Indonesian Journal of Electrical Engineering and Computer Science
Vol. 25, No. 3, March 2022, pp. 1406~1419
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v25.i3.pp1406-1419

Multi-feature based automatic facial expression recognition using deep convolutional neural network

Anjali Dixit, Tanmay Kasbe
Department of Computer Science, Oriental University, Indore, India

Article Info

Article history:

Received Oct 11, 2021
Revised Jan 12, 2022
Accepted Jan 20, 2022

Keywords:

Convolutional neural networks
Deep multi-task learning
Face detection
Facial feature vectors
Multi-feature learning
Recognition accuracy

ABSTRACT

Deep multi-task learning is one of the most challenging research topics widely explored in the field of recognition of facial expression. Most deep learning models rely on the class labels details by eliminating the local information of the sample data which deteriorates the performance of the recognition system. This paper proposes multi-feature-based deep convolutional neural networks (D-CNN) that identify the facial expression of the human face. To enhance the accuracy of recognition systems, the multi-feature learning model is employed in this study. The input images are preprocessed and enhanced via three filtering methods i.e., Gaussian, Wiener, and adaptive mean filtering. The preprocessed image is then segmented using a face detection algorithm. The detected face is further applied with local binary pattern (LBP) that extracts the facial points of each facial expression. These are then fed into the D-CNN that effectively recognizes the facial expression using the features of facial points. The proposed D-CNN is implemented, and the results are compared to the existing support vector machine (SVM). The analysis of deep features helps to extract the local information from the data without incurring a higher computational effort.

Corresponding Author:

Tanmay Kasbe
Department of Computer Science, Oriental University
175-B Veena Nagar, Near MR 10, Indore, Madhya Pradesh, India
Email: tanmay.kasbe@gmail.com

1. INTRODUCTION

Facial expression is one of the significant manners wherein humans display emotions. Humans' expressions and intentions are most effectively communicated through facial expression. Different circumstances of humans pose different ways of displaying the emotions such as anger, fear, surprise, disgust happiness, and sadness [1]-[3]. The identification of human emotion will lead to effective communication and also build a healthy relationship. Intelligent recognition of human emotion is gaining much interest among the researchers in coping up with recent technologies like internet of things (IoT) and any other smart environments. Intelligent personal assistants (IPAs) [4] make use of natural language processing that communicates with human languages. These languages are augmented with different sorts of emotions. This has increased the stages of interaction with human-level intelligence [5]. The advancement in artificial intelligence (AI) has been actively working on different domains to improve some specific standards and also identify the unwanted elements [6]. It is also merging with other different applications such as market-based analysis, spam detection, and so on.

Deep learning algorithms are widely employed in real-time applications dealing with the concepts of pattern recognition and classification [7]-[9]. In this study, recognition facial expression is one of the research areas that imbibe the concept of pattern recognition models. It has revolutionized with different facets of computing modules in which human-computer interaction (HCI) becomes a part of our today's life. Information on the searching for human expressions is an important component of HCI [8] that contributes to efficient communication. It lessens the cost and computational time and also enhances the reliability rate of the application system. On the other hand, it provides a potential solution via non-verbal communication that supports human-human communication. The analysis of facial expression is an interesting and challenging study that has impacted many real-time applications related to HCI and medical applications. Literature studies on emotion detection deal with identifying the facial features for each facial expression [10]. The report states that there are 68 facial marks available to detect the type of emotions in real-time like positive, negative, and expressionless. Different studies have given different facial features to recognize facial expressions. Even though remarkable progress has been made in facial expression recognition, in the aspect of performance measures, still achieving high accuracy becomes a doubtful part.

Many studies have been instantiated to develop multimodal information such as light intensity, face, and background changes using facial terms to estimate multimodal systems, computational complexity has been used. The computational complexity has been used to estimate the multimodal systems. The multimodel strategy has better performance regarding emotional recognition compared with the single modal strategy. However, the combination of facial expressions with other modalities, such as voice signals, gestures, and biosignal, did not support the development of smarter human machine interaction (HMI) [11]. The availability of HMI has disrupted many opportunities related to real-time applications.

The contributions of this study are filtering, and enhancement of the facial image is applied to the input image that assists to extract the discriminative local features of a facial expression. The use of different types of filtered enhanced facial images removes the feature dimension reduction and the feature overlapping issues. Consequently, it improves the computational complexity of the face recognition rate. The deployment of the deep convolutional neural networks (D-CNN) classifier on the enhanced facial images has discriminated against the correlation of those features, which is the main essential task. The local information of facial regions contains the saliency information of a face and thus, the face recognition rate is achieved effectively in this study. Figure 1 presents the block diagram of the proposed technique.

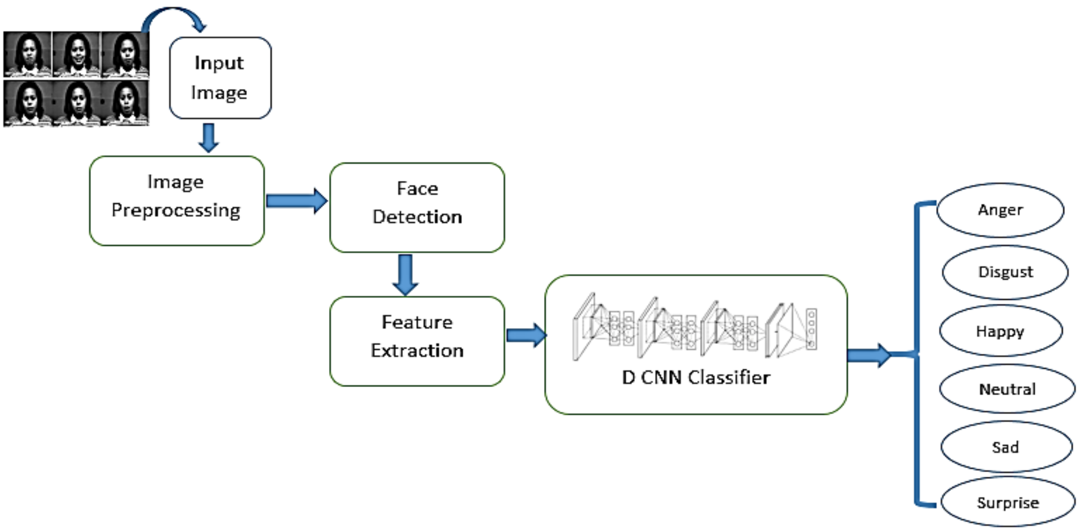

Figure 1. Overall block diagram of the proposed technique

The paper is organized as follows; section 2 presents the Literature Survey that discusses the reviews of existing techniques. Section 3 presents the research methodology that details the workflow of the proposed phases. Section 4 presents the experimental results and discussion which describes the achieved performance using the proposed techniques. Section 5 presents the conclusion that portrays the findings of the study.

