred images are removed from the dataset, as they may hinder the model's ability to detect relevant facial features accurately. Pre-processing methods are used in the study to minimize blur artefacts and enhance image quality. To lower noise and maintain image characteristics, filters such as Gaussian and median are employed. These filters increase fidelity and clarity, which increases the efficacy of blur detection and removal techniques. Furthermore, they strengthen the algorithm resistance to noise disruption, guaranteeing precise detection and rectification of blur artefacts in later phases of picture processing. Secondly, images with closed eyes are eliminated to ensure the model focuses on analyzing expressions when the eyes are visible. Thirdly, images with face occlusion, where a significant portion of the face is obscured by objects, hands, or other elements, are also removed to avoid misleading the model.

Lastly, to simplify the data representation and reduce computational complexity, the images are converted from RGB (Red-Green-Blue) color space to grey-scale. Grayscale images contain only one channel of intensity values instead of three channels in RGB. This simplification retains the relevant facial features while reducing the computational burden on the model during training and inference. Figures 2 and 3 show the steps in dataset creation and image preprocessing. Once these preprocessing steps are completed, the dataset is ready to be used for training the facial expression recognition model.

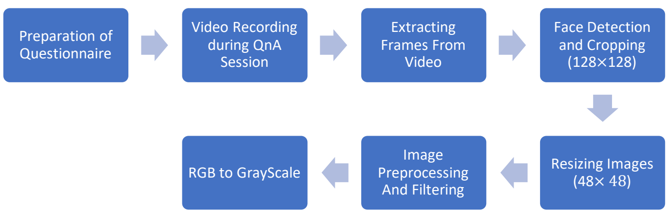

Figure 2. Dataset Creation Steps

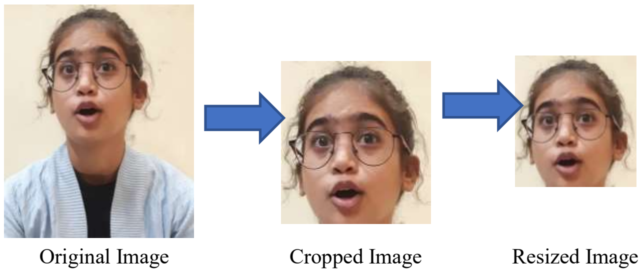

Figure 3. Cropping and resizing of an image

The SMM Facial Expression dataset contains neutral (947), happy (789), surprised (282), sad (330), and angry (482) expressions. The dataset is unbalanced, as shown by the number of images for each emotion. To train the model we need a balanced dataset. To make the dataset balanced, a resampling technique is used. There are two types of resampling – up-sampling and down-sampling. Figure 4 shows the frequency distribution of each emotion in the dataset.

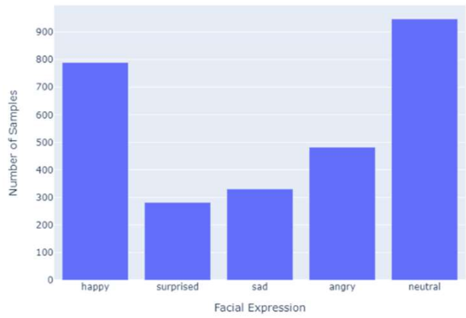

Figure 4. Data Distribution before up-sampling

Up-sampling helps in balancing the dataset by increasing the instances of rare classes. Up-sampling helps to lift the minority data by reducing the size of abundant class data. In our case, we have used the up-sampling technique to balance the dataset. After up-sampling dataset has neutral (947), happy (762), surprised (786), sad (791), angry (846) images of respective emotions. Figure 5 shows the frequency distribution of each emotion after the up-sampling.

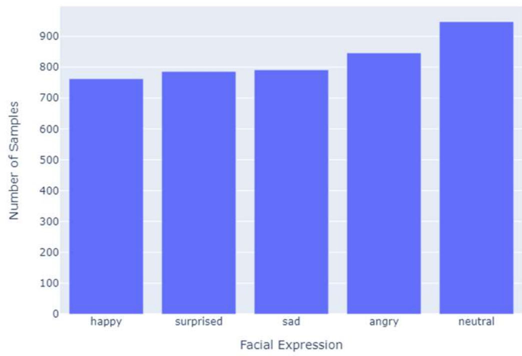

Figure 5. Data Distribution after up-sampling

Figure 6 shows sample images from an SMM Facial Expression dataset.

Neutral Happy Surprised Angry Sad

Figure 6. Sample images from an SMM Facial Expression dataset

IV. PROPOSED MODEL

The proposed neural network model detects facial expressions using a deep learning technique known as Convolutional Neural Networks (CNNs). The Convolutional Neural Network has had remarkable success in a variety of computer vision applications such as picture categorization, detecting objects, human behavior identification, recognition of facial expressions, face recognition, and so on. Convolution Neural network (CNN) is the most prevalent variety of deep learning architecture used in facial expression recognition. The primary components of CNN are convolution layers, pooling layers, and dense layers. The convolutional layer at different stages of the model extracts various local features in the input image. Computational costs also get reduced using CNN. CNN uses several kernels to mine features from the images. The kernel is a medium that traverses the input image, executes a dot product with a sub-region of the input image, and obtains the output as a matrix of dot products. The kernel is of different sizes viz. 3×3, 5×5, 7×7, etc. The convolution layer is defined as follows:

$$\mathrm{z}^{\mathrm{l}} = \mathrm{h}^{\mathrm{l}-1} * \mathrm{w}^{\mathrm{l}} \quad , \tag{1}$$

where the layer l's pre-activation output is denoted as z^l, h^{l-1} signifies activation of layer l, * specifies the discrete convolutional operator, and symbolizes convolution filter weights of layer l. CNN layers are responsible for the processing of the input images. It combines the input image with a filter and retrieves the features from the input image. The proposed model comprises 3-CNN layers having 256, 256, and 128 filters at each layer. The pooling layer is a key pillar of CNN. It is present in between the two convolution layers. It reduces the dimensionality of the feature vector. Max pooling and average pooling are the two types of pooling layers. In our approach, max pooling is used. For max pooling to function, the highest value from each pool is selected. Max pooling keeps the most prominent features of the feature map, and the final image is sharper than the original. Max pooling is defined as follows:

$$\mathrm{h}_{\mathrm{xy}}^{\mathrm{l}} = \max_{\mathrm{i}=0,\dots\mathrm{s}} \max_{\mathrm{j}=0,\dots\mathrm{s}} \mathrm{h}_{(\mathrm{x}+\mathrm{i})(\mathrm{y}+\mathrm{j})}^{\mathrm{l}-1} \,, \tag{2}$$

where h_{xy}^{l} represents the output at position xy of layer l, $h_{(x+i)(y+j)}^{l-1}$ are the values from the input feature map at the corresponding positions *(x+i)(y+j)* of layer l.

The dense layer means the fully connected layer is followed by the pooling layer and used as an output layer. It receives input from the previous layer. It means that the dense layer neurons are connected to all the neurons of its preceding layer. The proposed model consists of three dense layers of vector sizes 128, 64, and 32. To train the model, the kernel of size 33 and a batch size of 32 are used. Batch size helped to perform the grouping of datasets into batches. Number of epochs castoff are 100. A fully connected layer is calculated as follows:

$$z^{l} = W^{l}.h^{l-1} \,, \tag{3}$$

where z^l represents the output of dense layer *l*, and *l*, W^l is the weight matrix that contains learnable parameters of layer *l*, and h^{l-1} is the input to the dense layer. The dot operator denotes an activation function applied element-wise to the linear transformation. CNN features like batch normalization, activation, and dropout achieved rich performance. A sequential model used as layer-by-layer model building is preferable. Figure 7 shows the proposed model architecture.

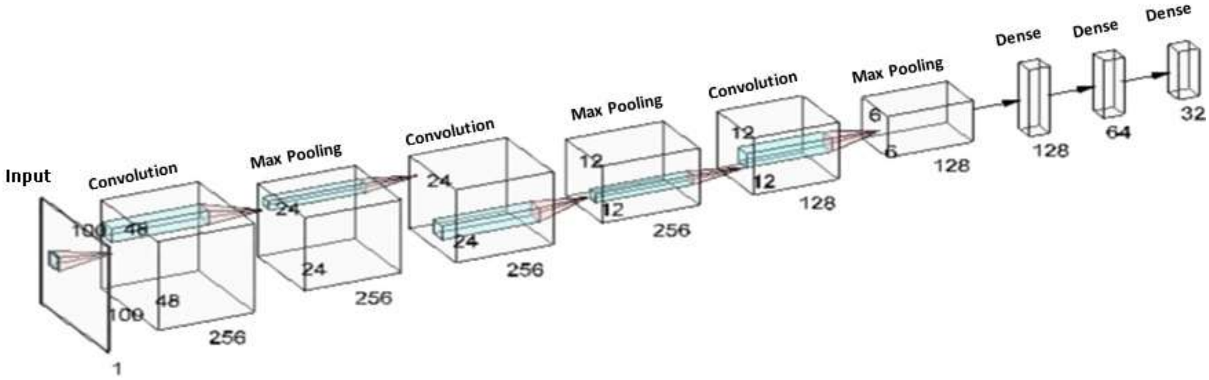

Figure 7. Design of the proposed model

Adam optimizer is used in the projected model, which helps to adjust the learning rate during model training. Adam is an adaptive learning rate method, and the weights of each neuron are updated by adjusting the learning rate adaptively for every single parameter in the model based on the history of gradients calculated for that parameter. This helps the optimizer converge faster and more accurately than the fixed learning rate. It utilizes the squared gradients for scaling the learning rate and it takes advantage of momentum by utilizing the gradient moving average rather than the gradient itself. The memory requirements of the Adam optimizer are very low and take less time to compute the weight values of neurons. A standard learning rate is used throughout model implementation. Sigmoid, tanh, leakyReLU, ReLU and ELU are different types of activation functions. In the proposed methodology, the ReLU Activation function is used during enactment to get better accuracy. The selection of the ReLU activation function was informed by a thorough review of the existing literature in the field of deep learning and neural network architectures. ReLU has been widely recognized for its effectiveness, which can impede the training of deep neural networks, particularly in deep convolutional architectures commonly used in image processing tasks. In tandem with our literature review, we conducted extensive experimentation to validate the suitability of ReLU. Performance of ReLU was systematically compared with the alternative activation functions, including sigmoid, tanh, and variants of the Leaky ReLU. Through rigorous analysis, it is observed that ReLU consistently outperforms alternative activation functions in terms of convergence speed, model accuracy, and generalization across diverse datasets and experimental conditions. Mathematically, the ReLU Activation function ($f(x)$) is given as follows:

$$f(x) = \max(0, x), \tag{4}$$

where x denotes the obtained nonlinear features from z^l. For the classification of expressions, the softmax activation function is obsolete. The loss function utilized in this multiclass classification problem is categorical cross-entropy. An image can be categorized as containing feelings of happiness, sadness, anger, surprise, and neutrality. The softmax function ($\sigma_i(z)$) is defined as follows:

$$\sigma_i(z) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}}, \tag{5}$$

where j and K denote the neuron number and final neuron in the softmax function. The proposed model works in the following manner. Initially, input is taken by the model. After pre-processing the images, the model gets trained. Feature extraction is an inbuilt process in CNN. Testing of the performance of the proposed model is done using the FER2013 and CK+ datasets and it is found that input image expression gets classified. Figure 8 illustrates the preprocessing and training of the model.

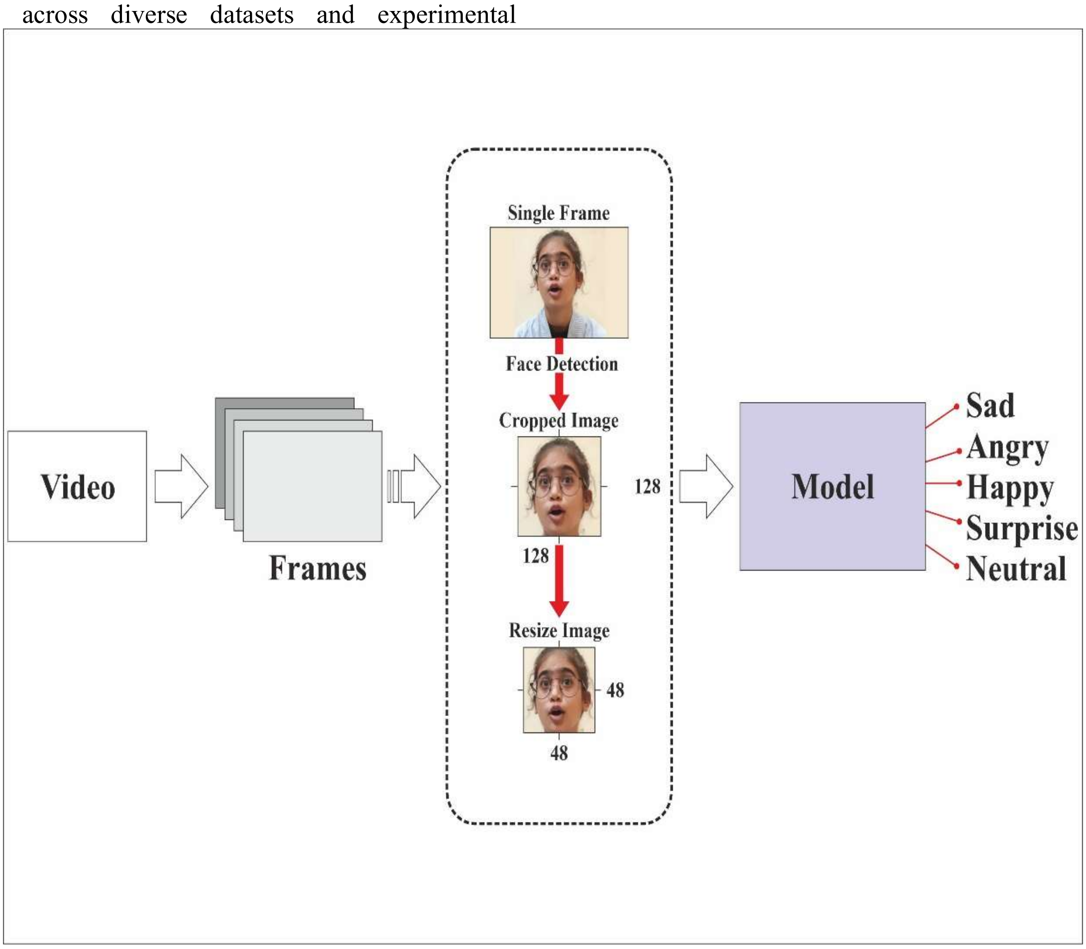

Figure 8. Proposed Methodology

V. EXPERIMENT OUTCOMES

This section gives you a transitory overview of the implementation analysis of the proposed model on available datasets for facial expression recognition. Firstly, we deliver a summary of databases used in this research, secondly, we perform the proposed model on two databases, and lastly, we compare the results with some of the promising recent works.

A. Databases

To evaluate the model, FER2013 and CK+ datasets are used. The steps taken to implement the proposed model are detailed below.

FER2013: It is a.csv file that stores the pixel values of each greyscale image of size 48X48. The dataset contains labeled images of seven different emotions namely Angry, Happy, Fearful, Disgusted, Sad, Surprised, and Neutral. Google image search API is used to create a database and faces are automatically registered [37]. FER2013 images vary, including those with low contrast, partial faces, faces with spectacles, and face blockage by hands. Model images of FER2013 are shown in Figure 9.

