Multi-Feature Integrated Concurrent Neural Network for Human Facial Expression Recognition

C. R. Dhivyaa[1*], K. Nithya[1], K. Karthika[1], S. Mythili[2]

[1] Department of Computer Science and Engineering, Kongu Engineering College, India
[2] Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, India
crdhivyait@gmail.com, knithya89@gmail.com, kndkarthi@gmail.com, mythilikarthikeyan911@gmail.com

Abstract

Facial expression helps to communicate between the people for conveying abundant information about human emotions. Facial expression classification is applied in various fields such as remote learning education, medical care, and smart traffic. However, due to the complexity and diversity of the facial emotions, the present facial expression recognition model causes a low recognition rate and it is hard to extract the precise features that are related to facial expression changes. In order to overcome this problem, we proposed Multi-feature Integrated Concurrent Neural Network (MICNN) which is significantly different from the single neural network architectures. It aggregates the prominent features of facial expressions by integrating the three kinds of networks such as Sequential Convolutional Neural Network (SCNN), Residual Dense Network (RDN), and Attention Residual Learning Network (ARLN) to enhance the accuracy rate of facial emotions detection system. Additionally, Local Binary Pattern (LBP) and Principal Component Analysis (PCA) are applied for representing the facial features and these features are combined with the texture features identified by the Gray-level Co-occurrence Matrix (GLCM). Finally, the integrated features are fed into softmax layer to classify the facial images. The experiments are carried out on benchmark datasets by applying k-fold cross-validation and the results demonstrate the superiority of the proposed model.

Keywords: Facial expression recognition, Feature extraction, Residual dense network, Attention residual learning network, Sequential convolutional neural network

1 Introduction

Facial Expressions convey nonverbal communication and it helps the listener to understand the non-spoken emotions from the human face. The universal facial expressions states like happiness, sadness, anger, surprise, fear, and disgust are extracted from human facial images to understand their emotions. Facial Expression is one of the important features which can be recognized by humans or machines. Humans can understand the emotion of a person who is present in front of them and it is possible in the case of one-to-one communication. In some cases, one human may interact with many people and in that situation it is difficult to grasp everyone's emotions individually. In the kinder garden school, classroom environment students are not able to express their feelings exactly through verbal communication. To understand their emotions, facial expression recognition technology is most helpful. To identify students' emotions in a group, artificial intelligent-based machines/robots are used, which are trained by the human input and it classifies what exactly the emotion means. Human face images along with their emotions are fed into the system for training purposes further it classifies the test image to any of the emotion categories.

Human Facial Expression Recognition (HFER) covers three predominant areas such as human face detection from images, feature extraction from segmented faces, and human emotion classification. HFER accuracy is improved through various researches by doing some changes in the above technique. During facial expression identification all the human face regions are highlighted and the expression-relevant features are extracted then emotions are identified by the aggregated facial features. In the state-of-art methods, many researchers suggest various feature extraction algorithms such as AdaBoost learning algorithm with Gabor filter [1], Viola-Jones object detection algorithm [2], Convolution Neural Network [3], Weighted Deep Convolution Neural Network [4-5], Residual Neural Network [6], Capsule Network [6], and so on for isolating the unique emotion regions from the face. Similarly, with the highlighted features the classifiers like SoftMax [6-7], Support Vector Machine [8] are used for recognizing the expression category. Most of the Convolution Neural Network-based approaches are more suitable for face identification and emotion recognition.

To extract the exact emotion from facial images, various feature extraction techniques like Local Binary Pattern (LBP) [8], Principle Component Analysis (PCA) are used for low-level expression feature extraction and Deep convolution neural network supports for high-level feature extraction. Along with this texture, feature analysis is performed for extracting the rich set of features for facial expression classification. Moreover, according to the research about classifying the facial expressions of human, the extraction of prominent low level and high-level features are essential. Inspired by the various research challenges in facial expression recognition we have created a novel feature extraction model to classify the expressions of humans.

Many researchers have suggested parallel neural network architecture to extract rich set of features for image

*Corresponding Author: C. R. Dhivyaa; E-mail: crdhivyait@gmail.com
DOI: 10.53106/160792642022112306009

classification. For breast cancer image classification, convolution neural network and recurrent neural network are modeled in parallel structure [9] to find the more number of features from the images. The two convolution neural networks [10] are arranged in a parallel architecture to improve the feature extraction capability. In a COVID-19 screening system [11], the chest X-ray images are analyzed by designing the dilated CNN in a parallel manner. Motivated by the previous studies, we have constructed the parallel model of sequential convolution neural network, residual dense network, and attention residual learning network for extracting huge number of significant features.

Our research focuses Multi-Feature Integrated Concurrent Neural Network for Human Facial Expression Recognition which consolidates the features at different levels and passes this information through the concurrent neural network for human emotion identification. The outline of our work is described as follows:

1) For improving the feature detection efficiency, in the preprocessing stage, pixel intensity of the image is changed and normalized by adjusting the contrast. Also, features at multi categories such as texture, LBP feature, and PCA are aggregated to form a unique feature map.

2) A Novel concurrent neural network performs feature extraction in three different convolution network architectures which are finally concatenated. This network architecture includes sequential convolution neural network, residual dense network, and attention residual learning network.

3) To identify the feature emotion SoftMax multiclass classifier is deployed.

The sections of the research work are arranged as follows: Section 1 introduces the Human Facial Expression Recognition methods; Section 2 elaborates the existing facial expression identification methods; Section 3 proposes a combination of three feature extraction techniques along with concurrent neural network; Section 4 explains the results of the proposed model, and Section 5 concludes the full text.

2 Related Works

Facial expression detection is mainly consisting of feature detection and identification of facial emotions that are expressed on the human face. The traditional feature detection methods include Local Binary Pattern (LBP) [12-13], histogram of oriented gradients [14], facial action units [15], Principal component analysis [16], and so on for detecting the relevant features from the face for expression classification. An artificial neural network-based ensemble classifier is employed in [17] for analyzing the multicultural facial expression where Local Binary Pattern (LBP), Principal Component Analysis (PCA), and uniform Local Binary Pattern (LBP) are applied for representing the facial features. [18] proposes multitask cascaded convolutional networks which have cascade detection features for complete face identification and captured face coordinates are sent to the facial expression classification model to recognize the facial emotions. This model uses residual modules and depth-wise separable convolutions to reduce the number of parameters that need 0.496GB memory to complete the facial expression classification with an accuracy of 67% on the FER-2013 dataset.