2. THE COMPREHENSIVE THEORETICAL BASIS AND THE PROPOSED METHOD

The study of various existing research methodologies on facial expression recognition techniques and the application of different deep learning algorithms for improving the performance of the recognition process. The study of related work includes the review of existing literature about different techniques related to the performance enhancement of facial expression recognition models. To interpret the emotional states of the human mind, facial expression recognition is one of the efficient ways to deal with it. The recognition of facial expression from a human face is a troublesome task since face expression varies for different humans [12], [13]. There are 46 components or action units involved in the facial action unit codes (FACS) that contribute to the movements of facial muscles. Likewise, prkachin and solomon pain intensity scale (PSPI) is the metric used for exploring the frame-by-frame analysis of the FACS [14]. Recently, machine learning and intelligent systems are combined to detect the facial expression of the human face with the least computational efforts. The main challenge is the accurate development of facial expression recognition models with different inputs and different applications. The next challenges related to the unbalanced with relatively complex data beyond the nature of facial images. The classes are uniquely for various image sets that seek a versatile expert system. Such a system needs to create a representative model that captures the inherent general characteristics to achieve an unbiased classification or prediction of pain from facial images.

The temporal-based facial actions were employed to detect the face expression. A larger range of facial conduct using the facial muscle actions that compute the expressions was studied [15]. However, the accuracy is still in the developmental stage. Gabor feature-based boosted classifiers then helped to improve this. It was compared with the gentle boost, support vector machines, and hidden markov models. Convolutional neural networks [16] have explored the efficiency of visual recognition images. In which results were processed using the deep multi-layer network for the prediction of saliency, and forwarded to CNN for facial expression recognition, and its visual saliency maps. The performance results show that the widespread use of deep learning networks is 65.39% compared to other techniques.

Temporal-based reinforced approaches were studied to improve the facial expression recognition study [17] active appearance model (AAM) was introduced to extract the shape and texture models of facial images. Here, 16 geometric features were extracted to recognize the emotions. These extracted patterns were given as input to the relevance vector machine (RVM) that classified the different emotional states by achieving a recognition rate of 95.6%. Multi-layer perceptron [18] has been presented to classify the emotional states. Certain texture features of the facial skin were extracted using the biological visionary approaches. The concept of perceived facial images (PFI) has yielded better results using constructive training algorithms. A similar analysis for safety and security purposes has been extended to biometric analysis detecting the facet modules [19]. The concept of component analysis (PCA), independent component analysis (ICA), linear discrimination analysis (LDA), and neural networks are used for segmentation and facial recognition methods.

A two-fold random forest classifier was introduced to explore the facial expression [20]. With the help of facial motion analysis, active appearance model (AAM), and lucas-kanade (LK) an optical flow tracker, the facial feature points are assessed. It was quietly used for exploring the facial landmarks and the extraction of facial points. The execution of two layers of RF has efficiently detected the facial expression. It was better than the support vector machine (SVM) classifier. A robust feature extraction model using ChanVese energy and Bhattacharyya distance functions [21]. To eliminate the noise level, the separation between face and background was removed for extracting the local information. The face images were extracted using the trained hidden markov model (HMM). The results show greater efficiency in the detection of FER in terms of recognition rate. A sparse learning-based facial expression model [22] was studied to display the effects of emotional states. Sparse local fisher discriminant analysis (SLFDA) has been suggested to resolve multimodal learning problems. It has discriminated the power of LFDA with non-zero elements based on the selected face regions. It was then combined with the sparsity of projection vectors that accomplished all the extracted features. It was better than other dimensionality reduction models.

Neural network approaches were studied to discriminate the emotional states with better pattern recognition models [23]. All phases of image processing techniques such as pre-processing, function extraction, segmentation, and classification have been renovated to achieve efficient facial expression recognition. The recognition phase was tested under a single-layer neural network, multi-layer neural network, convolutional neural network, backward propagation network, and the recurrent neural network. Compared to machine learning approaches, neural network algorithms proved to have better efficiency. A CNN with the machine learning techniques [24] was combined to improve the steps involved in the image processing systems. The accuracy of the systems was improved under facial image databases; however, the computational complexity and time were not focused. Different types of algorithms for facial expression recognition depending on the biometric system have been discussed [25]. These were tested and compared with the feature reduction models like PCA, discrete cosine transforms (DCT), and types of wavelets. Table 1 summarises a comparative analysis of recent techniques.

Table 1. Comparative analysis of the recent techniques

Reference	Methods and the facial features	Classification technique	Database and face recognition rate	Main results
Rao *et al.* [26]	CNN-related Features	Fully Connected Layer	YTF DB has given an accuracy rate of 96.52%, PaSC DB-Acc:95.67%YTC DB -Acc:97.82	Here, the frame data is eliminated in this study during the reinforcement learning process.
Wang [27]	3D- dynamic features	Classified based on Distance Measurement	ORL Database containing 300 images from 30 subjects that give precision probability 100%	It has drastically improved the face detection rate.
Ding and Tao [28]	Branch-based ensemble features	Trunk-Branch Ensemble CNN model (TBE-CNN)	PaSC Database has yielded a verification rate of 96.12%; COX Face Database has given an identification raof te 98.96%; Youtube Database has obtained a verification rate of 94.96%	It is better than the conventional CNN in terms of extracting the deep features
Mokhayeri *et al.* [29]	Domain-specific face synthesis (DSFS)	SRC Classifier	COX-S2V DB has yielded pAUC:0.916 and AUPR:0.775	Computational complexity has been significantly reduced.
Mokhayeri *et al.* [30]	A synthetic plus variational model	SRC Classifier	COX-S2V DB has achieved pAUC:0.905 and AUPR:0.776	A face recognition system at different angles has been analyzed.
Xiaa *et al.* [31]	Geometrical motion	Random walk classifier	It has experimented on the CASME and SMIC	The study revealed the micro-expression spotting on facial features. Geometric analysis has reduced facial dimensionality.
Li *et al.* [32]	The local binary pattern and Histogram Oriented Optical Flow (HOOF)	Threshold techniques	It was analyzed using CASME II and SMIC datasets. Compared to the other datasets, this system has given an exceeded result 27.98% higher than the existing system.	The system has increased the True Positive rate.
Wang *et al.* [33]	Main Directional Optical Flow (MDMD)	Threshold techniques	CAS (ME)2 has given recall, precision, and F1-score were 0.32, 0.35, and 0.33, respectively	Micro-expression from the video sequences has been improved in terms of a better error rate.
Davison *et al.* [34]	LBP, OF, and 3D-HOG	Threshold techniques	SAMM, CASME II	The focus of action units of facial muscles was speculated under 3D. This ignores the influence of the overall facial emotions, it emphasizes local facial muscle changes, thereby reducing computational complexity and improving detection accuracy
Li *et al.* [35]	LBP with differential image analysis	Threshold techniques	SAMM, CASME II	The local temporal patterns of facial muscles have helped to spontaneously locate facial emotions.