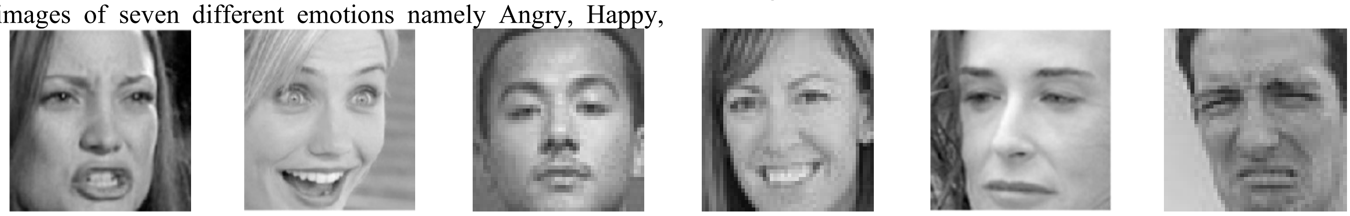

Figure 9. Model imageries from FER2013

CK+: This is the second dataset which contains 920 adapted gray scale images. Images are of seven different categories viz. anger, contempt, happiness, disgust, fear, sadness, and surprise. Images available are of size 48X48 png format. It is a laboratory-controlled facial expression classification database. Figure 10 shows model images from this dataset.

Figure 10. Model imageries from CK+

B Comparison and Analysis of Experiment

We demonstrate the effectiveness of the suggested model on the aforementioned datasets in this section. Every time, the model is trained using a subset of training data, and its accuracy is assessed using a subset of test data.

The system used to implement the model has the configuration of Intel® Core™ i5-8250U CPU @ 1.60GHz 1.80 GHz processor, 64-bit Windows10 Operating system with 20GB RAM. OpenCV and Python were used to implement the model. We have evaluated our proposed model on the FER2013 dataset. FER2013 is not a balanced dataset as compared to other datasets because it contains 4963 images of expressions like Anger, 8989 images of Happiness, 547 images of Disgust, 5121 images of Fear, 6071 images of Sadness, 6198 images of Neutrality, and 4002 images of Surprise used in the implementation. It is easily available on Kaggle and used in several types of research. We succeeded in obtaining an accuracy of 82.18% on the FER2013 dataset. Figure 11 shows the comparison of the proposed technique with some previous works on the FER2013 dataset.

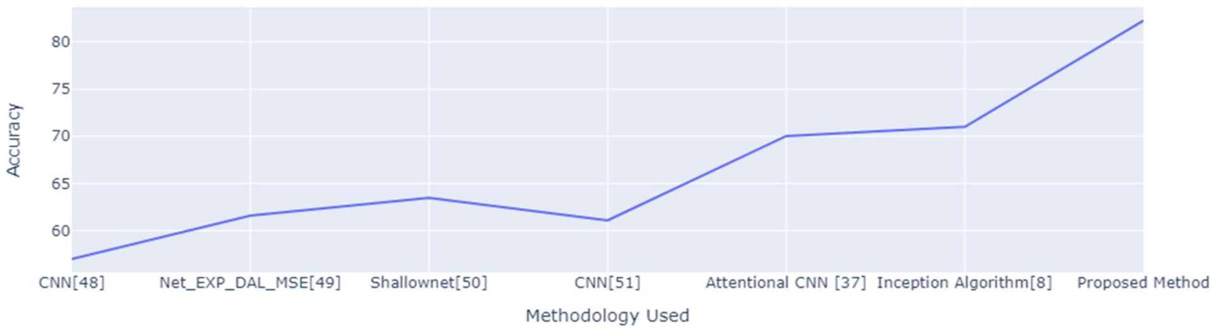

Figure 11. Accuracy comparison on the FER2013 dataset

CK+: CK+ datasets are used to test the model. The images are of seven different categories which are happiness, sadness, anger, disgust, contempt, fear, and surprise. We dropped classes such as disgust and contempt from the testing dataset because our proposed model was not trained on these classes. Available images are of size 48X48.png format. The result shows that the proposed model gives the best accuracy of 85.27% for all the expressions excluding sadness. Figure 12 shows the assessment of the proposed method with some earlier works on the CK+ dataset.