To improve the recognition capabilities of the system, fine-grained facial expression recognition [19] is built to design an end-to-end Multi-Scale Action Unit (AU)-based Network (MSAU-Net) for learning more robust facial representation by focusing the facial actions and aggregating the local features. This MSAU-Net is extended to a two-stream model (TMSAU-Net) for recognition with video by concatenating the module with an attention mechanism and a temporal stream branch for jointly learning both spatial and temporal features. For extracting the expression image features, fuzzy C-means clustering algorithm [20] is applied to the convolutional layer of CNN to upgrade the ability of the feature detection model by solving random initialization problem of the convolution kernel, and also softmax function is replaced with support vector machine to increase the detection capability of the model which improves the facial expression recognition rate.

Facial expression detection is essential to monitor the emotions of the teacher for improving the quality of the teaching and professional development. [21] combined Convolutional neural network (CNN) [22] and attention mechanism to detect the expressions of the teacher. Migration learning is also applied to rectify the overfitting problem in the training process of the network and furthermore, combination of InceptionResNetV2 and CBAM network is proposed to detect similar features from the facial expressions that outperform the network without attention mechanism. This network provides the 78% classification rate on EIDB-13 intensity-based facial expression dataset and 88% of accuracy rate on the RAF-DB public macro expression dataset. A novel joint deep learning of facial expression synthesis and recognition method is developed in [23] for effective face expression recognition where the facial expression synthesis generative adversarial network (FESGAN) is pre-trained for creating face images with various emotions then expression recognition network is jointly learned with the pre-trained FESGAN. The classification loss is computed to optimize both performances of recognition and generator of FESGAN.

The accurate features can be detected hardly when it is highly correlated with facial changes. In order to solve this problem, [24] proposes a parallel neural network fusing texture features for improving the recognition of facial expression. The convolution neural network, residual network, and capsule network are used to construct the parallel neural network, and analysis of texture is performed on input face image to find the large features from the images for classification with an accuracy of 98.14%, F1-score of 0.9801. A deep learning-based frequency neural network is developed in [25] for facial expression recognition which processes the images in the frequency domain to learn the features by constructing the multiple multiplication layers followed by high-level features are identified using summarization layer and Basic-FreNet is constructed by utilizing the multiplication layers and summarization layer based on the discrete cosine transform (DCT). To improve the performance of the basic frequency network, the Block frequency network is developed for learning the features, and dimension reduction is done by block sub-sampling. This is the first attempt at using a frequency-based network model for expression detection which attains superior results with low computational cost.

A joint framework includes facial emotion classification, face alignment, and face synthesis are consistently associated tasks and it can be achieved by proposing a novel end-to-end deep learning model [26] which jointly utilizes code of expression, code of geometry, and generated data for concurrent pose-invariant facial emotions recognition, face image synthesis, and face alignment. The weight-adapted convolution neural network (WACNN) is used in [27] for extracting discriminative expression representations to detect facial expressions. In this model, after pre-process the facial expression images, the principal component analysis is applied to identify the low-level features then high-level features are detected and identified by WACNN which is optimized by a hybrid genetic algorithm. The effectiveness of the model is evaluated on JAFFE, CK+, and static facial expressions that are analyzed by k-fold cross-validation and yields superior results than other state-of-the-art methods.

3 Proposed Methodology

The key focus of our research work is to design an automatic facial expression identification system that can classify the universal facial expressions like happiness, anger, fear, sadness, disgust, neutral, and surprise. The proposed model is consists of pre-processing the images, feature extraction, and facial expression classification. In pre-processing, the input images are pre-processed such as face detection, resizing, and normalization to identify the regions of interest for feature extraction. After pre-processing, the relevant features are extracted using LBP, PCA, and also texture features are detected for classifying the facial emotions. Further, the identified features are aggregated and passed as input to the multi-feature integrated concurrent neural network which is constructed by integrating the sequential convolutional neural network, residual dense network, and attention residual learning network for recognizing the facial expressions of humans and the framework of the proposed model is displayed in Figure 1.

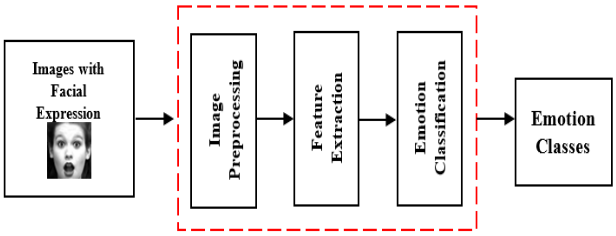

Figure 1. Human facial expression recognition

3.1 Pre-processing

In real-time, the facial expression image contains the key information of facial expression and also includes the noise of the surrounding environment, hair and so on which will inhibit the features of the facial expression recognition. Hence, pre-processing of facial expression images is necessary to remove irrelevant information in the input facial image for detecting high-quality features when performing facial expression operations. Pre-processing the input image mainly includes noise removal, face detection and normalization of the image. First, noise is reduced by applying hybrid median filter [28] and geometry normalization is performed to localize the face and extract the key areas of the face by applying the Viola-Jones face detector algorithm [29] which obtains the rich set regions of facial expression distribution shown in Figure 2 defined by the function Y(x),

$$Detection = \begin{cases} face, & if\ Y(x) > 0 \\ non-face, & otherwise \end{cases} \tag{1}$$

$$Y(x) = \sum \alpha_t y_t(x); \qquad y_t(x) = \begin{cases} +1, & if\ \ h_t(x) > \theta_t \\ -1, & otherwise \end{cases} \tag{2}$$

Where, $h_t(x)$ is direction of inequality, θ_t is threshold value, $y_t(x)$ is number of features, α_t is feature weight.

Figure 2. Geometric normalization

Then, the images are resized and intensity normalized for further processing. Resizing of the image is done by using the bicubic interpolation technique [30] to preserve the fine details of the facial expressions which have created a high-quality image with 150 pixels height and 128 pixels width. Furthermore, histogram equalization is applied to reduce the light intensity on facial expression features by adjusting the intensity values of the facial expression images.

The histogram equalization is denoted by the function g(x),

$$g(x) = \sum_{i=0}^{x} \frac{n_i}{n}, \tag{3}$$

where, x represnts range of intensity, n_i represents frequency of intensity i, n represents total frequencies.

After performing the pre-processing of images, the regions of interest are detected and it contains the components like eyes, nose and mouth which are expression concentration regions for feature extraction.