Facial expression recognition is one of the latest biometric technologies. It is treated as a computer application that intelligently detects (or) recognizes (or) verifies the person in the form of a digital image. Despite extensive research, the performance of recognition systems is still in the early stages of development. Facial expression is a sort of internal expression (emotional state) that greatly influences expression on the human face, which offers a natural expression of emotional state, communication, and social communication on prohibited activities. The expression of facial emotion is the best form of nonverbal communication that provides information about the emotional state, mindset, and intention. The design of intelligent facial expression recognition systems composes deformations of facial components and their spatial relations, (or) changes in the pigmentation of the face. Additionally, a few factors like facial appearance, the high false recognition rate, twin attacks, poor lighting, and obstacles have been affecting the performance of facial expression recognition.

This section includes the framework of the proposed technique used for the design of an automatic facial expression recognition technique and discusses the application of deep matching algorithms for developing the recognition model in detail. This process has five phases such as face detection, image preprocessing, segmentation of facial region, feature extraction, and classification. The proposed multi-feature-based automatic facial expression recognition using deep convolutional neural network comprises four phases. Each phase has a significant contribution towards identifying facial emotions. Figure 2 presents the workflow of the proposed D-CNN using a multi-feature extraction module.

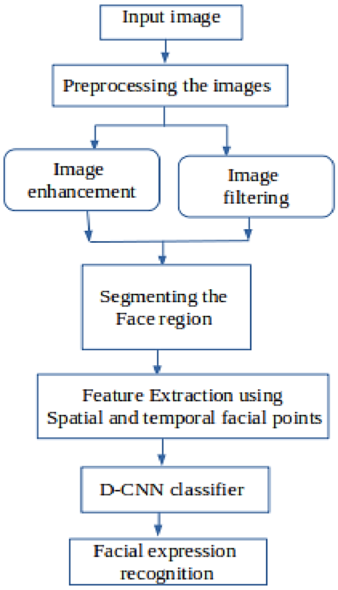

Figure 2. Proposed workflow

3. RESEARCH METHOD

3.1. Data collection and preprocessing

The extended cohn-kanade dataset (CK+) is collected from the publicly available datasets of size 256×256. The image is converted into a grayscale image so that the skin region appears brighter than the background. The algorithm then detects the skin region from the central point in the direction of the skin region. Generally, the images are subjected to random noise due to the performance of camera adjustment settings. Therefore, preprocessing methods play a vital role in this research study. To increase the quality and appearance of the images, image enhancement and filtering techniques are employed. These filters will reduce the peak-signal-to-noise-ratio (PSNR) and signal-to-noise ratio (SNR) of the image.

3.1.1. Image enhancement

The face images are enhanced using image sharpening techniques. The ultimate focus of the sharpening techniques is to enhance the visual quality of the images by increasing/decreasing the intensity level of brightness and contrast of an image. It appears like rings near the facial points by sharpening the transition of the image. It is profound near the edge of an image. This technique involves a few steps. It reads the input image from the dataset and converts RGB to the Grayscale image. Obtaining the histogram from the original image. The original and histogram-equalized images are integrated. Compare the image's performance parameters to the original image's performance parameters. Figure 3 presents the image enhancement process.

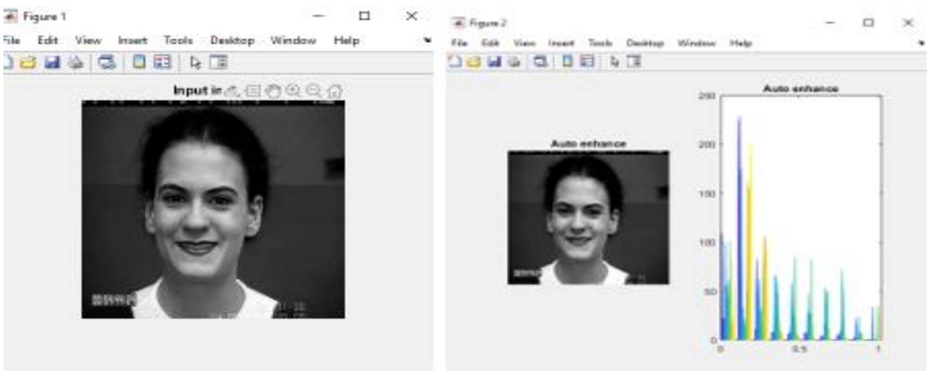

Figure 3. Image enhancement process

3.1.2. Image filtering

Image filtering is the process of improvising the appearance of the image. Different types of filters are used to increase contrast and provide a range of special effects to the image. The three filtering methods like Gaussian filter, Wiener filter, and adaptive mean filter are employed to preserve the edges of facial points.

1) Gaussian filter

Gaussian filter is generally utilized for removing the noise and smoothing. It requires computational assets and its productivity in implementing has been a motivation to study it. The gaussian operators are convolution operators, and convolution is used to achieve Gaussian smoothing. It is a 2D convolution operator that is utilized for image smoothing and noise removal. This Filter comprises two parameters: window dimensions and the standard deviation. If the value is huge, the picture smoothing impact will be higher. Filters of Gaussian smoothing are viable LPFs from the point of view of both the frequencies and spatial domains, are effective for implementation, and can be utilized adequately by engineers in practical applications of vision.

```
I1=fspecial ('gaussian’, [3 3],0.4); % choose the kernel size of 3x3
Iblur1 = imfilter (img, I1);
figure,
imshow(Iblur1)
title ('Gaussian Filter')
Iblur1=double(Iblur1);
[Gaussian_PSNR, Gaussian_SNR] = psnr (Iblur1, img);
```

