A Multi-Model Framework for Grading of Human Emotion Using CNN and Computer Vision

Praveen Kulkarni, Dayanand Sagar University, India

Rajesh T. M., Dayananda Sagar University, India

ABSTRACT

Emotion analysis is an area that has been widely used in the forensic crime detection domain, a mentoring device for depressed students, psychologically affected patient treatment. The current system helps only in identifying the emotions but not in identifying the level of emotions like whether the individual is truly happy/sad or pretending to be happy/sad. In this proposed work, a novel methodology has been introduced. The authors have rebuilt the traditional local binary pattern (LBP) feature operator to image the expression and combine the abstract characteristics of facial expression learned from the neural network of deep convolution with the modified features of the texture of the LBP facial expression in the full connection layer. These extracted features have been subjected as input for CNN Alex Net to classify the level of emotions. The results obtained in this phase are used in the confusion matrix for analysis of grading of emotions like Grade-1, Grade-2, and Grade-3 obtained an accuracy of 87.58% in the comparative analysis.

KEYWORDS

CNN Alex Net, Computer Vision, Decision Tree, Euclidean Distance, Fuzzy Logic, Grading of Emotions, KNN, LBP

INTRODUCTION

People communicate with one another mostly through speech, but they also use bodily gestures to emphasize specific points in their speech and to express feelings. Facial articulations, which are an important aspect of communication, are one of the most important ways people express their feelings. Despite the fact that nothing is said vocally, the signals we transmit and receive via nonverbal contact contain a lot of information. Facial articulations allow for nonverbal communication and play an important role in between-home relationships. In typical human-machine interfaces, programmed facial articulation responses can play an important role. It could be employed in both social research and clinical practice. Despite the fact that people can recognize facial joints nearly instantly, reliable machine recognition remains a challenge. In prior work, a throw analysis was performed using well-known and widely utilized algorithms (Kulkarni et al., 2020), and many studies focused just on

emotion recognition rather than grading. Previous work on grading emotion using a combination of LBP and the KNN algorithm (Kulkarni et al., 2021) yielded findings that were not promising, with an accuracy of only 79 percent. This has prompted us to develop the methods proposed in this paper. Decision trees and fuzzy algorithms were used to train the model, which took into account 19 aspects of the human face. Using machine learning methods, the emotion detection accuracy was not up to par.

Despite significant progress in the field of emotion analysis, in areas such as mentoring for depressed students, criminal evaluation systems, psychologically affected patient treatment, and musical therapy for desperate and mentally disabled patients, the assessment of the classification of the person's emotion expressed on his or her face is currently lacking. Grading human emotion can be useful in ensuring that the person remaining in front of the camera isn't just a two-dimensional representation, but is actually expressing a genuine emotion.

The proposed feature points were chosen through a thorough literature review of different feature points that contribute to happy and sad emotions. Emotions are graded by comparing the number of features that match the threshold value. The threshold value is determined by comparing the number of extracted features to the number of matched features. Using a happy emotion as an example, we have subcategorized happy emotions into grade 1 happy, grade 2 happy, and grade 3 happy emotions. Whereas a grade of 1 implies that the person is trying to be happy, and the features matched for this emotion among 19 features are about 6 to 8. That is, if 6 to 8 features are matched, we consider it grade 1. Similarly, grade 2 emotion is nothing, although the person is normally joyful. The threshold value has been set at 12 to 14 features. If more than 12 features match, we consider the person to be normal joyful. Similarly, more than 15 elements should match for the grade 3 feeling, which indicates that the person is extremely delighted. This allows us to discern the actual emotions in a person's face. The same procedure is followed with sad feelings. We proposed a novel methodology for calculating the performance of each model in distinguishing facial expression and rating the level of emotion in happy and sad emotions using LBP inputs by applying CNN and varying its depths. We used the MMI research database, the Japanese Female Facial Expression Database (JAFFE), and some real-time expression movies in our studies. We'll look at how the highlights are separated and changed for decision tree computations in this lesson. We'll look at component extraction calculations and approaches from a variety of sources.

BACKGROUND

Outward appearance-based emotions grouping moves the following grade the familiarity of the climate, precision and validity of connection occurring in the environmental factors, particularly to exhibit human-PC communication complexities as shown by (Rota et al., 2017) in his strategy identified with molecule groupings. To consider these contemplations, the two researchers and analysts from the local area are contributing significant endeavors to outward appearance-based emotion order methods and the writing is expanding with the progression of time. (Jain et al., 2017) presented a calculation dependent on development and most recent Deep Convolution Neural Networks (DNNs) that is made of different layers performing various capacities and profound leftover squares to accomplish various assignments of interest.

(Alreshidi et al., 2020) the researcher focused on the seven most common facial expressions that are employed often in everyday life. However, because of the framework's modular nature, it may be expanded to classify an infinite number of facial expressions. (Yan et al., 2017) presented a novel and vigorous discriminative multi-metric learning approach for outward appearance arrangement in various video. (Sun et al., 2017) presented a multi-channel profound neural organization that learns and assembles the spatial-fleeting descriptors for outward appearances distinguishing proof in static edges. (Lopes et al., 2017) Presented a powerful calculation for outward appearance ID that utilizes a unification of Convolution Neural Network and some novel pre-preparing factors for a similar reason. (Chen et al. 2017) proposed a hearty technique to deal with the critical test of face movements by

considering a vigorous arrangement of highlights to be specific Histogram of Oriented Gradients from three opposite planes to gather highlights related with surfaces from video information.(Zhang et al.,2017) Is presented from a novel change of facial tourist spots. Finding the qualities of facial highlights-based emotions characterization procedures, individuals in the field focused on outward appearance-based emotion order strategies for handheld brilliant gadgets including mobiles.

(Abdullah ,et al., 2021) Researchers compared the applications of deep learning for emotional identification of multimodal signals based on current findings. Multimodal affective computing systems are compared to unimodal solutions since they have greater classification accuracy. The quantity of emotions observed, characteristics extracted, classification algorithm, and database consistency all affect accuracy. (Lee et al., 2017) proposed profound organizations for setting mindful emotion acknowledgment that consider both human outward appearance and setting information in a consolidated style. (Nhuonget al., 2019) proposes a calculation for include extraction with the end goal of face acknowledgment. (le et al.2019) presents succession parts for emotion acknowledgment. (Erfana et al., 2017) presents an overview about the emotion knowledge of various calculations in the field. Our proposed strategy presents minimized portrayal of spatial data (Verma et al., 2019) that viably encode emotion data. Coordinating the qualities of both face identification and emotion grouping into a bound together model.

(Jaiswal et al.,2020) suggested a deep learning architecture based on convolutional neural networks (CNN) for emotion recognition from photos. The suggested method's performance is assessed using two datasets: the Facial Expression Recognition Challenge (FERC-2013) and the Japanese female facial emotion dataset (JAFFE). The suggested model achieves 70.14 and 98.65 percent accuracy for the FERC-2013 and JAFFE datasets, respectively. (Mehendale et al.,) proposed a new facial emotion identification system based on convolutional neural networks (FERC). The FERC is built on a two-part convolutional neural network (CNN): first one removes the background from both the image, while the second focuses on the extraction of facial feature vectors. The expressional vector (EV) is utilised in the FERC model to identify the five different forms of regular facial expressions. Using an EV of length 24 data, it was feasible to appropriately emphasize the emotion with 96 percent accuracy.