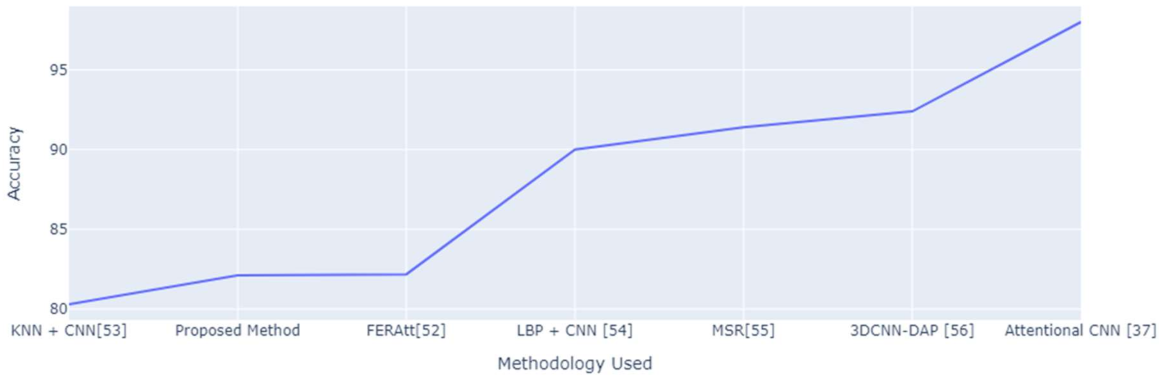

Figure 12. Accuracy comparison on the CK+ dataset

In the SMM Facial Expression Dataset, it is observed that after up-sampling out of 4132 images, 3991 images are classified correctly. Anger, Happiness, Neutrality, Sadness, and Surprise are different expressions captured and tested using the proposed method. Neutral expression gives the highest accuracy. Approximately 16% of the surprised expressions are recognized as happiness. From the confusion matrix, it is perceived that the model gets confused in angry and sad emotions. To overcome this problem, there is a necessity for improvement of the image capturing. In the proposed model, multiple convolutional layers are used which are responsible for filtering and scanning the feature matrix completely. Hence, we get better accuracy for the SMM facial expression dataset. The dataset shows an accuracy of 96.60%. The proposed system confusion matrix on the SMM Facial Expression dataset is shown in Figure 13.

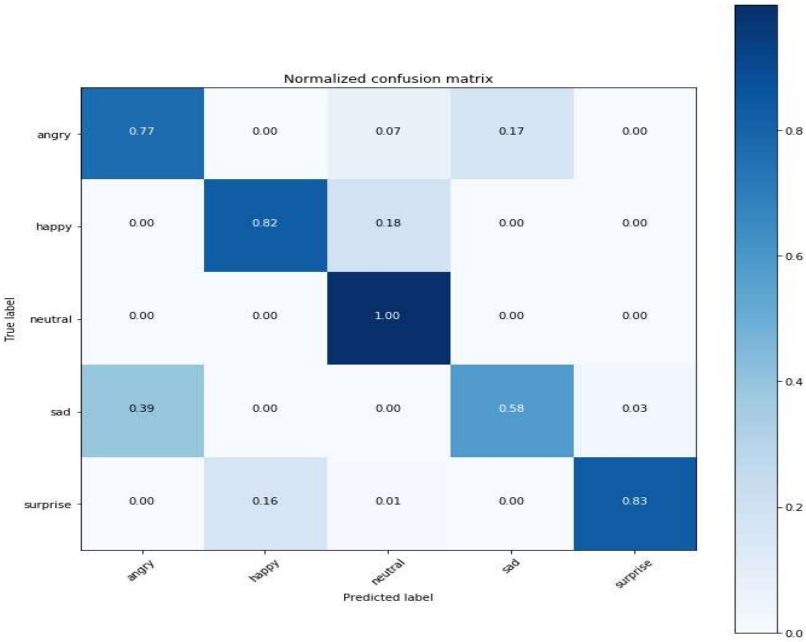

Figure 13. The proposed model confusion matrix using the SMM Facial Expression Dataset

Table 2 illustrates a comparison of the proposed method to recent work on facial expression recognition and classification. The performance of the proposed model was achieved by pre-processing the collected frames before the classification. Moreover, the proposed model applies efficient database development steps to address common challenges in facial emotion recognition systems, such as resizing, occluded image removal, noise removal, and blurred images. The proposed method utilizes the Adam optimizer in the existing CNN model for updating the weights of each neuron. This algorithm improves efficiency by reducing training time and considering the relevant relationships between data features. By combining these factors, the proposed method results in higher performance. This significantly reduces the iterations of the proposed CNN. This also demonstrates the efficient use of the developed system.

Table 2. Comparison of the proposed method to recent work on facial expression recognition and classification

Datasets	Classification Model	Accuracies	References
JAFFE	ALEXNET CNN	89%	Zhang et al. [11]
Karolinska Directed Emotional Faces (KDEF) datasets	CNN	98.91%	Tarnowski et al. [5]
CK+, FERG	Attentional CNN	98.0%, 99.3%	Minaee et al. [37]
FER2013	CNN	91.01%	Debnath et al. [45]
Kaggle's ICMP 2013-Fecial Expression Recognition (FER) dataset	CNN	66.7%	Khopkar et al. [46]
MMI	ALEXNET	77.9%	Mollahosseini et al. [51]
CMU Database	CNN	78%	Mehendale [13]
SMM Facial Expression Dataset developed in the current study	CNN	96.60%	Current Study

VI. CONCLUSION AND FUTURE ENHANCEMENT

The proposed CNN approach can process individual frames or groups of frames arriving from a camera with an accuracy rate of 96.60% for facial expression recognition from E2E (End-to-end) networks. The significant research contributions refer to: (1) the questionnaire developed to capture emotions while talking on a camera; (2) a labeled SMM Facial Expression Dataset created using a questionnaire with the help of a psychological counselor; (3) the CNN model developed to capture and classify individual frames arriving from a camera; (4) the predicted findings showing how successfully the suggested approach handles all the complexity and produces encouraging results; (5) it is perceived that the proposed model is applicable to train.csv files and images; (6) hence, it is proved that the SMM facial dataset is up to the mark like previous benchmarked datasets, and researchers can use it for further studies. Since most of the participants in the proposed study are speaking quietly while the video is being recorded, it is difficult to map facial expressions of fear and disgust. As a result, the database development assumes that the subjects are angry. Therefore, only five classes are used in the dataset.

In future enhancement, we can add images of expressions like fear and disgust and make the SMM facial expression dataset a complete dataset. Other advanced methods can also be applied to check the validity of the SMM facial expression dataset. One can also design a sensor to detect human emotions. The development of the biometric machine with facial expression recognition is also helpful for organizations.