2) Wiener filter

The Wiener filter is a filter that develops an estimated random process via a linear time-variant and the discrete wavelet transform over a noisy process. It helps the spatial domain to become a frequency domain. Each frame contains samples that convert to the domain of frequencies. The fourier discernment is one of the fastest algorithms (DFT):

$$D_k = \sum_{m=0}^{Nm-1} D_m \quad e_{\frac{-j2km}{Nm}}$$

where, k= 0, 1, 2.... N_{m-1}. The presence of DFT is estimated for all the computational processes which are useful.

```
h=fspecial ('gaussian’, [256,256],40);
h=mat2gray(h); %normalize [0,1]
img=imresize (img, [256 256]);
out_img_Wiener=image Restoration (img, h);
figure,
imshow(J)
title ('Weiner Filter')
out_img_Wiener=double(out_img_Wiener);
[Weiner_PSNR, Weiner_SNR] = psnr (out_img_Wiener, img);
```

3) Adaptive median filter

The median filter is a rank-order filter whose noise-reducing effects depend on the size and shape of the filtering mask. The complexity of the algorithm involved in getting the median value. Each pixel of an image is examined pixel by pixel, with each value substituted with the median value of the neighboring pixels. The term "Window" refers to the neighbor’s patterns. The pixel-by-pixel window scans throughout the whole image. The median value is calculated when the pixels of the window are first sorted into numeric order and then the pixel is replaced with the median value. Figure 4 presents the image filtering process, which shows how different filters, such as the Gaussian filter, adaptive median filter, and Weiner filter, improve the appearance of the image.

```
J = amedfilt2_calc (img);
figure,
imshow (J)
title ('Adaptive median Filter')
J=double (J);
[Adaptive_PSNR, Adaptive_SNR] = psnr (J, img);
```

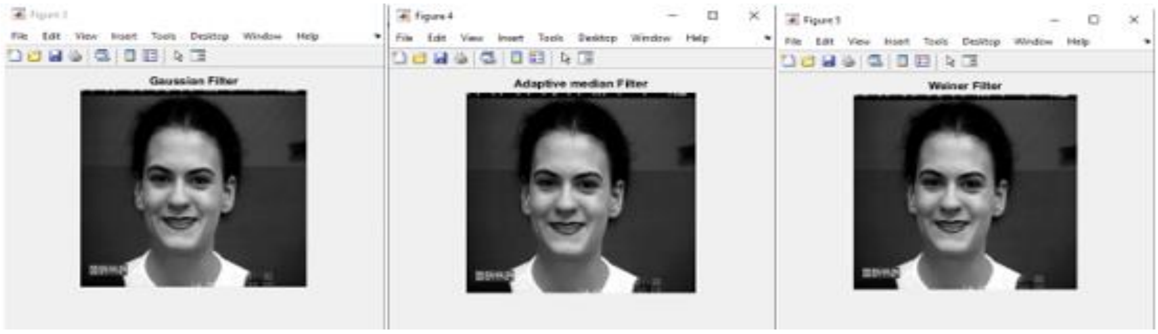

Figure 4. Image filtering process

3.2. Segmentation

The technique of segmenting a digital image into many segments is known as image segmentation. The Viola-Jones approach is proposed here for face detection and inverse discrete cosine transform (IDCT) is proposed here to segment the facial region of an image. The proposed IDCT is defined based on the foreground and background frame subtraction. Using this process, the segmentation will be processed which gives enhanced results when compared to existing DCT. Based on conditions, the segmented images are gathered. A regional local binary pattern (LBP) approach is used to gather the face points.

3.2.1. Viola Jones (VJ)-face detector technique

P. Viola and M. Jones have developed a human face detector that has achieved better detection accuracy. The Viola-jones detector performed better than the other detection algorithm in terms of accuracy metrics. It followed a sliding window approach that explored the locations of the pixels in an image that helped the window to capture the human face. Though the sliding window approach is a time-consuming process, moreover, the detector enhanced the detection speed. Integral images, which increase the speed of the convolution process, were introduced to boost detection speed. Depending on the window size, the complexity level of the computational process is declared, and selecting the relevant feature. The report states that the Adaboost algorithm was mainly employed on the small set of feature vector selection processes. It is very helpful in face detection concepts. Furthermore, by introducing the concepts of detection cascades, which follow the multistage detection pipeline while focusing on the face targets, the computational complexity is minimized.

```
function v = visort (v,N)  %#eml
temp = v;
for i=1:N-1
    m = v(i);
    k = 1;
        for j = i+1:N
        if v(j) < m
            m = v(j);
            k = j-i+1;
        end
    end
    for j = 1:k-1
        v(i+j) = temp(i+j-1);
    end
    v(i) = m;
    for j=1:N
        temp(j) = v(j);
    end
end
```

3.2.2. Inverse discrete cosine transform (IDCT)

In our proposed approach, we have employed an inverse-DCT (I-DCT) technique to improve the segmentation process. Discrete cosine transforms (DCT) are used to extract the foreground objects in this case. In normal cases, DCT is used for the compression process. Here we employ DCT for segmenting the object concerning the foreground and background subtraction. Here we consider the DCT attributes of decorrelation, preserving features, and complexity reduction to extract the background and foreground to segment the object. The utilization of DCT block transformation combined with different procedures brings about great retrieval efficiency. Much of the time, DCT could be viewed as a stage of preprocessing pursued by a pretty much complex strategy for the extraction of basic highlights. Then, user-defined functions are employed for DCT which makes the segmentation process more effective and the same will be then utilized in the next stage of feature extraction. The steps involved in the modified DCT are given as: i) The Initial frame is first selected; ii) Then each of the frames is compared with the first frame; and iii) The frame with the object and the empty frame is subtracted using the binary segmentation with morphological operations to obtain the image of the object alone.

3.3. Automatic extraction of regional lbp features

This step extracts the facial points of each facial expression like happy, sad, anger, and so on. Here, a regional local binary pattern (LBP) approach is employed to collect the facial points. Local binary pattern (LBP) is applied for extracting textural oriented features. It is an operator that defines the invariance of gray-scale texture measures. The texture is analyzed with multiple scales of combining neighborhood pixels. It is simple and robust for a computational study. The steps in local binary pattern are: Let the central pixel be estimated as, (X_c, Y_c) for color pixels, and the gray pixel is denoted as g_c. The LBP is calculated as:

$$LBP^{P,R}(x_c, y_c) = \sum_{i=1}^{P} s\left(g_i^{P,R} - g_c\right)^{2^{i-1}}$$

where the constraints of s(x) are given as:

$$s(x) = \{1, if x \geq 0\}; \{0, if x < 0\}$$

3.4. Classification using D-CNN

These feature vectors i.e facial texture points are fed as input to the deep convolutional neural networks. D-CNN is fundamentally utilized in convolving an image along with kernels to obtain feature maps. The weights within the kernels help to connect every unit of the feature map to prior layers. These kernel weights are used at the time of dataset training to enhance the input characteristics. The weights that require training within the convolutional layers are lesser than those for layers that are fully connected since the kernels are typical to each unit of the specific feature map. Feature vectors of each facial point are fed to D-CNN. CNN's functionality can be divided into four major categories.

i) LBP facial features will be fed to the input layer.
ii) The output for the neurons associated with the input local regions will be determined by the convolutional layer through the computation of a scalar product between the regions associated with the volume of the input and the weights of the neurons.
iii) The pooling layer then downsampled the input, reducing the number of parameters for that particular activation.
iv) The fully connected layer will then generate scores of the classes (based on the activations) that will be used in the classification process.