Most of the previous work focused on recognizing prototype expressions of six basic emotions. Instead of classifying a facial image into one of the facial expression categories, Focus needs to be on the problem of estimating the depth of facial expression to further assess emotions. Another major drawback is the facial feature which are been automatically generated has been considered for analysis purpose which results in poor accuracy. As a solution for this problem, fusion of algorithms which extract more number of facial features along with CNN classifier features which results gives more accuracy. Also this approach helps in grading of emotions.

PROPOSED METHODOLOGY:

Lot of algorithms to detect facial emotions such as Haar Cascades which identifies the face & cuts parts of the face as the main regions of the eyes, mouth & nose.Proposed architecture as shown in figure 1 using CNN and different other algorithms by varying its depths to calculate the performance of each model for detecting the degree of facial expression in happiness and sadness using databases such as MMI research database videos, female Japanese facial expression database (JAFFE) & various real time expression videos are also used.Considering videos related to happiness and sadness ratings for classification purposes.Video pre-processing is performed using the LBP technique to determine the distance between the face points. In this research 19 geometric points are considered using the Euclidean distance between these points compared to the amount of facial change for different degrees of sadness and happiness.However, regions such as eyes, mouth, ears and nose clipped according to the viola-Jones algorithm, which system is also used Cascade Haar to get parts of face that are very important for detecting emotions.The results obtained after the classification phase are used in

confusion matrix for analysis of grading of emotions like Grade 1- pretending to be happy/sad emotion, Grade 2- normal happy/sad emotion and Grade 3- really happy/sad emotion as shown in figure 1.

De-Noising of Images

The process of recovering a signal from noisy photographs is known as image de-noising. De-noising is used to reduce undesirable noise from an image so that it may be analysed more effectively.

The additive noise in the input image p(y) with noise can be articulated by an equation:

$$P(y) = u(y) + \eta(y), y \in \Omega \tag{1}$$

From the above equation, u(y) is the original image which is not having noise. y is set of pixels, & η (y) is an additive noise item, which represents the impact of noise. Ω is a group of pixels, which is the whole image. Figure 2 shows the sample image of de-noising using Gaussian filter.

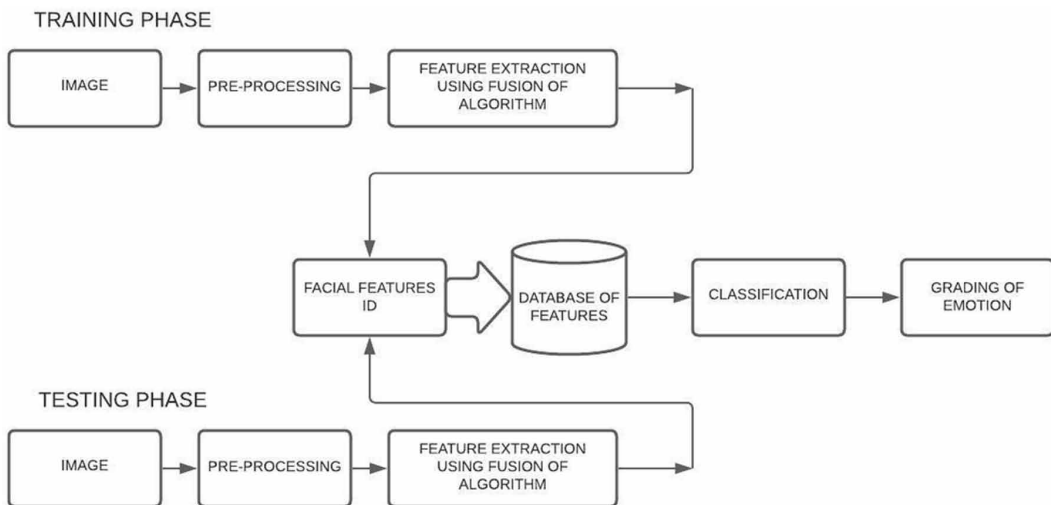Figure 1. Proposed Architecture for grading of emotion

Figure 2. De-noising the image using Gaussian filter

Fusion of Algorithms

Rebuilding the Traditional Local Binary Pattern (LBP) feature operator to image of the expression and combine the abstract characteristics of facial expression learned from the neural network of deep convolution with the modified features of the texture of the LBP facial expression in the full connection layer extracting 19 facial features.

Viola Jones Algorithm

Viola-Jones is intended for front face, so it is capable to detect the best frontally instead of faces facing sideways, up or down. The Viola-Jones algorithm first detects the face in the gray scale image and then finds its position in the color image. Viola-Jones outlines a box (as you can see right) and looks for a face inside the box:

$$f(x,y)= \sum\{f(x1,y1) + f(x2,y2)\} \tag{2}$$

From the above equation f(x,y)- is an input image where foreground and background are classified as f(x1,y1) and f(x2,y2).here f(x1,y1) is considered as face part in foreground and rest of portion of image is considered as background as shown in figure 3.

Extracting the Geometric Face Features Using HAAR Cascade Technique

It is an object detection algorithm used to identify faces in a real-time image. Principal entities are selected using the Adaboost algorithm. Each subtant is then checked for the presence or absence of a face using a cascade of classifiers. The discovery algorithm uses a cascade of classifiers that use Haar-like entities. Therefore, it is also called a detector based on Haar Cascades:

$$f(x1,y1) = \sum\{g(x11,y11)+........+g(x19,y19)\} \tag{3}$$

From the above equation f(x1,y1) is an input face image using HAAR cascade technique 19 face features like eyes ,nose ,etc. have been clustered as g(x11,y11)+........+g(x19,y19). Figure 4 shows the extracted face features using fusion of algorithms.

Feature Points Distance Calculation Using LBP Algorithm and Euclidean Distance

The Local Binary Pattern algorithm is simple but very effective texture operator that labels the pixels of an image at the neighborhood threshold of each pixel and treats the result as a binary number. The Local Binary Pattern algorithm is a simple but very efficient texture operator that labels the pixels of an image at the neighborhood threshold of each pixel and treats the result as a binary number. Redundant plot neighborhood points extracted from LBP coordinates will be taken and calculated for those distances as shown in Figure 5:

Figure 3. Face Location Identification

Input image

Output image
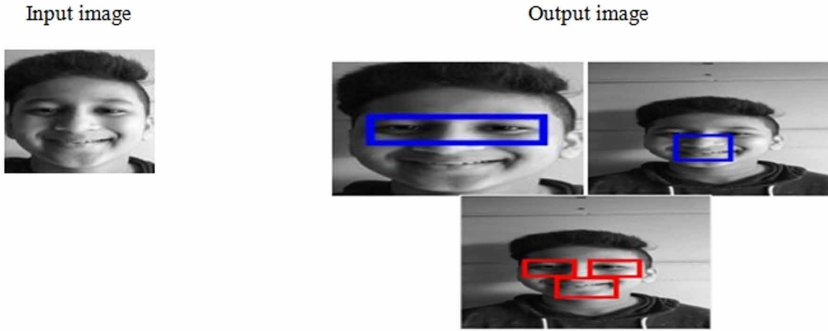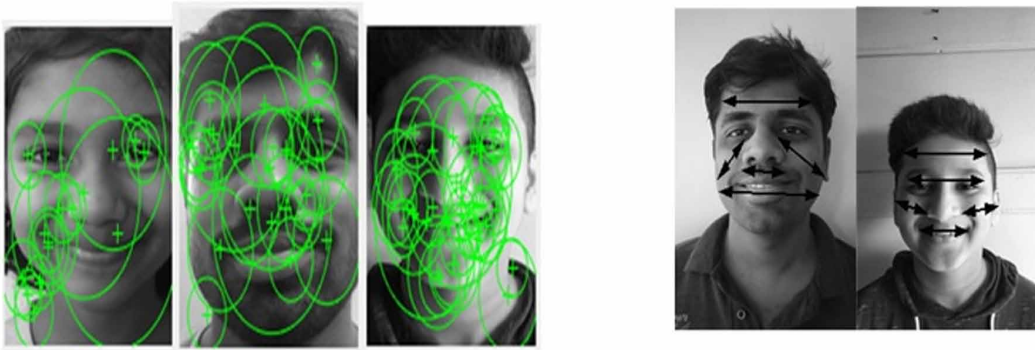
Figure 4. Fusion of Algorithms to Extract Face Features

Figure 5. Feature Points Distance Calculation

$$D= \sum \{g(x11, y11) + \ldots\ldots + g(x19, y19)\} = \sqrt{\sum_{i=1}^{k} \left(xi - yi\right)^2} \tag{4}$$

Sigma of $g(x11,y11)+\ldots\ldots+g(x19,y19)$ is the given 19 features of face like eyes, nose, etc. which is the subjected to calculate the distance of each features using LBP and Euclidean distance $\sqrt{\sum_{i=1}^{k} \left(xi - yi\right)^2}$ as shown in equation (4).

This work presents a novel methodology by fusion of algorithms like LBP and Euclidian's distance for feature extraction of human emotion to get a maximum of 19 features of human face as shown in table 1. Dependent on Convolution Neural Networks (CNN) for facial demeanor acknowledgment using 19 feature points from the face image to classify the grades of happiness and sadness. The dimensions of the image in our study are 128×128×1. The contribution to our framework is a picture; at that point, CNN foresee the facial demeanor mark which ought to the one of these names like degrees of happiness and sadness.

The proposed feature points were chosen through a thorough literature review of various feature points that contribute to happy and sad emotions. Emotions are graded by comparing the number of features that match the threshold value given. The threshold value is determined by comparing the number of extracted features to the number of features that have been matched. Using a pleasant emotion as an example, we have divided happy emotions into three categories: grade 1 happy, grade

Table 1. Landmark points identified on face

Facial Landmark Number	Face Landmark	Class
FL1	Left Eye	LY
FL2	Right Eye	RY
FL3	Upper Nose	UN
FL4	Lower Nose	LN
FL5	Upper Lip	UL
FL6	Nasion	NS
FL7	Lower Lip	LL
FL8	Right Cheek	RC
FL9	Left Cheek	LC
FL10	Right Eyebrow	RB
FL11	Left Eyebrow	LB
FL12	Forehead	F
FL13	Chin	C
FL14	Right Ear	RE
FL15	Left Ear	LE
FL16	Right Jowl	RJ
FL17	Left Jowl	LJ
FL18	Right Temple	RT
FL19	Left Temple	LT

2 happy, and grade 3 happy emotions. Grade 1 implies that the person is feigning happiness, and the qualities matched for this emotion among 19 features are approximately 6 to 8. In other words, if 6 to 8 features coincide, we consider it a grade 1. In the same manner, grade 2 feeling is nothing more than a person who is normally pleased. The threshold value has been set at 12 to 14 features. If more than 12 characteristics fit, we consider the person to be normal and happy. Similarly, more than 15 elements should match for the grade 3 feeling, which indicates that the person is extremely delighted. This allows us to see the actual feelings on someone's face. The same procedure is used to deal with negative emotions.

Classification and Grading of Human Emotions

KNN

In this algorithm, training samples will be vectors in multidimensional feature space, each with class tag. The training phase of the algorithm consists only in storing feature vectors and labels of training sample classes. In the classification phase, k is a user-defined constant, and an unlabeled vector is classified by assigning the most common tag among the training k examples closest to that query point. A distance metric commonly used for continuous variables is the Euclidean distance. The nearest k-neighbor classifier will be seen as allocating the nearest k-neighbors a display style 1/k weight and all other 0 weights. This can be generalized to the closest weighted neighboring classifiers. Previous work on grading of emotion by fusing LBP and KNN algorithm (kulkarni et al., 2021) and the results were not promising with accuracy of 79% in grading of human emotions.

Decision Tree

In this algorithm a regulated learning strategy that predicts a worth given a bunch of contributions by "learning" rules dependent on a bunch of preparing information. To lay it out plainly, it is a huge tree of if-then-else rules(Zhang et al., 2017). The decision-making measure gets going at the foundation of the tree and drops by noting a progression of yes-no inquiries. Toward the finish of this if-then-else chain, it shows up at a solitary anticipated mark. This is the yield of a decision tree.

Fuzzy Logic

Fuzzy Logic is close to constant scientific issues and control issues. So in the framework that had utilized a Mamdani-type Fuzzy Rule Based System. It comprises of two segments (Özerdem et al., 2017). Database of fuzzy framework contains scaling factors for info and yield. It additionally contains the membership capacities that determine the importance of etymological terms. It has all overseeing decides that lead to effective articulation acknowledgment. Fuzzy membership function is as shown in below equation (5) $\mu_{tr}(x)$:

$$\mu_{tr}(x) = \begin{cases} \dfrac{1}{(r-p)}(x-r)+1 \, p < x \le r \\ \\ -\dfrac{1}{(q-r)}(x-r)+1 \, r < x \le q \end{cases}$$

(5)

For perceiving the appearance of any static picture, requiring just eight fundamental facial activity components like temple, eyebrow, eye, nose, lips, teeth, cheek, and jawline(Meng et al., 2016). These data sources are gone from 0 to1 values. These facial activity components are planned to their individual fuzzy sets by input membership work (IMF).

The FFCIS rules are listed according to the knowledge of the psychologist. The triangular fuzzy function is used as a membership function. Figure 6 provides an example of the FFCIS input membership functions. A set of rules for each FFCIS is stored in the engine based on blurry knowledge. Ten FFCIS outputs are forwarded to the fuzzy emotional inference system.