References

[1] M. Albert, *Nonverbal Communication*, Routledge, 2017, https://doi.org/10.4324/9781351308724.

[2] Y. L. Tian, T. Kanade, and J. F. Cohn, "Recognizing lower face action units for facial expression analysis," *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (Cat. No. PR00580), Grenoble, France, 2000, pp. 484-490, https://doi.org/10.1109/AFGR.2000.840678.

[3] I. M. Revina and W. R. S. Emmanuel, "A survey on human face expression recognition techniques," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, issue 6, pp. 619–628, 2021, https://doi.org/10.1016/j.jksuci.2018.09.002.

[4] V. Umarani and M. Srilakshmi, "Face emotion recognition: A brief review," *International Journal of Creative Research Thoughts*, vol. 9, issue 11, pp. 766–771, 2021, [Online]. Availablenat: www.ijcrt.org.

[5] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017, https://doi.org/10.1016/j.procs.2017.05.025.

[6] M. H. Siddiqi et al., "A *Brief Review of Facial Emotion Recognition Based on Visual Information*," *Proceedings of the 2018 JEEMA CeTechNxT*, vol. 5, issue 1, pp. 196–201, 2018.

[7] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, and Z. Lai, "Using the idea of the sparse representation to perform coarse-to-fine face recognition," *Information Sciences*, vol. 238, pp. 138–148, 2013, https://doi.org/10.1016/j.ins.2013.02.051.

[8] S. Nithya Roopa, "Emotion recognition from facial expression using deep learning," *International Journal of Engineering and Advanced Technology*, vol. 8, special issue 6, pp. 91–95, 2019, https://doi.org/10.35940/ijeat.F1019.0886S19.

[9] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach," *Knowledge-Based Systems*, vol. 110, pp. 1–14, 2016, https://doi.org/10.1016/j.knosys.2016.07.011.

[10] P. L. C. Courville, A. Goodfellow, I. J. M. Mirza, Y. Bengio, FER-2013 Face Database. Universite de Montreal: Montréal, QC, Canada, 2013.

[11] Z. Zhang and M. Li, "Research on facial expression recognition based on neural network," *Proceedings of the 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, Xi'an, China, 2020, pp. 78-81, https://doi.org/10.1109/ICCNEA50255.2020.00025.

[12] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression Patrick," *Proceedings of the 2010 IEEE Computer Society Conference Workshops on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 94-101, https://doi.org/10.1109/CVPRW.2010.5543262.

[13] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Applied Sciences*, vol. 2, issue 3, pp. 1–8, 2020, https://doi.org/10.1007/s42452-020-2234-1.

[14] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, "Facial expression recognition using a pairwise feature selection and classification approach," *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016, pp. 5149-5155, https://doi.org/10.1109/IJCNN.2016.7727879.

[15] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on UB KinFace database," *Proceedings of the Workshops on Computer Vision and Pattern Recognition CVPR 2011*, Colorado Springs, CO, USA, 2011, pp. 60-65, https://doi.org/10.1109/CVPRW.2011.5981801.

[16] Z. Yue, F. Yanyan, Z. Shangyou, P, Bing, "Facial expression recognition based on convolutional neural network," *Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, 2019, pp. 410-413, https://doi.org/10.1109/ICSESS47205.2019.9040730.

[17] N. Christou and N. Kanojiya, "Human facial expression recognition with convolution neural networks," In: X.S. Yang, S. Sherratt, N. Dey, A. Joshi (Eds.), *Proceedings of the Third International Congress on Information and Communication Technology, Advances in Intelligent Systems and Computing*, Springer, Singapore, 2019, vol. 797. https://doi.org/10.1007/978-981-13-1165-9_49.

[18] H. Zhang, A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," *IEEE Access*, vol. 7, pp. 159081–159089, 2019, https://doi.org/10.1109/ACCESS.2019.2949741.

[19] N. Aifanti, C. Papachristou, and A. Delopoulos, "The MUG facial expression database," *Proceedings of the 11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, Desenzano del Garda, Italy, 2010, pp. 1-4.

[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, issue 5, pp. 807–813, 2010, https://doi.org/10.1016/j.imavis.2009.08.002.

[21] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME 2005*, vol. 2005, issue July, pp. 317–321, 2005, https://doi.org/10.1109/ICME.2005.1521424.

[22] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, Santa Barbara, CA, USA, 2011, pp. 921-926, https://doi.org/10.1109/FG.2011.5771374.

[23] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, 2011, pp. 2106-2112, https://doi.org/10.1109/ICCVW.2011.6130508.

[24] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, Valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, issue 1, pp. 18–31, 2019, https://doi.org/10.1109/TAFFC.2017.2740923.

[25] S. Li, W. Deng, and J. P. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2584-2593, https://doi.org/10.1109/CVPR.2017.277.

[26] C. Shan, S. Gong, P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, issue 6, pp. 803–816, 2009, https://doi.org/10.1016/j.imavis.2008.08.005.

[27] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, vol. 32, issue 5, pp. 784–794, 2010, https://doi.org/10.4218/etrij.10.1510.0132.

[28] H. Wang, J. Gao, L. Tong, and L. Yu, "Facial expression recognition based on PHOG feature and sparse representation," *Proceedings of the 2016 35th Chinese Control Conference (CCC)*, Chengdu, China, 2016, pp. 3869-3874, https://doi.org/10.1109/ChiCC.2016.7553957.