3.2 Feature Extraction

Feature extraction is essential for classifying facial expressions and plays a crucial role in improving the accuracy of the classifier [31]. It is required to detect the most significant features which highly contribute to the facial expression representation. In this research, we have extracted the prominent features from the pre-processed image using LBP and PCA for performing dimensionality reduction to reduce the irrelevant features which may affect the classifier's accuracy. We also focus on texture features based on the gray levels which are very important in emotion classification and the various abundant texture features are obtained using gray-level co-occurrence matrix (GLCM).

3.2.1 Local Binary Pattern (LBP)

The Local Binary Pattern (LBP) [32-34] extracts the additional texture features from the Grayscale images. In LBP construction each pixel is considered as centre pixel and its corresponding 3 x 3 neighbours are used to calculate its two-dimensional output based on the threshold value. To convert the grayscale values of an image into a binary patch image, the intensity of each pixel is compared with its 8 surrounded neighbours. If the centre pixel's intensity value is greater than its neighbour then the neighbour pixel value is set to zero or it is set as one for converting to the binary value as shown in Figure 3.

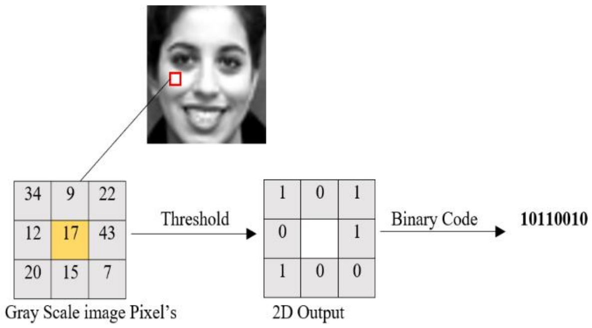

Figure 3. Original LBP operation

The resultant binary value is converted to its corresponding decimal value for obtaining the 2D LBP output it is depicted in Figure 4.

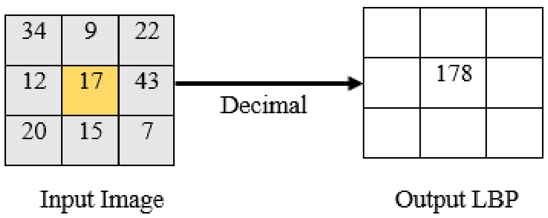

Figure 4. LBP computation for the highlighted pixel

The LBP operator [35] is calculated for the given pixel (x_c, y_c) as follows,

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n s(i_n - i_c) . \quad \textbf{(4)}$$

where, c represents centre pixel and the value of n varies from 0 to 7 for each of its neighbour. The variables i_n and i_c are gray scale value at neighbour pixel (n) and gray scale value at centre pixel (c). S(k) is 1 if k ≥ 0 and 0 otherwise. The sample images from ck+ dataset is converted into LBP images as shown in Figure 5.

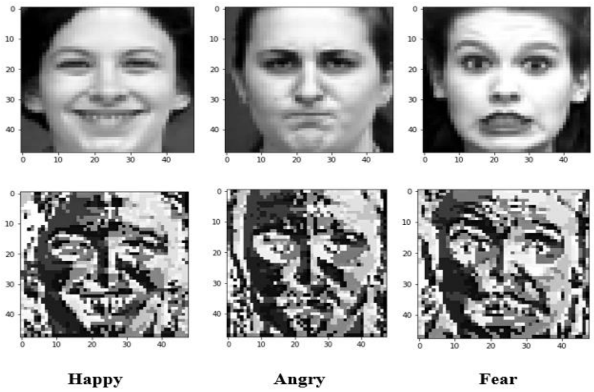

Figure 5. LBP feature for different emotions

3.2.2 Principle Component Analysis (PCA)

PCA is a statistical method that extracts the important facial features from the image without any information loss. It reduces the high dimensional features into low dimensions and the entire facial image is considered as a set of optimal Eigenvectors or principle components. The two-dimensional images are converted as one dimensional vector by combining each row or column. Let's consider the training images in the dataset as M and the training set as X. where X=[x_1, x_2,…, x_M] and by subtracting the mean image from the eigenvector the images are mean centred and the mean image K is defined as follows,

$$K = \frac{1}{M} \sum_{i=1}^{M} x_i . \quad \textbf{(5)}$$

Here x_i defines the rows and columns of a vector and the covariance matrix M_c is defined as follows,

$$M_c = YY^T . \quad \textbf{(6)}$$

For the eigen vector matrix M_c, $Y = \{\varphi_1, \varphi_2, ...\varphi_n\}$ and ϕ_i is the difference between mean image and vector that is represented as follows, $\phi_i = x_i - K$

The dimensionality of the vector is further reduced by,

$$M_c = Y^T Y . \quad \textbf{(7)}$$

3.2.3 Texture Features

In general, various objects have enormous texture features on their faces. It provides relevant information in the spatial arrangement of intensity in an image. It is predominant to learn the textures in the field of image processing for classification. In this work, we extract nine texture features [36] such as energy, entropy, correlation, contrast, homogeneity, variance, mean, dissimilarity, and maximum probability using gray-level co-occurrence matrix (GLCM) for classifying the facial expression images.

Energy:

The Angular Second Moment (ASM) is energy that measures degree of pixel pair repetitions and it can acquire uniformity of gray level for each image and roughness of texture. If all the values of matrix are same then the value of energy is minimum and if all the GLCM values are different then the energy value is maximum. It can be defined by,

$$ASM = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(p(i.j))^2 . \quad \textbf{(8)}$$

Where, i and j are position coordinates of the pixel and N is the dimension of the number of gray levels

Entropy:

The entropy feature computes the degree of randomness which is used to characterize the texture of the image and also measures the degree of uniformity between pixels within the image. The value of entropy is larger with a high gray level values. It can be presented as,

$$Entropy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\log(p(i,j)) . \quad \textbf{(9)}$$

Correlation:

The correlation feature of the image measures how a pixel is correlated with its neighbouring pixels over the entire image. When the matrix elements’ values are same, the correlation value is maximum and if the difference between the values of matrix is very high then the value of correlation is minimum. It can be represented as,

$$COR = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j} . \quad \textbf{(10)}$$

Where, μ_i represents mean values of rows. μ_j represents mean values of columns. σ_i and σ_j are standard deviations of rows and columns respectively.