FCM used low-quality image functions; the facial spots are coordinated. The low-quality function is understandable only by machine, while the high-quality function is close to human intelligence. Figure 7 shows the diffuse function for the mouth part. This implies that FEIS performed human
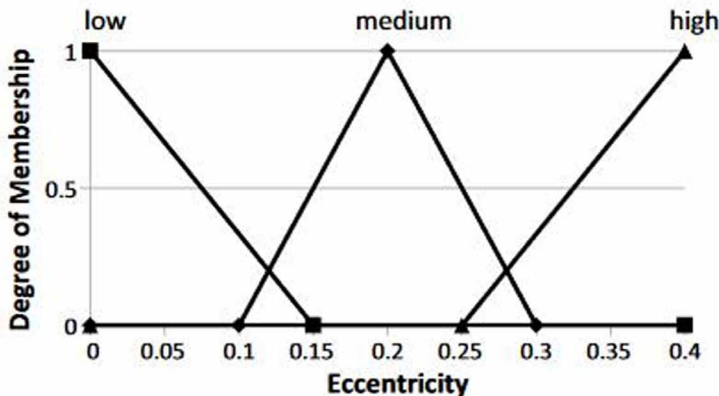
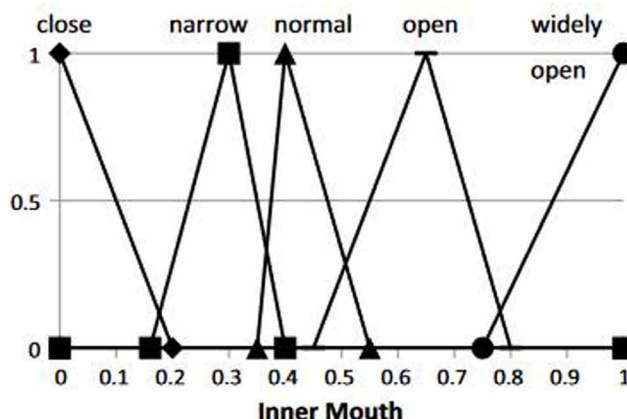**Figure 6. Fuzzy triangular function**

Figure 7. Fuzzy function for Mouth part

natural recognition through a set of rules on facial components and provided the good accuracy of the recognition results.

CNN Alex Net

Proposed methodology using CNN by varying its depths to calculate performance of each model for facial expression grade detection in happiness and sadness using databases like research MMI database videos(Meng et al., 2016), Japanese Female Facial Expression (JAFFE) Database(Feng et al.,2017) and some real time expression images are used as shown in figure 8. Considering videos related to grades of happiness and sadness for classification purpose. The preprocessing of the video is done by using LBP technique to determine the distance between the face points. In this research 19 geometrical points are considered using Euclidian distance between these points with respect to amount of change in face for different grades of sadness and happiness.

The most famous method for classification of images is Convolutional neural network (CNN). The concept of multi-layer perceptron (MLP) and CNN both are not same because in CNN had hidden layers, which are called as convolutional layers(Yu et al., 2015)(Zhang et al., 2016).In this research work proposed grade two neural network structures. The principal grade suggested is foundation expulsion (Feng et al., 2017), used to remove feelings from a picture, as demonstrated.

Here traditional CNN network module is utilized to separate essential expressional vector. The expressional vector (EV) created by finding significant facial purposes of significance. EV stays straight forwardly identified with changes in articulation (Mehmood et al., 2017). The EV is gotten utilizing an essential perceptron unit applied on foundation eliminated face picture. In this proposed

Figure 8. Real time dataset for grading of happy and sad emotions

Figure 9. Proposed architecture using CCN
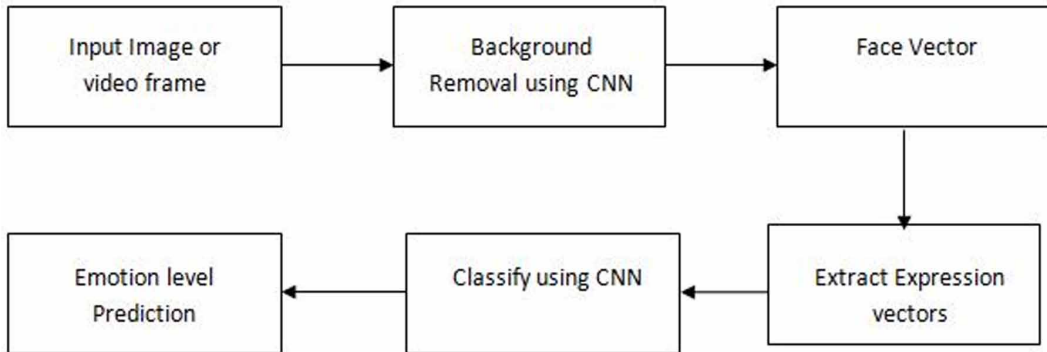
FERC model, additionally having a non-convolutional perceptron layer as the last stage. Every one of the convolutional layers gets the information (or picture), changes it, and afterward yields it to the following grade. The change in convolution activity, as demonstrated (Batbaatar et al., 2019). All the convolutional layers utilized are fit for design identification. Inside each convolutional layer, four channels were utilized. The information picture took care of to the initial segment CNN (utilized for foundation expulsion) by and large comprises of shapes, edges, surfaces, and items alongside the face.

The edge detector, circle detector, and corner detector channels are utilized toward beginning of convolutional layer 1. When face has been identified, the second part CNN channel gets facial highlights, like eyes, ears, lips, nose, and cheeks. The edge recognition channels utilized in this layer are appeared as shown in figure 10.

Equation 6 explains the output after the convolution result:

$$(M \times M) * (F \times F) = (M-F+1) \times (M-F+1) \quad 6)$$

where:

M $\times$ M is an input image
F $\times$ F is a filter size

The second-part CNN comprises of layers with 3×3 portion matrix, e.g., (0.25, 0.17, 0.9; 0.89, 0.36, 0.63; 0.7, 0.24, 0.82). These numbers are chosen somewhere in the range of 0 and 1 at first. These numbers are streamlined for EV location, in view of the ground truth that had, in the administrative preparing dataset (Zhang et al,. 2017). Here, utilizing smaller than normal mum mistake disentangling

Figure 10. Face Emotion Detection using CNN Architecture
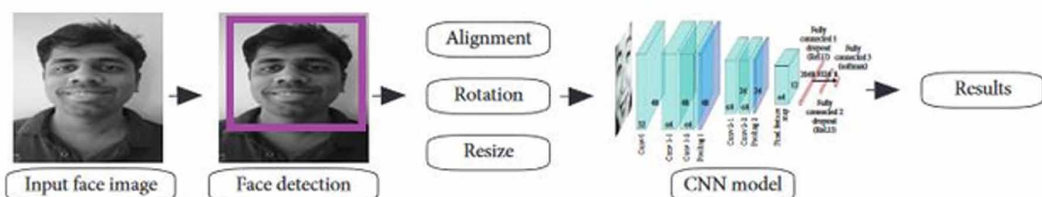
to advance channel esteems. The stride is denoted as S. The value of S indicates every move of a window after every operation. Figure 11 shows the way of striding. By default, the string value is 1.

The padding value P=0 which, adds P zeroes in the boundaries of the input image. Using valid mode of padding. The equation 7 explains the padding of an image with 0's:

$$(M+2P-F+1)x(M+2P-F+1) \tag{7}$$

where:

- M is the size of an input
- F is the filter size
- P is the padding 0's

When the channel is tuned by administrative learning, it is then applied to the foundation eliminated face (i.e., on the yield picture of the initial segment CNN), for recognition of various facial parts (e.g., eye, lips. nose, ears, and so forth) to create the EV matrix, taking all things together 24 different facial highlights are extricated. The EV highlight vector is only estimations of standardized Euclidian distance between each face part.

Choosing Key Frame Extraction

FERC works with a picture just as video input. In the event that, when the contribution to the FERC is video, at that point the distinction between particular edges is processed with the preprocessing technique LBP (Vella et al., 2017). The maximally steady edges happen at whatever points the intra-outline distinction are zero. At that point for these steady casings, a canny edge locator was applied, and afterward the totaled number of white pixels was determined. Subsequent to looking at the amassed wholes for every single stable casing, the edge with the maximum collected entirety is chosen since this casing has greatest subtleties according to edges (more edges more subtleties). This edge is then chosen as a contribution to FERC. The rationale behind picking this picture is that foggy pictures have least edges or no edges.