Once the training of all facial features is completed, the test facial points are fed to the classifier. Finally, the emotion of the given input database is done. Figure 5 presents the architecture of the proposed approach. It presents the detailed flow diagram of the proposed technique.

Proposed algorithm:
Step 1: Read the input image from the dataset in ‘png’ format.
Step 2: Convert RGB to Grayscale image.
Step 3: Obtaining the histogram from the original image.
Step 4: Compare the performance parameters image's performance parameters to the original image's.
Step 5: Apply a 2D convolution Gaussian operator that is utilized for image smoothing and noise removal.
Step 6: Each pixel of an image is explored pixel by pixel, substituting each value for the median value of the neighboring pixels.
Step 7: The pixel-by-pixel window scans throughout the whole image.
Step 8: The median value is calculated when the pixels of the window are first sorted into numeric order and then the pixel is replaced with the median value.
Step 9: Apply Viola jones’s algorithm for face detection.
Step 10: Image reconstruction by inverse of modified DCT.
Step 11: Initial frame is first selected.
Step 12: Set values less than 5 to 0 in DCT matrix.
Step 13: Then each of the frames is compared with the first frame.
Step-14: The frame with the object and the empty frame are subtracted using the binary segmentation with morphological operations to obtain the image of the object alone.

Step 15: Apply the Gabor filter.
Step 16: Create the gray level co-occurrences matrices (GLCMs).
Step 17: Derive Statistics from GLCM.
Step 18: Test feature = [Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Gab means, Gab_max, Gab_min];
Step 19: These feature vectors i.e facial textural points are fed as input to the deep convolutional neural networks. The LBP facial features will be fed to the input layer.
Step 20: The output for the neurons associated with the input local regions will be determined by the convolutional layer through the computation of a scalar product between the regions associated with the volume of the input and the weights of the neurons.
Step 21: The pooling layer then downsamples the input, reducing the number of parameters for that particular activation.
Step 22: The fully connected layer will then generate scores for the classes (based on the activations) that will be used in the classification process.
Step 23: Define the network layers.
Step 24: Specify the training options.
Step 25: Train the network using training data.
Step 26: Applying confusion matrix to interpreted performance (i.e., accuracy, sensitivity, specificity, recall, and precision)

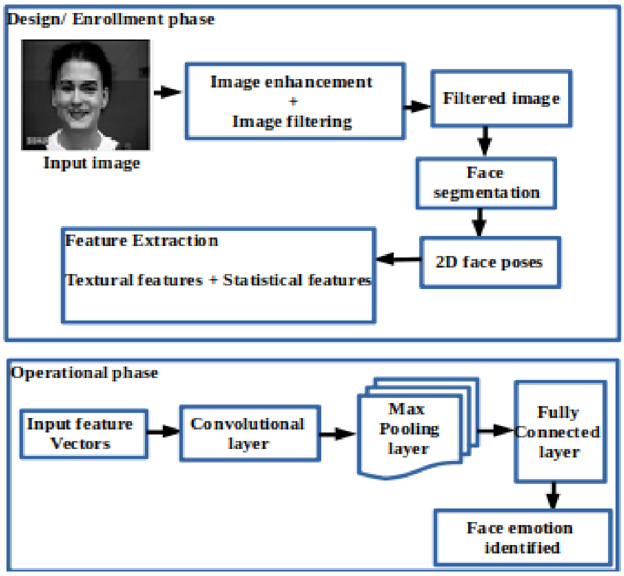

Figure 5. Proposed architecture

4. RESULTS AND DISCUSSIONS

This section explains the analysis of the simulation results and performance evaluation of the proposed methodology. The proposed D-CNN method is implemented in the MATLAB programming language. The dataset is collected from the public repository, named extended cohn kanade (CK+) that contains face images expressing happiness, sadness, anger, and so on. The dataset consists of 593 images with 123 concepts. Each input video has a different facial expression from neutral to peak levels with 640x490 or 640x480 pixels. Initially, the dataset is split into training and testing datasets. The D-CNN makes use of training datasets and further, it is evaluated from the testing dataset. The proposed D-CNN is compared with the existing SVM.

Figure 6 present the input image, which is selected from the testing dataset. Figure 7 presents the image enhancement process. It is the first step of the implementation process. It employs an auto-enhancement method that incorporates image brightness, contrast, blur, and smoothing effects. The intensity level of each effect is adjusted resulting in the visual quality of the image.

Figure 8 presents the Gaussian filtering process. As we know, PSNR and SNR are the most quality metric employed to inquire about the quality level of an image. This process reduces the blurriness of the image. By doing so, the contrast and edge of the image are preserved. The PSNR value is 46.9594 and SNR is 40.6113.

The Figure 9 presents the adaptive mean filtering process. It eliminates the salt-and-pepper noise that enhances the quality of non-noise pixels. The PSNR value is 38.7072 and SNR is 32.3591. Figure 10 represents the Wiener filter process. It removes the additive and spectra noise under a random process. The PSNR value is 9.1759 and SNR is 2.8278.