Contrast:

Contrast feature is a measurement of pixel intensities and its neighbour over the entire image. It also computes the local variants in the image. It is determined by difference in brightness and color of each pixel with other pixels in the same window. It can be written as,

$$Contrast = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|i-j|^2 p(i,j) . \quad \textbf{(11)}$$

Homogeneity:

Homogeneity feature is commonly known as Inverse Different Moment (IDM) and it can reflect roughness of image textures. It decreases as contrast of image increases with constant energy. It can be computed by,

$$IDM = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{p(i,j)}{1+(i-j)^2} . \quad \textbf{(12)}$$

Variance:

Variance feature of the image is calculated by dispersion of the gray values total dissemination of the image and it can be measured by,

$$Variance = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)(i-\mu_i)^2 . \quad \textbf{(13)}$$

Mean:

Mean feature of image is sum average of the gray values total dispersion of the image and it can be calculated by,

$$Mean = \sum_{i=2}^{2N} i p_{x+y}(i+j) , \quad \textbf{(14)}$$

where, $p_{x+y}(i+j)$ is the joint probability density in x and y directions.

Dissimilarity:

The dissimilarity feature of the image computes the distance between the pairs of objects in the Region of Interest and measures mean difference between gray levels in distribution of image. It can be formulated by,

$$Dissimilarity = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|i-j|p(i,j) . \quad \textbf{(15)}$$

Maximum Probability:

The maximum probability computes the maximum likelihood of pixels of the interest and it can be defined by,

$$Maxprobability = \max . p(i,j)\ forall(i,j) . \quad \textbf{(16)}$$

3.3 Multi-feature Integrated Concurrent Neural Network (MICNN)

In this research work, we integrate the three kinds of neural networks namely sequential convolutional neural network (SCNN), residual dense network (RDN), and attention residual learning network (ARLN) to construct the concurrent neural network for improving the accuracy of the facial emotion classification. The proposed MICNN combines the features obtained from the above-mentioned networks and produces the fused features which are finally applied to facial emotion classification. The architecture of the proposed concurrent neural network is shown in Figure 6. First, we conduct LBP, PCA features extraction of original face expression images and have extracted 9 GLCM based texture features which are integrated using convolutional layers with 3 x 3 kernel size of 16,32,64,128. Next, the integrated features are considered for further feature extraction by the SCNN, RDN, and ARLN w. Further, two fully connected layers are applied to extract the 256-dimension feature vectors and extract the 128 dimension feature vectors respectively. Finally, the three categories of 128 dimension feature vectors are combined and these features are again aggregated by a fully connected layer with 128 dimensions. With these integrated features, the softmax layer is used to recognize facial expressions accurately.

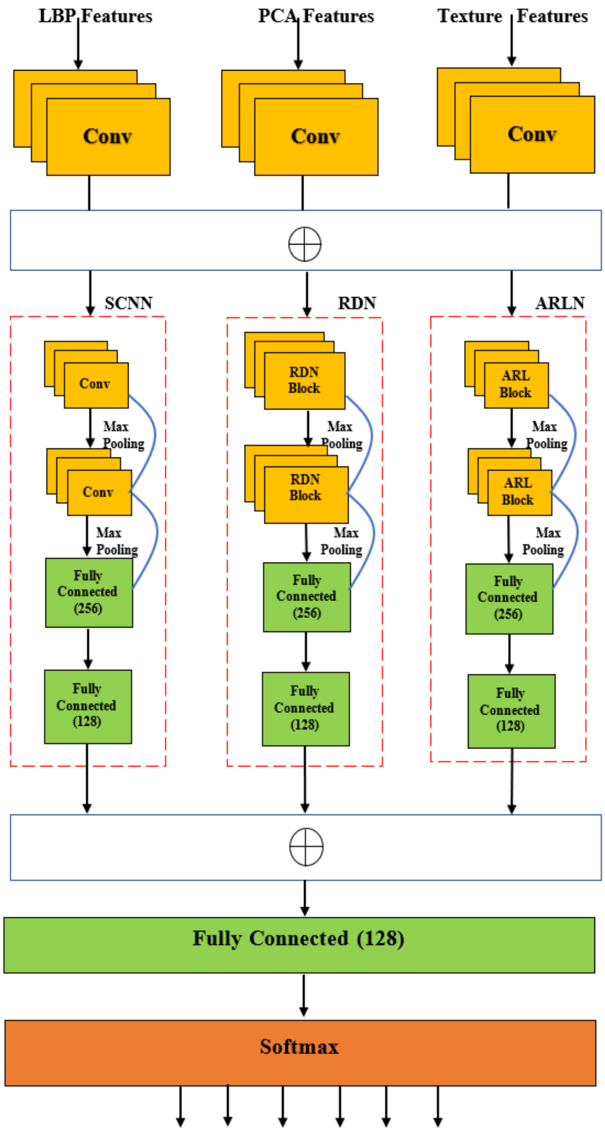

Figure 6. Architecture of proposed MICNN model

3.3.1 Sequential Convolutional Neural Network (SCNN)

The sequential convolutional neural network is the first section of the proposed concurrent neural network and it is the basic convolutional neural network that consists of input, convolutional, pooling, and dense layers. Convolutional operation is conducted by the convolutional layer, overfitting problem is solved by pooling layer and the extracted features are aggregated by dense layer to obtain the result for the final facial expression classification. As shown in Figure 7, the features are aggregated using a convolutional neural network and a fully connected layer for emotion recognition. In the SCNN, max pooling is applied after each convolutional layer for reducing the dimension of the feature map. At last, two fully connected layers with dimensions (256,128) are used to produce the feature map with dimensions of 128.