Figure 11. Filtering Process
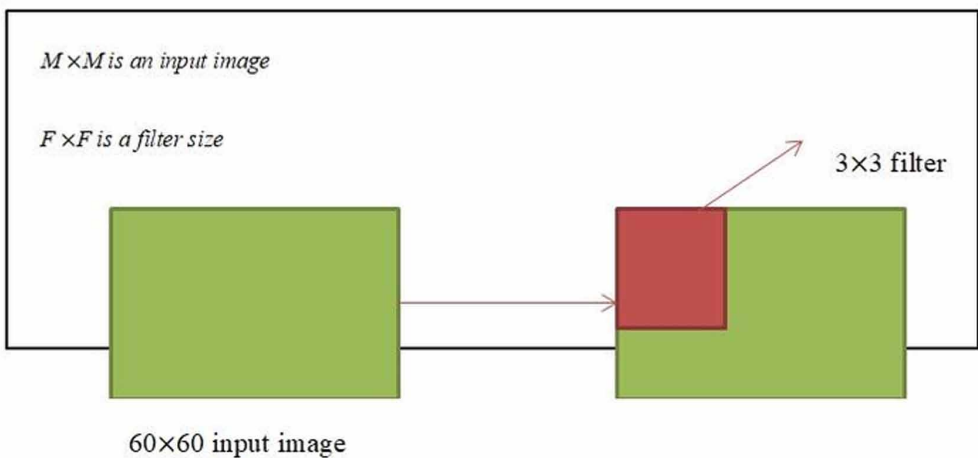
Edge Enhancement of Key Frame

After obtaining the input image, skin tone discovery calculation (Savchenko et al., 2017) is applied to separate human body parts from the picture. This skin tone-identified yield picture is a paired picture and utilized as the component, for the primary layer of foundation evacuation CNN (additionally alluded to as the initial segment CNN in this composition).

This skin tone recognition relies upon the sort of info picture. In the event that the picture is the shaded picture, image is converted into gray for feature extraction, Edge enhanced image is given as CNN input as shown in figure 13. Here 19 feature points are considered to detect grades of happiness and sadness from face images (Kamarol et al, 2017). To remove the background noise from the input image in CNN as shown in figure 12 we can use Hough transform method, equation is as follows:

$$H\left(\theta,\rho\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A\left(x,y\right)\delta\left(\rho - x\cos\theta - y\sin\theta\right)dx\ dy \tag{8}$$

Convolution Filter

As demonstrated in figure 14 for every convolution activity, the whole picture is partitioned into covering 3×3 lattices, and afterward the relating 3×3 channel is convolved over each 3×3 grid acquired from the picture (Ghimire et al., 2017). The sliding and taking spot item activity is called 'convolution' and thus the name 'convolutional channel.' Grade 1 happiness (acting to be happy) is shown in figure 15.

We used Leaky Rectifier Linear Unit (Leaky ReLU)(Massey et al., 2017) as follows:

$$f\left(x\right) = \max\left(x, \frac{x}{20}\right) \tag{9}$$

Figure 12. Background Removal

Figure 13. Edge enhanced image given as CNN input

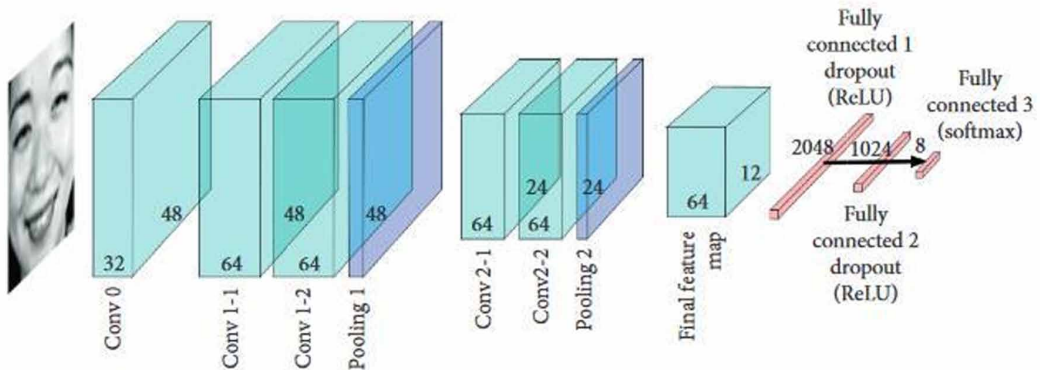Figure 14. Convolution Filtering for deep learning

Figure 15. Grade 1 Happiness (Acting to be Happy)

where the threshold value 20 is selected using the FER-2013 validation set. Leaky ReLU is chosen over ordinary ReLU to solve the dying ReLU problem. Instead of giving zero when x < 0, leaky ReLU will provide a small negative slope.

During the convolution, dab result of both 3×3 framework is processed and put away at a comparing area, e.g., (1,1) at the yield, as demonstrated. When the whole yield network is determined, at that point this yield is passed to the following layer of CNN for another round of convolution (Yang et al., 2017)) (Escalera et al, 2018). The last layer of face include separating CNN is a straightforward perceptron, which attempts to improve estimations of scale factor and example relying on deviation from the ground truth (Lee et al, 2014)(Cai, et al, 2018).

Comparison Analysis on Classification and Grading of Human Emotions

A suitable tree was chosen from the start due to a combination of its ability to prepare information with little effort and its reputation for performing wonderfully in virtually all situations. The first task was to use Sci-Kit Learn to create a decision tree. The bounds of a conventional choice tree classifier were fine-tuned. The emphmin testing portion boundary was the lone border that showed a recognisable distinction while adjusting. The general precision was barely 0.309 with the boundary set to 40. Because cutting-edge models have accuracies greater than 0.6, it was decided to examine fluffy logic as a whole.

If a single outward appearance was to develop, fluffy feeling was also prepared to bring about a single feeling force. In the second scenario, the presentation of FEIS was examined using two distinct classifiers that performed essential feeling grouping. The suggested framework was tested under the same conditions as the FEC and CNN classifiers (Takalkar et al, 2018) using a dataset of 304 images and the precision rate of the proposed framework was compared to that of other classifiers. FEIS 0.90, while FEC got the least normal precision rate 0.81. Despite the fact that FEIS got a somewhat lower precision rate than CNN, it is imperative to feature the way that FEIS can play out numerous acknowledgments.

In this segment, two arrangements of investigations are intended to confirm the presentation of the proposed strategy. The primary gathering of examinations breaks down the presentation of the calculation and checks that the preparation season of the calculation is lower than that of the customary CNN calculation model(Ma et al., 2019)(Tang et al., 2013)(Wang et al.,2012). The test information comes from the Fer-2013 articulation data set. The second gathering of tests is utilized to check that the acknowledgment pace of the calculation has expanded under complex foundation(Whitehill et al., 2006). The trial information comes from the MMI facial articulation data set and LF blended arrangements of W datasets (Savchenko et al., 2017), (Massey et al., 2017). The Fer-2013 outward appearance data set contains 28,709 preparing pictures and 7,178 test pictures, every one of which is a 48×48 dim scale picture. Each face is pretty much in the center of the image. Figure 16 shows the percentage of detected emotions expressed by 50 different people. In this way, in the analysis, the picture information can be straightforwardly contribution to the organization for preparing with no other pre-handling.