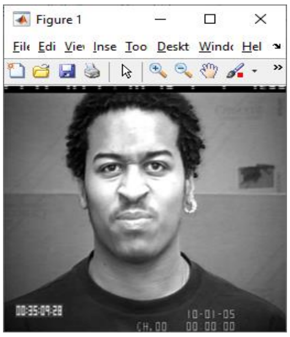

Figure 6. Input image

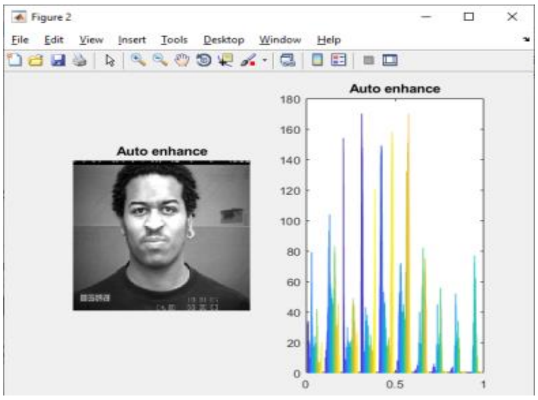

Figure 7. Image enhancement process

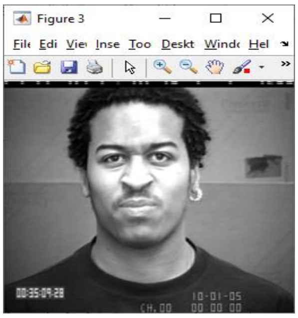

Figure 8. Gaussian filtering process

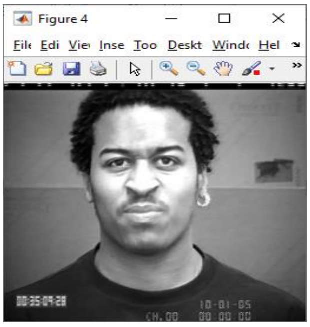

Figure 9. The adaptive mean filtering process

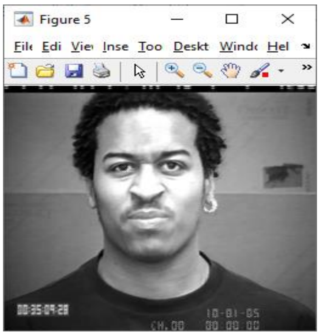

Figure 10. Wiener filter process

Figure 11 represents the segmentation process. Here, the face region is extracted from other background images. By the application of the VJ technique, the face image is cropped with the best features model. Figure 12 presents the inverse-DCT that exactly crops the facial points for identifying the facial expression. Finally, the output of inverse-DCT is fed into the D-CNN that recognizes given facial expression i.e., anger, which is shown.

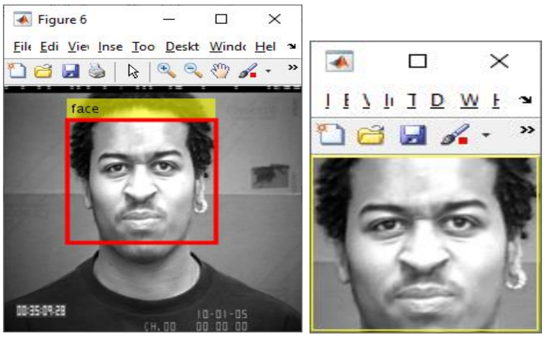

Figure 11. Segmentation process

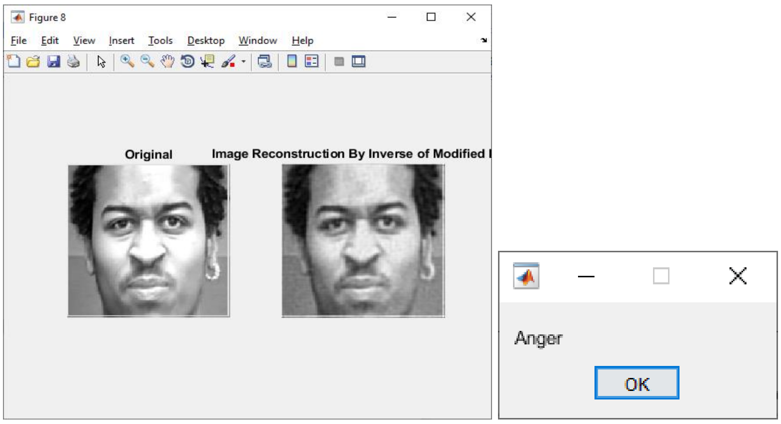

Figure 12. Inverse-DCT

The configuration matric is used for representing the result and to calculate the performance parameters such as accuracy, recall, sensitivity, specificity, and precision. A confusion matrix is generated on all criteria's to empirically evaluate the performance of developed authentication schemes. Where:

i) True positive (TP) = Number of original authentication credentials that have been authenticated successfully by the authentication system.
ii) True negative (TN) = Number of fake authentication credentials that have been rejected successfully by the authentication system.
iii) False positive (FP) = Number of fake authentication credentials that have been authenticated by the authentication system.
iv) False negatives (FN) = Number of original credentials that have been rejected by the authentication system.

For the validation of the proposed technique and the existing technique 593 image authentication credentials has been used. Fake credentials have a random chance of any one of the parameters used for authentication such as change of image part or change of random number. Some instances have been applied for all the image parameters for both techniques. After validation class labels are classified into four categories as shown in Table 2. The same has been applied to all the parameters of the image for the techniques.

Table 2. Confusion matrix representation

	Actual Authenticated	Actual Rejected
Predicted authenticated	True positive (TP)	False positive (FP)
Predicted rejected	False negative (FN)	True negative (TN)

Evaluation metrics are the criteria that define a specific quantitative evaluation of a developed system or proposed technique. This work is evaluated on evaluation metrics such as precision, recall, false positive rate (FPR), overall accuracy, and F-measure which are as follows:

a) Accuracy: Measure the number of absolutely, correctly classified instances either positive or negative divided by the entire number of instances. The equation for computing the overall accuracy is:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) * 100 \quad (1)$$

b) Precision (P): It represents the rate of cases of authentication credentials authenticated successfully that are the original credentials. The equation to calculate it is as follows:

$$Precision\ (P) = TP/\ (TP + FP)\ * 100 \quad (2)$$

c) Recall: (TPR): It is also called the True positive rate (TPR). It is the rate of the number of original credentials authenticated correctly. The TPR defines how many correct positive results occur among all positive samples available during the test. TPR is equivalent to sensitivity and its equation is:

$$Recall/TPR = TP/(TP + FN)\ * 100 \quad (3)$$

d) Specificity: It is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called true negative rate (TNR). The equation for computing the specificity:

$$Specificity\ (SP) = TN/\ (TN + FP)\ * 100 \quad (4)$$

Table 3 shows the achieved performance parameters for existing and new strategies, and Figure 13 shows a graphical representation of the performance parameters obtained from both existing and proposed techniques. Several exhaustive tests have been carried out to ensure that the desired accuracy is minimized. An in-depth study of many aspects of the generation of SVM and DCNN with an emphasis on parameters that affect the accuracy of FER classification. Specifically, a dependency on the number of base group classifiers, various SVM techniques, and various aggregation schemes. The ultimate goal is to make some guidelines to build efficient CNN ensembles. It has been discovered, for example, that utilizing various sources of variability is critical for improving overall accuracy. Pre-processing and pre-trainer procedures may provide for this purpose sufficient variability across the basic classifiers, whereas the use of various seeds does not appear to be an effective solution. Several images of six basic expressions are considered for testing the proposed techniques and achieved results are explained in the Table 4 concerning all the expressions. Figure 14 shows the graphical representation of the result achieved for different expressions. The proposed technique finally gives results after applying configuration matrices for the calculation of accuracy, recall, sensitivity specificity, and precision for each image.