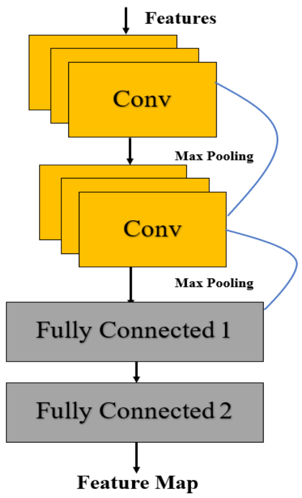

Figure 7. Structure of SCNN block

3.3.2 Residual Dense Network (RDN)

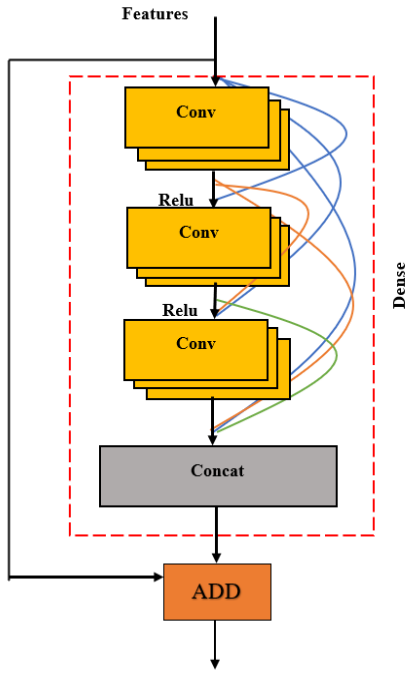

Figure 8. Structure of RDN Block

The Residual Dense Network (RDN) [37] is the next section of our proposed concurrent neural network where the convolution layer is replaced by Residual Dense Network (RDN) which finds the rich prominent local features via densely connected layers of convolution kernels and also obtains higher performance over the original residual network. Each RDN contains local feature fusion, densely connected layers, and local residual learning and it is directly connected from the state of previous RDN to all layers of present RDN shown in Figure 8 which leads to a contiguous memory mechanism. The extracted features from the RDN block are combined using two fully connected layers with dimensions 256 and 128 respectively to produce 128 dimensions of the feature vector. After each residual dense block, the max pooling operation is performed for avoiding the overfitting problem and also reduces the dimensionality of the generated feature map. The feature map of i^{th} convolutional layer of d^{th} RDN block can be represented as,

$$F_{d,i} = \sigma(W_{d,i}[F_{d-1}, F_{d,1}, \ldots\ldots F_{d,i-1}]) , \tag{17}$$

where, σ is ReLU, F_{d-1} and F_d are the input and output respectively and both have H0 feature maps, $W_{d,i}$ is corresponding weights of each layer $F_{d-1}, F_{d,1}, \ldots\ldots F_{d,i-1}$ refers to the feature maps concatenation from previous blocks. Assume $F_{d,i}$ consists of H feature maps then the final constructed feature map is H0+(i-1)*H.

3.3.3 Attention Residual Learning Network (ARLN)

The Attention Residual Learning Network (ARLN) [38] is the final part of our proposed concurrent neural network. In comparison with SCNN, residual learning is adopted to identify the degradation problem and an attention mechanism is employed to raise the discriminative expression representation ability. The convolution operation of CNN [39] is replaced by Attention Residual Learning where it merges both residual and attention learning mechanisms for obtaining the target features present in the image. Further, the features are combined using two fully connected layers for classifying the facial expressions in the image. The attention residual feature map is shown in Figure 9 and it is defined as follows,

$$y = a + f(a) + \beta.\delta[f(a)].a , \tag{18}$$

where, $f(a)$ is residual feature map, $\beta.\delta[f(a)].a$ is attention feature map, a is input and y is feature map, β is learnable weighting factor.

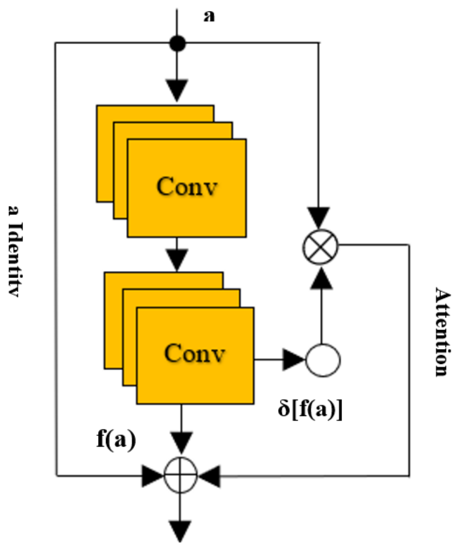

Figure 9. Architecture of ARLN model

4. Experiments of Facial Expression Recognition

Our proposed Multi-feature Integrated Concurrent Neural Network (MICNN) is constructed by combining the SCNN, RDN and ARLN. In the experiment, we have made training for each network at the earliest and then trained the concurrent neural network to demonstrate the superiority of the proposed model. The performance of the proposed network is analyzed with various cross-validation and compared with different approaches on a standard benchmark dataset.

4.1 Experimental Data

To assess the extensive performance of the proposed MICNN model, the experiments are conducted over two standard benchmark datasets namely Extended Cohn-Kanade (CK+) and FER 2013.

CK + Dataset: The Extended Cohn-Kanade (CK+) dataset contains 981 images of 123 different persons. It is mainly classified under seven different emotion categories such as anger, contempt, disgust, fear, happiness, sadness, and surprise. All the images are captured in the laboratory environment and available in grayscale with the pixel size of 640x490 or 640x480. Under each of the emotion categories, it contains 135, 54, 177, 75, 207, 84, and 249 images respectively and the sample images are shown in Figure 10.

Figure 10. Sample images from CK+ dataset

FER 2013 Dataset: It is an open-source dataset that is available in the form of a .csv file as well as in image format which contains a total of 35,887 images of seven different emotion categories. FER 2013 categorized 28709 images for the purpose of training and 3589 images for testing the model. All the available images are in the grayscale format with the standard size of 48 x 48 and all these images are grouped under seven different emotion categories such as anger, disgust, fear, happy, sad, surprise, and normal. The training dataset contains 3995, 436, 4097, 7215, 4830, 3171, and 4965 images for each of the above emotion categories respectively. Sample images from FER 2013 dataset are shown in Figure 11.

Figure 11. Sample images from FER2013 dataset

4.2 Experimental Settings and Training Details

Regarding facial expression classification [40], uniformly we divide the input images into the size of 48 and expand the various types of facial emotions with appropriate approaches to realize the balance in the number of seven categories of facial expressions. For labeling each class, one-hot encoding is conducted to make label values such as 0,1,2,3,4,5,6 for happiness, anger, fear, sadness, disgust, neutral, and surprise respectively which conform to the standards of softmax classifier. After every convolution layer in the network, the ReLU activation function is applied to learn faster and perform better. A python-based deep learning open source library known as Keras [41] is used for implementing the proposed concurrent neural network and it runs on the TensorFlow machine learning platform. Our model is trained with an ADAM optimizer to minimize the loss function and applied sparse categorical cross-entropy loss function for measuring the fluctuation between target and output of the proposed network for handling the multi-class label problem by adjusting the learning rate dynamically with an initial rate of 0.0001. The proposed network is trained with 1000 epochs for identifying the facial expressions. All the networks are trained with the same experimental settings for evaluation. The experiments of the proposed research work are conducted on a PC with a 2.40GHz 11th Gen Intel(R) Core(TM) i5-1135G7 CPU, 8G RAM, and an NVIDIA Computational Accelerator support up to 5120 Cores, 32 GB GPU memory. K-fold cross-validation is applied to evaluate the potential of the proposed algorithm.

4.3 Performance Metrics

We evaluate our proposed model using standard performance metrics known as precision, recall, F1-score and accuracy to classify the facial expressions of human face. These metrics are used to calculate the accuracy of the facial recognition system.