As can be seen from Figure 14, both the proposed calculation and the CNN calculation unite after a specific number of emphases. At the point when the model combines, the acknowledgment rates are 85.54% and 77.78%, separately. It tends to be seen that the proposed calculation improves by almost 8 rate focuses contrasted with the CNN calculation (Sun et al, 2016). Subsequently, the proposed calculation has certain points of interest in outward appearance acknowledgment rate.

Figure 16. Percentage of detected emotions expressed by 50 different people
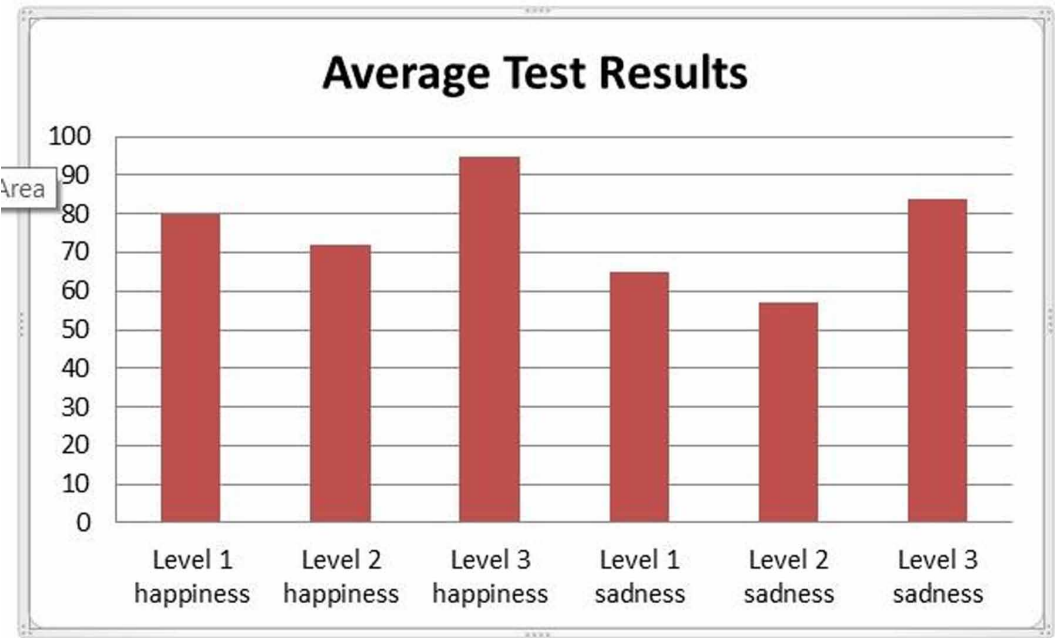
The two models combine after a specific number of epochs. At the point when the proposed model is iterated to multiple times, the model starts to join. The CNN calculation combines after 15,000 cycles (Tang et al, 2013)(Xu M. et al., 2015). This shows that the proposed calculation can accomplish good outcomes after less cycles, in other words, the preparation speed of the proposed strategy on the preparation set is 1.5 occasions quicker than that of CNN calculation.

The results with different datasets are given in Table 2.confusion matrix for decision tree, Fuzzy Based Classification, Alex net CNN is shown in table 3, table 4, and table 5. Grade 1- pretending to be happy/sad emotion, Grade 2- normal happy/sad emotion and Grade 3- really happy/sad emotion.

Algorithm Comparison for Grading of Emotion Face Detection with respect to different classification algorithms is given in table 6. For training and testing phases we have used 15000 images which are extracted from real time 3-5 second videos and MMI video data base. As part of this work, fusion of algorithms has been done to gain more accuracy in grading of human emotions. For emotion feature extraction, algorithms like LBP and Euclidian's distance have been fused to get a maximum of 19 features of human face. Deep learning CNN Alex Net gained an accuracy of 87.58%, Sensitivity 99.46% Specificity 89.80% which is maximum compared to the results obtained using KNN ,Decision Tree, Fuzzy logic Algorithms. Accuracy of deep learning over other algorithms is shown in figure 18. Detection analysis of Deep learning trained on three datasets is shown in figure 17.

The results obtained after the classification phase are used in confusion matrix for analysis of grading of emotions like Grade 1- pretending to be happy/sad emotion, Grade 2- normal happy/sad emotion and Grade 3- really happy/sad emotion. The proposed strategy is contrasted and CNN and FRR-CNN calculations (Xie et al., 2017). The test information comes from Fer-2013 outward appearance informational collection. Table 5 records their cognizance rate examination of the three calculations, and the time correlation on the test set and the preparation set. The preparation time and test season of preparing set demonstrated in Table 5 allude to the time used to deal with a cluster of

Table 2. Results with different datasets

Dataset	Image training data	Images tested	MSE
MMI database videos	1200	350	0.724
JAFFE	1700	570	0.5873
Real time	2000	640	0.3251

Table 3. Confusion Matrix for Decision Tree

	Happiness (grade1)	Happiness (grade2)	Happiness (grade3)	Sadness (grade1)	Sadness (grade2)	Sadness (grade3)
Happiness (grade1)	72.33%	20.21%	7.46%	0	0	0
Happiness (grade2)	18.45%	65.79%	15.76%	0	0	0
Happiness (grade3)	5.62%	22.87%	71.51%	0	0	0
Sadness (grade1)	0	0	0	62.57%	28.45%	8.98%
Sadness (grade1)	0	0	0	16.85%	70.25%	12.9%
Sadness (grade3)	0	0	0	8.12%	22.73%	69.15%

Table 4. Confusion Matrix for Fuzzy Based Classification

	Happiness (grade1)	Happiness (grade2)	Happiness (grade3)	Sadness (grade1)	Sadness (grade2)	Sadness (grade3)
Happiness (grade1)	77.25%	18.12%	4.63%	0	0	08
Happiness (grade2)	11.27%	70.64%	18.09%	0	0	0
Happiness (grade3)	2.89%	21.22%	75.51%	0	0	0
Sadness (grade1)	0	0	0	68.73%	27.55%	3.72%
Sadness (grade1)	0	0	0	8.25%	74.18%	17.57%
Sadness (grade3)	0	0	0	1.76%	20.71%	77.53%

Table 5. Confusion Matrix for Alex net CNN

	Happiness (grade1)	Happiness (grade2)	Happiness (grade3)	Sadness (grade1)	Sadness (grade2)	Sadness (grade3)
Happiness (grade1)	79.27%	15.45%	5.28%	0	0	0
Happiness (grade2)	4.03%	73.54%	22.43%	0	0	0
Happiness (grade3)	4.69%	12.84%	82.47%	0	0	0
Sadness (grade1)	0	0	0	71.48%	19.12%	9.4%
Sadness (grade1)	0	0	0	15.39%	69.46%	15.15%
Sadness (grade3)	0	0	0	3.41%	18.42%	78.17%

pictures. In this paper, three dataset pictures are handled in groups. Below graph figure 19 represents the model training accuracies and loss using the datasets.