Table 3. Performance metrics analysis-existing and proposed

Metric	Proposed (D-CNN)	Existing (SVM)
Accuracy	91.22	76.30
Recall	86.36	86.36
Sensitivity	86.36	86.37
Specificity	91.68	75.45
Precision	67.50	20.28

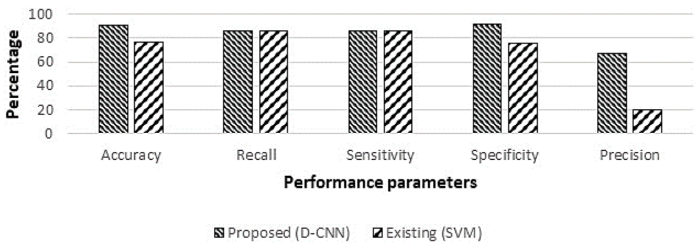

Figure 13. Performance chart-existing and proposed

Table 4. Performance matric for six different expressions

Metric	Anger	Disgust	Happy	Neutral	Sad	Surprise
Accuracy	93.7008	92.0792	93.0435	93.7984	94.3463	98.5946
Recall	97.6834	94.1748	96.7614	97.7860	98.2659	98.4416
Sensitivity	97.6834	94.1748	96.7614	97.7860	98.2659	98.4416
Specificity	91.6501	91.6501	91.6501	91.6501	91.6501	91.6501
Precision	85.7627	69.7842	81.1659	86.3192	89.0052	90.0238

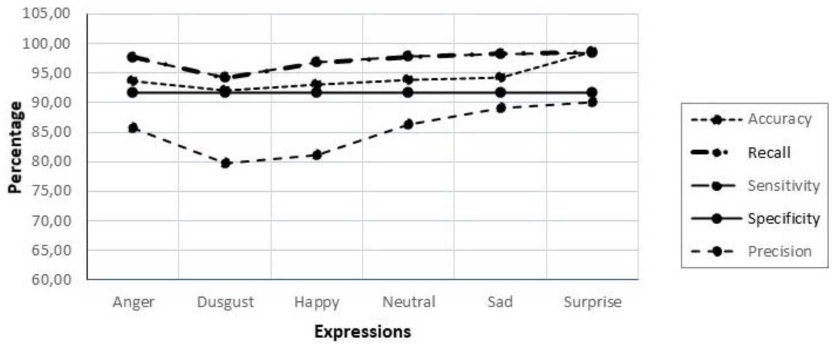

Figure 14. Curve represents the performance parameters for different expressions

5. CONCLUSION

In this research paper, we have proposed a multi-feature-based deep convolutional neural networks (D-CNN) that identifies the facial expression of the human face. The literature states that the local information is being misused in the case of recognition processes that degrade the performance of the recognition systems. A multi-feature learning model is employed in this study to overcome the false recognition rate issue. The collected images are preprocessed and enhanced via three filtering methods: Gaussian, Wiener, and adaptive mean filtering. The preprocessed image is then segmented using a face detection algorithm. The detected face is further applied with LBP that extracts the facial points of each facial expression. These are then fed into the D-CNN that effectively recognizes the facial expression using the features of the facial poi. The proposed D-CNN is implemented with the existing SVM. Compared to the SVM, the analysis of deep features helps to extract the local information of the data without incurring a higher computational effort.

REFERENCES

[1] F. Albregtsen “Statistical texture measures computed from gray level co-occurrence matrices,” *Image processing laboratory, Department of informatics, university of OSLO*, p. 5, 2008.

[2] D. Baswaraj, A. Govardhan, and P. Premchand, “Active contours and image segmentation,” *Global Journal of Computer Science and Technology*, vol. 12, no. 11, 2012.

[3] A. Bouzalmat, N. Belghini, A. Zarghili, and J. Kharroubi, “Face Detection and Recognition using Back Propagation Neural network and Fourier Gabor filters,” *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 2, no. 3, pp. 15, 2011, doi: 10.5121/sipij.2011.2302.

[4] P. Carcagnì, M. D. Coco, M. Leo, and C. Distante, “Facial expression recognition and histograms of oriented gradients: a comprehensive study,” *Springer Plus*, vol. 4, no. 1, pp. 1, 2015, doi: 10.1186/s40064-015-1427-3.

[5] X. Chen and W. Cheng, “Facial expression recognition based on edge detection,” *International Journal of Computer Science & Engineering Survey (IJCSES)*, vol. 6, no. 2, pp. 1, 2015, doi: 10.5121/ijcses.2015.6201.

[6] A Ferreira, R. L. Silva, W. C. Celeste, T. F. B. Filho, and M. S. Filho, “Human–machine interface based on muscular and brain signals applied to a robotic wheelchair,” *16th Argentine Bioengineering Congress and the 5th Conference of Clinical Engineering, Journal of Physics*, vol. 90, 2007, pp. 012-094, doi: 10.1088/1742-6596/90/1/012094.

[7] H. S. Dadi and G. M. Pillutla, “Improved Face Recognition Rate Using HOG Features and SVM Classifier,” *IOSR Journal of Electronics and Communications Engineering*, vol. 11, no. 4, pp. 34-44, 2016, doi: 10.9790/2834-1104013444.

[8] A. P. Gosavi and S. R. Khot, “Facial expression recognition using principal component analysis,” *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 4, pp. 2231-2307, 2013.

[9] B. He. *et al.*, “Fast face recognition via sparse coding and extreme learning machine,” *Cognitive Computation*, vol. 6, no. 2, pp. 264-277, 2014, doi: 10.1007/s12559-013-9224-1.

[10] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” IEEE *Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 2, pp. 13-19, 2012.

[11] J.-S. Jang, K. H. Han, and J. H. Kim, “Face Detection using Quantum Inspired Evolutionary Algorithm,” *In Evolutionary Computation, CEC2004. Congress on IEEE*, vol. 2, pp. 2100-2106, 2004.

[12] A. B. Ashraf *et al.*, “The painful face–pain expression recognition using active appearance models,” *Image and Vision Computing*, vol. 27, no. 12, pp. 1788-1796, 2009, doi: 10.1016/j.imavis.2009.05.007.