Precision measures the number of facial expressions correctly classified among classified ones and it is defining as follows

$$Precision = \frac{TP}{TP + FP}. \quad \textbf{(19)}$$

Recall identifies the number of facial expressions correctly classified among total number of expected expressions and it can be represented by,

$$Recall = \frac{TP}{TP + FN}. \quad \textbf{(20)}$$

F1-score computes the weighted average of precision and recall and indirectly it highlights missed expected recognized expressions. It is important metrics to determine the overall performance of the proposed model. it can be computed by,

$$F1 - score = \frac{2 * precision * recall}{precision + recall}. \quad \textbf{(21)}$$

Confusion matrix is used to calculate the classification error rate and visualize the performance of the classifier. The diagonal elements of the confusion matrix specify the number of accurately classified expressions for each label and the off-diagonal of the matrix elements specify the classification error or misclassification. The accuracy of the classification can be evaluated by,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad \textbf{(22)}$$

TP=True Positive; TN=True Negative; FP=False Positive; FN=False Negative.

4.4 Results and Discussion

The performance of the proposed model is evaluated with various k-fold cross validation to analyse the algorithm and also the results of the proposed work is compared with different state-of-the-art methods on standard benchmark datasets.

4.4.1 Performance Evaluation with Cross Validation

In the research work, the proposed MICNN model is assessed by applying the k-fold cross-validation method where the dataset is split into k parts and we use the value of k is 5. In this validation for each test, 4 of them are used as training sample sets, and the remaining one is used as a testing sample dataset. We have conducted experiments on two standard benchmark datasets namely the CK+ dataset and FER 2013 dataset to estimate the proposed model.

The experiment is repeated five times and the mean value is considered as the final data value of the result. As a lot of training data is needed to train the model, the original images in datasets are transformed symmetrically to double the number of training images for learning. With the enriched data, the experiments are performed on the benchmark datasets and the results are illustrated in Table 1, where k represents the kth part in the k-fold cross-validation.

In order to highlight the capability of our model Multi-feature Integrated Concurrent Neural Network (MICNN), we applied k-fold cross-validation for four various kinds of network such as SCNN, RDN, ARLN, and our model MICNN on two benchmark datasets. As per analysis made in Table 1, it is evident that our proposed model attains an average accuracy of 98.12 % than the other networks on the CK+ dataset and our model achieves an average accuracy of 84.80 % than other networks on the FER 2013 dataset.

Table 1. Comparison with other networks on benchmark dataset

k-fold	CK+ Dataset				FER 2013 Dataset			
	SCNN	RDN	ARLN	**MICNN**	SCNN	RDN	ARLN	**MICNN**
k=1 (%)	89.9	89.5	92.0	97.1	74.5	75.1	77.1	84.5
k=2 (%)	92.1	92.1	95.5	99.6	75.1	79.1	79.9	86.1
k=3 (%)	89.0	93.4	94.5	97.0	77.2	78.9	78.4	85.1
k=4 (%)	90.1	94.8	92.9	99.5	76.1	77.8	79.9	84.2
k=5 (%)	89.1	92.5	91.2	97.1	73.2	74.1	75.3	83.9
Average(%)	90.0	92.5	93.2	**98.1**	75.2	77.0	78.1	**84.8**

Experiments conducted over CK+, FER 2013 benchmark datasets are analyzed by confusion matrix which gives the average accuracies of k fold (k=5) cross-validation shown in Figure 12 and Figure 13 respectively. In the CK+ dataset, our MICNN model has attained an accuracy rate of 100% for happiness and more than 95% for the emotions such as anger, disgust, fear, contempt, sadness, and surprise. In FER 2013 dataset, the MICNN model gives the highest recognition result of 87 % for happiness and sadness. The lowest recognition result is 83% for angry and normal. All other emotions such as disgust, fear, and surprise have reached the accuracy rate of 86%, 84 %, and 84% respectively.

The loss function and accuracy of the proposed model is plotted for the different number of iterations to acquire efficient experimental results on a standard benchmark dataset. As shown in Figure 14 and Figure 16 the loss function of the proposed model is decreased with an increasing number of epochs and it is slowly decreased while it reached iteration 200.

It clearly shows that our model gives better classification results to identify the appropriate emotions. The training and testing accuracy of the MICNN model is depicted in Figure 15 and Figure 17. Accuracy increases suddenly up to the 200th iteration and then improved gradually. After the 650th iteration, the accuracy remains stable without any further changes. The mean accuracy reached 98% and 84% for CK + dataset and FER2013 dataset respectively.

	Anger	Disgust	Fear	Happy	Contempt	Sadness	Surprise
Angry	.97	0	.01	0	0	.02	0
Disgust	0	.98	.01	0	.01	0	0
Fear	0	0	.99	0	0	.01	0
Happy	0	0	0	.1	0	0	0
Contempt	.01	0	.01	0	.97	.01	0
Sadness	0	.01	0	0	.01	.98	0
Surprise	0	0	0	.01	.01	0	.98

Figure 12. Confusion matrix on CK + dataset

	Angry	Disgust	Fear	Happy	Normal	Sadness	Surprise
Angry	.83	.9	.4	0	.2	.2	0
Disgust	0	.86	.8	0	.5	.1	0
Fear	.8	.2	.84	0	.6	0	0
Happy	0	.5	0	.87	0	0	.8
Normal	0	0	.5	.12	.83	0	0
Sadness	.2	.4	.7	0	0	.87	0
Surprise	0	0	.1	.9	.6	0	.84