[29] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764-779, 2002, https://doi.org/10.1109/TPAMI.2002.1008383.

[30] M. H. Alkawaz, D. Mohamad, A. H. Basori, and T. Saba, "Blend shape interpolation and FACS for realistic avatar," *3D Research*, vol. 6, issue 1, 2015, https://doi.org/10.1007/s13319-015-0038-7.

[31] M. H. Siddiqi, M. Alruwaili, J. H. Bang, and S. Y. Lee, "Real time human facial expression recognition system using smartphone," *IJCSNS International Journal of Computer Science and Network Security*, vol. 17, issue 10, pp. 223–230, 2017.

[32] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018, https://doi.org/10.1016/j.neucom.2017.08.043.

[33] S. H. Wang, P. Phillips, Z. C. Dong, and Y. D. Zhang, "Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm," *Neurocomputing*, vol. 272, pp. 668–676, 2018, https://doi.org/10.1016/j.neucom.2017.08.015.

[34] M. Asad, S. O. Gilani, and M. Jamil, "Emotion detection through facial feature recognition," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 12, issue 11, pp. 21–30, 2017, https://doi.org/10.14257/ijmue.2017.12.11.03.

[35] M. Merlin Steffi and J. John Raybin Jose, "Comparative analysis of facial recognition involving feature extraction techniques," *International Journal of Computer Sciences and Engineering*, vol. 6, issue 2, pp. 81–86, 2018, [Online]. Available at: http://www.ijcseonline.org/spl_pub_paper/NCTT-2018-18.pdf.

[36] V. M. Alvarez, R. Velazquez, S. Gutierrez, and J. Enriquez-Zarate, "A method for facial emotion recognition based on interest points," *Proceedings of the 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*, San Salvador, El Salvador, 2018, pp. 1-4, https://doi.org/10.1109/RICE.2018.8509055.

[37] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, issue 9, 3046, 2021, https://doi.org/10.3390/s21093046.

[38] D. Goren and H. R. Wilson, "Quantifying facial expression recognition across viewing conditions," *Vision Research*, vol. 46, issue 8–9, pp. 1253–1262, 2006, https://doi.org/10.1016/j.visres.2005.10.028.

[39] N. Samadiani, G. Huang, B. Cai, W. Luo, C. Chi, Y. Xiang, and J. He, "A review on automatic facial expression recognition systems assisted by multimodal sensor data," *Sensors*, vol. 19, issue 8, 1863, 2019, https://doi.org/10.3390/s19081863.

[40] Z. Song, "Facial expression emotion recognition model integrating philosophy and machine learning theory," *Frontiers in Psychology*, vol. 12, 2021, https://doi.org/10.3389/fpsyg.2021.759485.

[41] K. Lekdioui, Y. Ruichek, R. Messoussi, Y. Chaabi, and R. Touahni, "Facial expression recognition using face-regions," *Proceedings of the 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, Morocco, 2017, pp. 1-6, https://doi.org/10.1109/ATSIP.2017.8075517.

[42] K. R. Kulkarni and S. B. Bagal, "Facial expression recognition," Proceedings of the 12th 2015 International Conference on Information Processing (ICIP), 2015, pp. 535-539. https://doi.org/10.1109/INFOP.2015.7489442.

[43] G. N. Matre and S. K. Shah, "Facial expression detection," *Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Computing Research*, Enathi, India, 2013, pp. 1-3, https://doi.org/10.1109/ICCIC.2013.6724242.

[44] J. Avanija, K. R. Madhavi, G. Sunitha, S. C. Sangapu, and S. Raju, "Facial expression recognition using convolutional neural network," *Proceedings of the 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, 2022, pp. 336–341, https://doi.org/10.1109/ICAITPR51569.2022.9844221.

[45] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Scientific Reports*, vol. 12, issue 1, pp. 1–18, 2022, https://doi.org/10.1038/s41598-022-11173-0.

[46] A. Khopkar and A. A. Saxena, "Facial expression recognition using CNN with Keras," *Bioscience Biotechnology Research Communications*, vol. 14, issue 5, pp. 47–50, 2021, https://doi.org/10.21786/bbrc/14.5/10.

[47] F. Yao, and L. Qiu, "Facial expression recognition based on convolutional neural network fusion SIFT features of mobile virtual reality," *Wireless Communications and Mobile Computing*, vol. 14, issue 1–6, pp. 253–267, 2021, https://doi.org/10.1155/2021/5763626.

[48] V. Tümen, Ö. F. Söylemez, and B. Ergen, "Facial emotion recognition on a dataset using convolutional neural network," *Proceedings of the 2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Turkey, 2017, pp. 1-5, https://doi.org/10.1109/IDAP.2017.8090281.

[49] Y.-K. Zhai and J. Liu, "Facial *Expression Recognition based on Transfer ring Convolutional Neural Network*," *Journal of Signal Processing*, pp. 729–738, 2018.

[50] Y. Fang, *Research of Facial Expression Recognition Based on Convolutional Neural Networks*, Dissertation, Xidian University, 2017.

[51] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 2016, pp. 1-10, https://doi.org/10.1109/WACV.2016.7477450.

[52] P. D. M. Fernandez, F. A. G. Pena, T. I. Ren, and A. Cunha, "FERAtt: Facial expression recognition with attention net," *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 837-846, https://doi.org/10.1109/CVPRW.2019.00112.