CONCLUSION

In this paper, LBP and a convolutional neural network are used to combine multiple algorithms for feature extraction in order to score human emotion. Previously, features created automatically using CNN were employed for categorization, but the results were less accurate. To improve the accuracy of our suggested unique work, we used 19 features retrieved using a combination of methods such as LBP and Euclidian distance, as well as the CNN Alex net methodology. Furthermore, the proposed computation achieves the highest normal recognition rate of 87.58 percent in distinguishing pleasure

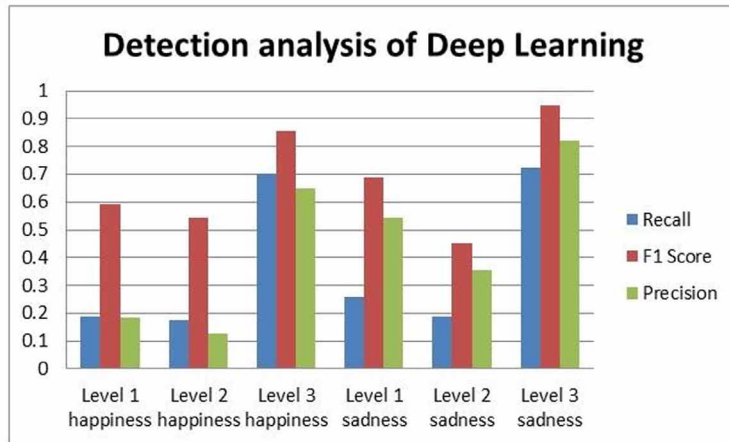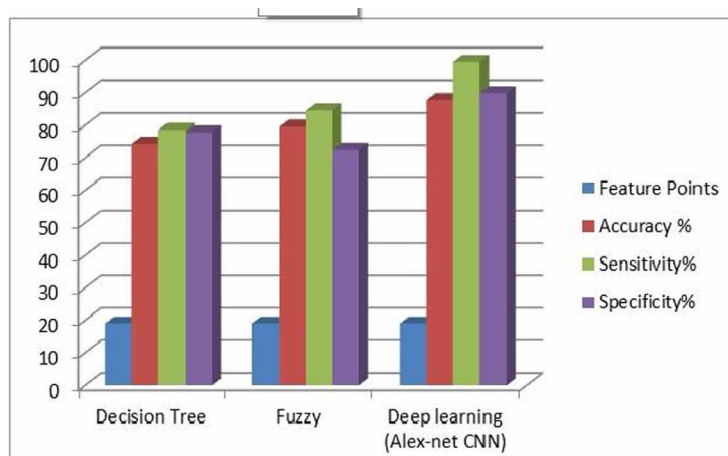**Figure 17. Detection analysis of Deep learning trained on three datasets**

**Table 6. Algorithm Comparison for Grades in Emotion Face Detection**

Algorithm	Feature Points	Accuracy %	Sensitivity%	Specificity%
Decision Tree	19	74.26	78.57	77.77
Fuzzy	19	79.57	84.69	72.40
Deep learning (Alex-net CNN)	19	87.58	99.46	89.80

Figure 18. Accuracy of deep learning over other algorithms

and melancholy grades. The threshold for grading human emotion into grade 1, 2, 3 is considered with respect to the number of features matched considering the 19 features. When all of the models are integrated with the aggregator, the aggregator exceeds state-of-the-art approaches in terms of emotion grades. By achieving outstanding results on MMI database videos, JAFFE database videos, and some real-time expression video datasets, our experiments demonstrate a clear advantage of

Figure 19. Training loss and accuracy using Alex-net
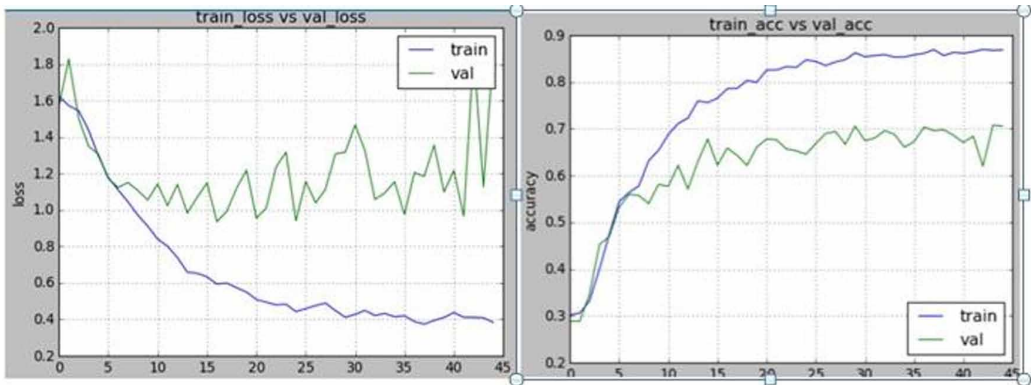
aggregating deep learning using Alex-net and CNN models in achieving better accuracy for grades of emotion such as Grade 1- pretending to be happy/sad emotion, Grade 2- normal happy/sad emotion, and Grade 3- really happy/sad emotion detect. The presented method could be applied in a decision support system for neuropsychiatric diseases assessment in the future. Many of these patients have trouble interpreting emotions through facial expressions for various reasons. This method will aid in the interpretation of facial expressions in order to provide better treatment options.

REFERENCES

Abdullah, S. M. S. A., Ameen, S. Y. A., Sadeeq, M. A., & Zeebaree, S. (2021). Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(02), 52–58.

Alreshidi, A., & Ullah, M. (2020). Facial emotion recognition using hybrid features. In Informatics (Vol. 7, No. 1, p. 6). Multidisciplinary Digital Publishing Institute.

Batbaatar, E., Li, M., & Ryu, K. H. (2019). Semantic-emotion neural network for emotion recognition from text. *IEEE Access : Practical Innovations, Open Solutions*, 7, 111866–111878. doi:10.1109/ACCESS.2019.2934529

Cai, J., Chang, O., Tang, X. L., Xue, C., & Wei, C. (2018, July). Facial expression recognition method based on sparse batch normalization CNN. In *2018 37th Chinese Control Conference (CCC)* (pp. 9608-9613). IEEE.

Escalera, S., Baró, X., Guyon, I., Escalante, H. J., Tzimiropoulos, G., Valstar, M., Pantic, M., Cohn, J., & Kanade, T. (2018). Guest editorial: The computational face. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), 2541–2545. doi:10.1109/TPAMI.2018.2869610

Ghimire, D., Jeong, S., Lee, J., & Park, S. H. (2017). Facial expression recognition based on local region specific features and support vector machines. *Multimedia Tools and Applications*, 76(6), 7803–7821. doi:10.1007/s11042-016-3418-y

Jaiswal, A., Raju, A. K., & Deb, S. (2020). Facial Emotion Detection Using Deep Learning. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.

Kamarol, S. K. A., Jaward, M. H., Kälviäinen, H., Parkkinen, J., & Parthiban, R. (2017). Joint facial expression recognition and intensity estimation based on weighted votes of image sequences. *Pattern Recognition Letters*, 92, 25–32. doi:10.1016/j.patrec.2017.04.003

Kulkarni, P., & Rajesh, T. M. (2020). Analysis on techniques used to recognize and identifying the Human emotions. *Iranian Journal of Electrical and Computer Engineering*, 10(3), 3307. doi:10.11591/ijece.v10i3.pp3307-3314

Kulkarni, P., & Rajesh, T. M. (2021). Video Based Sub-Categorized Facial Emotion Detection Using LBP and Edge Computing. *Rev. d'Intelligence Artif.*, 35(1), 55-61.