Figure 13. Confusion matrix on FER2013 dataset

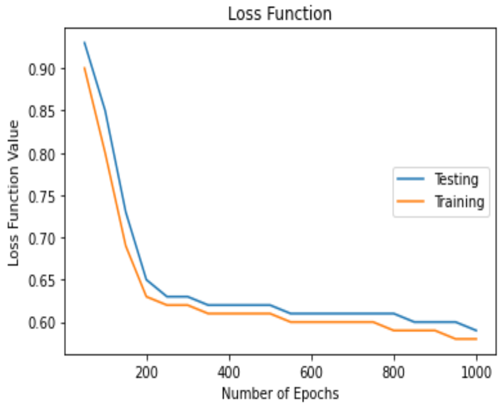

Figure 14. Loss function of MICNN on CK+

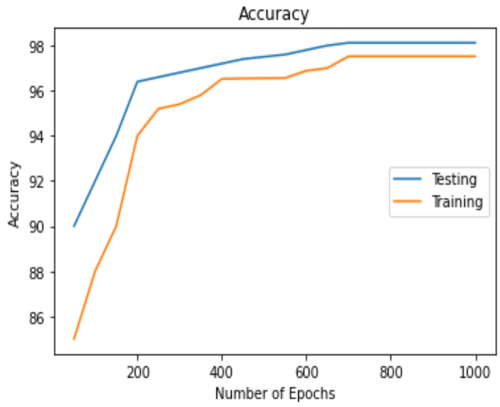

Figure 15. Accuracy of MICNN on CK+

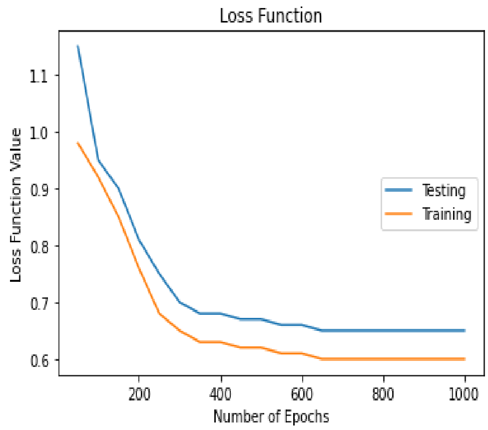

Figure 16. Loss function of MICNN on FER2013

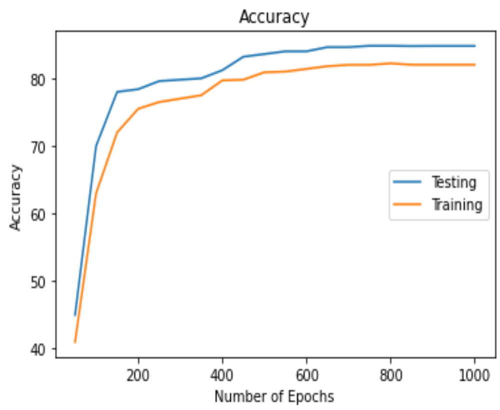

Figure 17. Accuracy of MICNN on FER2013

4.4.2 Comparison with Various Approaches on Benchmark Dataset

The result of the proposed work is analyzed with the most advanced existing methods as depicted in Table 2 and Table 3, which mainly focus on detecting the discriminative expression features by numerous methods and this is also one of the inspirations of the research work.

To learn better emotional features of face images, we proposed Multi-feature Integrated Concurrent Neural Network (MICNN) for classifying the facial expressions of humans. From the experimental results present in Table 2, our multi-feature integrated model provides competitive results than other methods with an accuracy of 98.12% and reached the F1-score value of 0.9825 on the CK+ dataset.

Table 2. Comparison of experimental results for several methods on CK+ dataset in terms of accuracy and F1-Score

Source	Methods	Accuracy (%)	F1-Score
M. Wu, et al. [27]	WACNN	91.00	0.9216
Yang, et al. [43]	DeRL	96.54	0.9730
G. Lu, et al. [44]	Seven CNN	84.56	0.8533
Fan, et al. [45]	LZMHI	88.30	0.8753
B. F. Wu, et al. [46]	WCRAFM	89.84	0.9010

In addition, the proposed model is evaluated on FER 2013 dataset and the results of the state-of-art method are compared with the research work shown in Table 3, it can be seen that the recognized accuracy of the MICNN attained 84.80 % and F1-score value achieved 0.8415, depicting enormous advantageous of the proposed model in multi-class tasks for facial expression classification.

Table 3. Comparison of experimental results for several methods on FER2013 dataset in terms of accuracy and F1-Score

Source	Methods	Accuracy (%)	F1- Score
Zhang et al. [5]	Deep model	75.52	0.751
Liang et al. [13]	MSAU-Net	79.12	0.783
Shi et al. [20]	FCNN	83.86	0.8275
Zhang et al. [26]	Unified Deep	82.60	0.8350
Pramerdorfer et al. [42]	CNN	74.15	0.752
Proposed Model	MICNN	84.80	0.8415

5 Conclusion

The Multi-feature Integrated Concurrent Neural Network (MICNN) is proposed to enhance the classification rate of facial emotions by getting facial expression information from the extracted discriminative expression features. The proposal pre-processes the facial expression images for identifying the human faces from the images then Local Binary Pattern (LBP) and Principle Component Analysis (PCA) are applied to find the low-level features from the pre-processed face images. Further, these features are fused with texture features identified by GLCM to enrich the feature expressions for classification. Besides, the three common neural networks namely sequential convolutional neural network (SCNN), residual dense network (RDN), and attention residual learning network (ARLN) are integrated for extracting the more relevant features which are again aggregated to recognize the facial expressions. Then, classification of the facial expressions is conducted using the softmax layer. Experiments on CK+ and FER 2013 are accomplished by applying k-fold cross-validation to assess the ability of the proposed algorithm and are also analyzed with various approaches. Results prove that the accuracy of the facial expression classification is increased by making parallel settings of three different networks for effectively integrating the features from the images. In comparison with other singular network, the accuracy can be improved but the quantity of the parameters increases the complexity of the network. However, the parallel operations reduce the running time of the network which can remedy the problem of excessive complexity. Our proposed work can upgrade the development of the human-machine intercommunication systems and also promotes the aggregation of natural and body languages effectively.

Acknowledgement

The authors acknowledge and thank the Department of Science and Technology (Government of India) for sanctioning the research grant (Ref. No.SR/FST/COLLEGE-096/2017 dated 16.01.2018) under Fund for Improvement of S&T Infrastructure (FIST) program for completing this work.

References

[1] H.-A. Li, J. Fan, J. Zhang, Z. Li, D. He, M. Si, Y. Zhang, Facial Image Segmentation Based on Gabor Filter, *Mathematical Problems in Engineering*, Vol. 2021, Article No. 6620742, February, 2021.

[2] W. Lu, M. Yang, Face Detection Based on Viola-Jones Algorithm Applying Composite Features, *International Conference on Robots & Intelligent System (ICRIS)*, Haikou, China, 2019, pp. 82-85.

[3] A. T. Lopes, E. de Aguiar, A. F. De Souza, 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, January, 2017.

[4] Liu W, Zhou L, Chen J, Face Recognition Based on Lightweight Convolutional Neural Networks, *Information*, Vol. 12, No. 5, pp. 1-18, March 2021.

[5] Z. Zhang, P. Luo, C.-C. Loy, X. Tang, Learning social relation traits from face images, *IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 3631-3639.