[53] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," *Proceedings of the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, London, UK, 2017, pp. 123-128, https://doi.org/10.1109/SERA.2017.7965717.

[54] S. Sawardekar and S. R. Naik, "Facial expression recognition using efficient LBP and CNN," *International Research Journal of Engineering and Technology (IRJET)*, issue June, pp. 2273–2277, 2018.

[55] S. Rifai, Y. Bengio, A. Courville, P. Vincent, M. Mirza, "Disentangling factors of variation for facial expression recognition," In: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid, (Eds.), Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol. 7577, 2012, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33783-3_58.

[56] M. Liu, S. Li, S. Shan, R. Wang, X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," *Proceedings of* the Asian Conference on Computer Vision, 2014, 143-157. https://doi.org/10.1007/978-3-642-33783-3_58.

[57] F. Adjailia, et al., "Integration of 2D textural and 3D geometric features for robust facial expression recognition," *Computing & Informatics*, vol. 40, issue 5, pp. 988–1007, 2021. https://doi.org/10.31577/cai_2021_5_988.

[58] J. Haddad, O. Lezoray, P. Hamel, "3D-CNN for facial emotion recognition in videos," In: Bebis, G., et al. Advances in Visual Computing. ISVC 2020. Lecture Notes in Computer Science, vol. 12510, 2020. Springer, Cham. https://doi.org/10.1007/978-3-030-64559-5_23.

[59] M. M. Oliver, and E. A. Alcover, "UIBVFED: Virtual facial expression dataset," Plos one, vol. 15, issue 4, e0231266, 2020. https://doi.org/10.1371/journal.pone.0231266.

[60] Y. Lin, and H. Xie, "Face gender recognition based on face recognition feature vectors," *Proceedings of the 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2020. https://doi.org/10.1109/ICISCAE51034.2020.9236905.

[61] M. K. Scheuerman, K. Wade, C. Lustig, and J. R. Brubaker, "How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis," *Proceedings of the* ACM Hum.-Comput. Interact, vol. 4, issue CSCW1, Article 58, pp. 1-35, 2020. https://doi.org/10.1145/3392866.

[62] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, Q. Ji, "Cross-domain AU detection: Domains, learning approaches, and measures," *Proceedings of the Int Conf Autom Face Gesture Recognit*, May 2019. https://doi.org/10.1109/FG.2019.8756543. Epub 2019 Jul 11. PMID: 31749665; PMCID: PMC6867108.

[63] M. Olszanowski, et al., "Warsaw set of emotional facial expression pictures: A validation study of facial display photographs," *Frontiers in Psychology*, vol. 5, Frontiers Media SA, 2015. https://doi.org/10.3389/fpsyg.2014.01516.

[64] P. Mehta, A. Kumar and S. Bhattacharjee, "Fire and gun violence based anomaly detection system using deep neural networks," *Proceedings of the2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 199-204, https://doi.org/10.1109/ICESC48915.2020.9155625.

[65] S. Tripathi, S. Tripathi, and H. Beigi, "Multi-modal emotion recognition on IEMOCAP dataset using deep learning," arXiv preprint arXiv:1804.05788, 2018.

***MAYURI M. BAPAT** pursuing her Ph.D. from Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India. in Computer Science. She is working as an Assistant Professor at MAEER's MIT Arts, Commerce and Science College, Alandi(D), Pune. She is having an experience of 15+ years of graduate and postgraduate level teaching. Her research area of interest is Natural Language Processing, Digital Image Processing, and Machine learning. She has published 15 research papers in different national and international conferences and journals. She has completed the BCUD project which received a grant of 1.10L.*

***DR. CHANDRASHEKHAR H. PATIL** received his PhD in Computer Science under the domain of Digital Image Processing, from the Department of Computer Science, Bharati Vidyapeeth Deemed University, Pune. He is an Associate Professor at the Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India. He has over 17+ years of academic experience in research, innovation, and UG/PG-level teaching. His six (01) Indian patents are under the granting stage, one (01) International patent and one(01) Indian patent has been granted, and he authored/co-authored over 52+ referred journals, book chapters, and conference papers in reputed international journals and conferences indexed by Scopus/SCIE/UGC. The current publication citations of his research are more than 94+, 48+, and 29+, with an h-index of 4, 2, and 1 as per Google Scholar, Scopus, and Web of Science, respectively. His specialization is OCR, Hyperspectral-Multispectral Remote Sensing, GIS, and Remote Sensing of Soil, crops, urban areas, etc. His research interests include Geospatial Technology, Digital Image Processing, Pattern Recognition, Big Data Analysis, the Internet of Things (IoT), Machine Learning, and Artificial Intelligence. He is a reviewer for reputed journals and calendar-conferences.*

***PROF. SHANKAR MALI** is currently working as a Professor at the Department of Computer Science, Dr. Vishwanath Karad MIT World Peace University. He is also handling additional responsibilities as a Director of IQAC at MIT-WPU. He has experience of 23 years teaching MCA and M. Sc. (Computer Science) students. He received a Ph. D. degree from Solapur University in Computer Science (2011) and qualified SET in Computer Science in year 2004. Two students have completed Ph. D. under his guidance from Savitribai Phule Pune University, Pune. Presently guiding two Ph.D. students at MIT World Peace University, Pune. His research areas include Digital Image Processing, Pattern Recognition, and Natural Language Processing*