[13] R. Kharghanian, A. Peiravi, and F. Moradi, “Pain detection from facial images using unsupervised feature learning approach,” *38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2016, pp. 419-422.

[14] K. M. Prkachin and P. E. Solomon, “The structure, reliability and validity of pain expression: Evidence from patients with Shoulder Pain,” *Pain*, vol. 139, no. 2, pp. 267-274, 2008, doi: 10.1016/j.pain.2008.04.010.

[15] M. F. Valstar and M. Pantic, “Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics,” *BNAIC, Belgian-Dutch Conference on Artificial Intelligence*, pp. 367, 2008.

[16] Viraj Mavani, S. Raman, and K. P. Miyapuram, “Facial Expression Recognition using Visual Saliency and Deep Learning,” *ArXiv preprint arXiv:1708.08016*, 2017, doi: 10.1109/ICCVW.2017.327.

[17] K. T. Song and C.-Y. Lin, “Robust facial emotion recognition using a temporal-reinforced approach,*” 14th International Conference on Control, Automation and Systems (ICCAS 2014) on IEEE*, 2014, pp. 804-807, doi: 10.1109/ICCAS.2014.6987889.

[18] H. Boughrara, M. Chtourou, C. B. Amar, and L. Chen, "Facial expression recognition based on a MLP Neural network using constructive training algorithm," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 709-731, 2016, doi: 10.1007/s11042-014-2322-6.
[19] S. Tikoo and N. Malik, "Detection, segmentation and recognition of face and its features using neural network," *Journal of Biosensors and Bioelectronics*, vol. 7, no. 2, arXiv preprint arXiv:1701.08259, 2017, doi: 10.4172/2155-6210.1000210.
[20] X. Pu, K. Fan, X. Chen, L. Ji, and Z. Zhou, "Facial expression recognition from image sequences using twofold random forest classifier," *Neurocomputing*, vol. 168, pp. 1173-1180, 2015, doi: 10.1016/j.neucom.2015.05.005.
[21] M. H. Siddiqi, R. Ali, A. M. Khan, E. S. Kim, G. J. Kim, and S. Lee, "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Multimedia Systems*, vol. 21, no. 6, pp. 541-555, 2015, doi: 10.1007/s00530-014-0400-2.
[22] X.-H. Wang, A. Liu, and S. Q. Zhang, "New facial expression recognition based on FSVM and KNN," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 21, pp. 3132-3134, 2015, doi: 10.1016/j.ijleo.2015.07.073.
[23] M. M. Kasar, D. Bhattacharyya, and T. H. Kim, "Face recognition using neural network: A review," *International Journal of Security and its Applications*, vol. 10, no. 3, pp. 81-100, 2016, doi: 10.14257/ijsia.2016.10.3.08.
[24] A. T. Lopes, E. D. Aguiar, A. F. D. Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610-628, 2016, doi: 10.1016/j.patcog.2016.07.026.
[25] A. P. and M. K. Tewari, "Study of Different Algorithms for Face Recognition," *(Doctoral dissertation)*, 2010.
[26] Y. Rao, J. Lu, and J. Zhou, *Attention-aware deep reinforcement learning for video face recognition*, IEEE International Conference on Computer Vision, ICCV, 2017, pp. 39510-3960.
[27] M. Ju-Wang, *Face feature dynamic recognition method based on intelligent image*, International Conference on Virtual Reality and Intelligent Systems, ICVRIS, Changsha, China, 2018, doi: 10.1109/ICVRIS.2018.00022.
[28] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002-1014, 2018, doi: 10.1109/TPAMI.2017.2700390.
[29] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Domain-specific face synthesis for video face recognition from a single sample per person," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 757-772, 2019, doi: 10.1109/TIFS.2018.2866295.
[30] F. Mokhayeri and E. Granger, "A paired sparse representation model for robust face recognition from a single sample," *Computer Vision and Pattern Recognition, arXiv*, vol. 100, pp. 107-129, 2020, doi: 10.1016/j.patcog.2019.107129.
[31] Z. Xiaa, X. Feng, J. Peng, X. Peng, and G. Zhao, "Spontaneous micro-expression spotting via geometric deformation modeling," *Computer Vision and Image Understanding*, vol. 147, pp. 87-94, 2016, doi: 10.1016/j.cviu.2015.12.006.
[32] X. Li *et al.*, "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods," *IEEE Transactions* on *Affective Computing*, vol. 9, no. 4, pp. 563-577, 2017, doi: 10.1109/TAFFC.2017.2667642.
[33] S.-J. Wanga, S. Wu, X. Qian, J. Li, and X. Fu, "A main directional maximal difference analysis for spotting facial movements from long-term videos," *Neurocomputing*, vol. 230, pp. 382-389, 2017, doi: 10.1016/j.neucom.2016.12.034.
[34] A. Davison, W. Merghani, C. Lansley, C.-C. Ng, and M. H. Yap, "Objective micro-facial movement detection using FACS - based regions and baseline evaluation," *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG).* IEEE, 2018, pp. 642-649, doi: 10.1109/FG.2018.00101.
[35] J. Li, C. Soladie, R. Seguier, S.-J. Wang, and M. H. Yap, "Spotting micro-expressions on long videos sequences," *14th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2019, pp. 1-5, doi: 10.1109/FG.2019.8756626.

BIOGRAPHIES OF AUTHORS

Anjali Dixit received her Master of Science in Software Systems at Birla Institute of Technology and Science, Rajsthan University in 2010. She was with Shri Vaishnav Institute of science and Technology Indore (Madhya Pradesh), from 1998 to 2002. From 2002 to 2007, she worked in Industry. She is currently working in Birla Institute of Technology and Science, K K Birla Goa Campus, Goa since 2007. She has five years of industrial and over 18 years of academic and administrative experience. She can be contacted at email: anjalindixit1802@gmail.com.

Tanmay Kasbe is an Associate Professor in the Department of Computer Science and Engineering at Oriental University, Indore. He has completed BCA (2003), MCA (2007) & PhD (2019) in Computer Science specialization in Artificial Intelligence and Machine Learning. He has more than 12 years of teaching and research experience. Dr. Tanmay Kasbe has published research papers in SCI/Scopus indexed international conferences and journals and is associated with various international journals as a reviewer and editorial board member. His research areas include Artificial Intelligence, Machine Learning, Image Processing, Disease Diagnosis, and Wireless networks. He can be contacted at email: tanmay.kasbe@gmail.com.