Kumar, N., & Bhargava, D. (2017). A scheme of features fusion for facial expression analysis: A facial action recognition. *Journal of Statistics and Management Systems*, 20(4), 693–701. doi:10.1080/09720510.2017.1395189

Lee, S. H., Plataniotis, K. N., & Ro, Y. M. (2014). Intra-class variation reduction using training expression images for sparse representation based facial expression recognition. *IEEE Transactions on Affective Computing*, 5(3), 340–351. doi:10.1109/TAFFC.2014.2346515

Li, J., Zhang, D., Zhang, J., Zhang, J., Li, T., Xia, Y., Yan, Q., & Xun, L. (2017). Facial expression recognition with faster R-CNN. *Procedia Computer Science*, 107, 135–140. doi:10.1016/j.procs.2017.03.069

Li, S., & Deng, W. (2018). Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, 28(1), 356–370. doi:10.1109/TIP.2018.2868382 PMID:30183631

Li, W., Tsangouri, C., Abtahi, F., & Zhu, Z. (2018). A recursive framework for expression recognition: From web images to deep models to game dataset. *Machine Vision and Applications*, 29(3), 489–502.

Ma, H., & Celik, T. (2019). FER-Net: Facial expression recognition using densely connected convolutional network. *Electronics Letters*, 55(4), 184–186.

Magudeeswaran, V., & Singh, J. F. (2017). Contrast limited fuzzy adaptive histogram equalization for enhancement of brain images. *International Journal of Imaging Systems and Technology*, 27(1), 98–103. doi:10.1002/ima.22214

Mao, Q., Rao, Q., Yu, Y., & Dong, M. (2016). Hierarchical Bayesian theme models for multipose facial expression recognition. *IEEE Transactions on Multimedia*, 19(4), 861–873. doi:10.1109/TMM.2016.2629282

Massey, R., Williams, L., Smit, R., Swinbank, M., Kitching, T. D., Harvey, D., Jauzac, M., Israel, H., Clowe, D., Edge, A., Hilton, M., Jullo, E., Leonard, A., Liesenborgs, J., Merten, J., Mohammed, I., Nagai, D., Richard, J., Robertson, A., & Tittley, E. et al. (2015). The behaviour of dark matter associated with four bright cluster galaxies in the 10 kpc core of Abell 3827. *Monthly Notices of the Royal Astronomical Society*, 449(4), 3393–3406. doi:10.1093/mnras/stv467

Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), 1–8.

Mehmood, R. M., Du, R., & Lee, H. J. (2017). Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors. *IEEE Access : Practical Innovations, Open Solutions*, *5*, 14797–14806. doi:10.1109/ACCESS.2017.2724555

Meng, H., Bianchi-Berthouze, N., Deng, Y., Cheng, J., & Cosmas, J. P. (2015). Time-delay neural network for continuous emotional dimension prediction from facial expression sequences. *IEEE Transactions on Cybernetics*, *46*(4), 916–929. doi:10.1109/TCYB.2015.2418092 PMID:25910269

Özerdem, M. S., & Polat, H. (2017). Emotion recognition based on EEG features in movie clips with channel selection. *Brain Informatics*, *4*(4), 241–252. doi:10.1007/s40708-017-0069-3 PMID:28711988

Savchenko, A. V. (2017). Deep neural networks and maximum likelihood search for approximate nearest neighbor in video-based image recognition. *Optical Memory and Neural Networks (Information Optics)*, *26*(2), 129–136. doi:10.3103/S1060992X17020102

Sun, B., Li, L., Zhou, G., & He, J. (2016). Facial expression recognition in the wild based on multimodal texture features. *Journal of Electronic Imaging*, *25*(6), 061407.

Takalkar, M., Xu, M., Wu, Q., & Chaczko, Z. (2018). A survey: Facial micro-expression recognition. *Multimedia Tools and Applications*, *77*(15), 19301–19325. doi:10.1007/s11042-017-5317-2

Takalkar, M., Xu, M., Wu, Q., & Chaczko, Z. (2018). A survey: Facial micro-expression recognition. *Multimedia Tools and Applications*, *77*(15), 19301–19325.

Tang, Y. (2013). *Deep learning using linear support vector machines.* arXiv preprint arXiv:1306.0239.

Tzimiropoulos, G., & Pantic, M. (2017). Fast algorithms for fitting active appearance models to unconstrained images. *International Journal of Computer Vision*, *122*(1), 17–33. doi:10.1007/s11263-016-0950-1 PMID:32269419

Wang, Z., & Ying, Z. (2012, May). Facial expression recognition based on local phase quantization and sparse representation. In *2012 8th International Conference on Natural Computation* (pp. 222-225). IEEE.

Whitehill, J., & Omlin, C. W. (2006). Haar features for FACS AU recognition. In 7th international conference on automatic face and gesture recognition (FGR06). IEEE.

Xie, S., & Hu, H. (2017). Facial expression recognition with FRR-CNN. *Electronics Letters*, *53*(4), 235–237. doi:10.1049/el.2016.4328

Xu, M., Cheng, W., Zhao, Q., Ma, L., & Xu, F. (2015). Facial expression recognition based on transfer learning from deep convolutional networks. In *2015 11th International Conference on Natural Computation (ICNC)* (pp. 702-708). IEEE.

Yang, B., Xiang, X., Xu, D., Wang, X., & Yang, X. (2017). 3D palmprint recognition using shape index representation and fragile bits. *Multimedia Tools and Applications*, *76*(14), 15357–15375. doi:10.1007/s11042-016-3832-1

Yu, Z., & Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 435-442). ACM.

Zhang, F., Mao, Q., Shen, X., Zhan, Y., & Dong, M. (2018). Spatially coherent feature learning for pose-invariant facial expression recognition. *ACM Transactions on Multimedia Computing Communications and Applications*, *14*(1s), 1–19. doi:10.1145/3241059

Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., & Yan, K. (2016). A deep neural network-driven feature learning method for multi-view facial expression recognition. *IEEE Transactions on Multimedia*, *18*(12), 2528–2536.

Zhang, Y., Li, Y., Xie, B., Li, X., & Zhu, J. (2017). Pupil localization algorithm combining convex area voting and model constraint. *Pattern Recognition and Image Analysis*, *27*(4), 846–854. doi:10.1134/S1054661817040216

Praveen Kulkarni is Asst. Professor in Christ School of Engineering and Technology, Bangalore. He holds a master's degree in Computer Science. he is currently pursuing his PhD at Dayananda Sagar University Bangalore. He has 8 years of teaching experience. His area of interest in research includes Digital Image Processing, Computer Vision, Data Analytics.

Rajesh T. M. is an Assistant Professor in the Department of CS&E, SoE, Dayananda Sagar University Bangalore. He has completed his Ph. D from Jain University, Bengaluru. He has a total of 8+ years of experience in industry, research, and academics domains. His area of interest in research includes Digital Image Processing, Pattern Recognition, Video Analytics, Computer Vision Data Analytics and IoT.