[6] Z. Xi, Y. Niu, J. Chen, X. Kan, H. Liu, Facial Expression Recognition of Industrial Internet of Things by Parallel Neural Networks Combining Texture Features, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 4, pp. 2784-2793, April, 2021.

[7] A. Deopa, A. Sinha, A. Prakash, R. K. Sinha, Facial Expression Recognition using Convolutional Neural Network and SoftMax function on Captured Images, *International Conference on Communication and Electronics Systems (ICCES)*, 2019, pp. 273-279.

[8] B. Niu, Z. Gao, B. Guo, Facial Expression Recognition with LBP and ORB Features, *Computational Intelligence and Neuroscience*, Vol. 2021, Article No. 8828245, January, 2021.

[9] H. Yao, X. Zhang, X. Zhou, S. Liu, Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification, *Cancers*, Vol. 11, No. 12, pp. 1-14, December, 2019.

[10] S. Yang, G. Peng, D-PCN: Parallel Convolutional Networks for Image Recognition via a Discriminator, *Computer Vision and Pattern Recognition*, March, 2018. https://arxiv.org/abs/1711.04237

[11] N. K. Chowdhury, M. M. Rahman, M. A. Kabir, PDCOVIDNet: A parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images, *Health Information Science and Systems*, Vol. 8, No. 1, Article No. 27, pp. 1-14, December, 2020.

[12] L. Zhong, Q. Liu, P. Yang, J. Huang, D. N. Metaxas, Learning multiscale active facial patches for expression analysis, *IEEE Transactions on Cybernetics*, Vol. 45, No. 8, pp. 1499-1510, August 2015.

[13] L. Liang, C. Lang, Y. Li, S. Feng, J. Zhao, Fine-Grained Facial Expression Recognition in the Wild, *IEEE Transactions on Information Forensics and Security*, Vol. 16, pp. 482-494, 2021.

[14] N. Farajzadeh, M. Hashemzadeh, Exemplar-based facial expression recognition, *Information Sciences* Vol. 460-461, pp. 318-330, September, 2018.

[15] S. F. Wang, S. Wu, G. Z. Peng, Q. Ji, Capturing feature and label relations simultaneously for multiple facial action unit recognition, *IEEE Transactions on Affective Computing*, Vol. 10, No. 3, pp. 348-359, July-September, 2019.

[16] S. L. Happy, A. Routray, Automatic facial expression recognition using features of salient facial patches, *IEEE Transactions on Affective Computing*, Vol. 6, No. 1, pp. 1-12, January-March, 2015.

[17] G. Ali, A. Ali, F. Ali, U. Draz, F. Majeed, S. Yasin, T. Ali, N. Haider, Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis, *IEEE Access*, Vol. 8, pp. 134950-134963, July, 2020.

[18] N. Zhou, R. Liang, W. Shi, A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection, *IEEE Access*, Vol. 9, pp. 5573-5584, December, 2020.

[19] C. Chen, X.-Y. Wang, J. Chen, Q. Liu, S.-L. Peng, An Active Security System Based on AR Smart Classes and Face Recognition Technology, *Journal of Internet Technology*, Vol. 23, No. 2, pp. 245-253, March, 2022.

[20] M. Shi, L. Xu, X. Chen, A Novel Facial Expression Intelligent Recognition Method Using Improved Convolutional Neural Network, *IEEE Access*, Vol. 8, pp. 57606-57614, March, 2020.

[21] K. Zheng, D. Yang, J. Liu, J. Cui, Recognition of Teachers' Facial Expression Intensity Based on Convolutional Neural Network and Attention Mechanism, *IEEE Access*, Vol. 8, pp. 226437-226444, December, 2020.

[22] S. Umer, R. K. Rout, C. Pero, M. Nappi, Facial expression recognition with trade-offs between data augmentation and deep learning features, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 13, No. 2, pp. 721-735, February, 2022.

[23] Y. Yan, Y. Huang, S. Chen, C. Shen, H. Wang, Joint Deep Learning of Facial Expression Synthesis and Recognition, *IEEE Transactions on Multimedia*, Vol. 22, No. 11, pp. 2792-2807, November, 2020.

[24] C. Li, M. Liang, W. Song, K. Xiao, A Multi-Scale Parallel Convolutional Neural Network Based Intelligent Human Identification Using Face Information, *Journal of Information Processing Systems*, Vol. 14, No. 6, pp. 1494-1507, December, 2018.

[25] Y. Tang, X. Zhang, X. Hu, S. Wang, H. Wang, Facial Expression Recognition Using Frequency Neural Network, *IEEE Transactions on Image Processing*, Vol. 30, pp. 444-457, 2021.

[26] F. Zhang, T. Zhang, Q. Mao, C. Xu, A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment, *IEEE Transactions on Image Processing*, Vol. 29, pp. 6574-6589, 2020.

[27] M. Wu, W. Su, L. Chen, Z. Liu, W. Cao, K. Hirota, Weight-Adapted Convolution Neural Network for Facial Expression Recognition in Human–Robot Interaction, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 51, No. 3, pp. 1473-1484, March, 2021.

[28] G. Kavitha, I. L. Aroquiaraj, Improving Face Recognition Rate with Preprocessing and Enhanced Viola Jones Face Detection Algorithm, *International Journal of Scientific & Technology Research*, Vol. 8, No. 10, pp. 3471-3475, October, 2019.

[29] P. Viola, M. J. Jones, Robust real-time face detection, *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137-154, May, 2004.

[30] S. Yuan, M. Abe, A. Taguchi, M. Kawamata, High accuracy bicubic interpolation using image local features, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E90-A, No. 8, pp. 1611-1615, August, 2007.

[31] S.-G. Yu, S.-E. Kim, K. H. Suh, E. C. Lee, Effect of Facial Shape Information Reflected on Learned Features in Face Spoofing Detection, *Journal of Internet Technology*, Vol. 23, No. 3, pp. 517-525, May, 2022.

[32] Y. Kortli, M. Jridi, A. Alfalou, M. Atri, A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF for security and face recognition, *Pattern Recognition and Tracking*, pp. 1-7, April, 2018.

[33] T. J. Alhindi, S. Kalra, K. H. Ng, A. Afrin, H. R. Tizhoosh, Comparing LBP, HOG and Deep Features for Classification of Histopathology Images, *International Joint Conference on Neural Networks* (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-7.

[34] A. S. Mubarak, S. Serte, F. Al-Turjman, Z. Ameen, M. Ozsoz, Local binary pattern and deep learning feature extraction fusion for COVID-19 detection on computed tomography images, *Expert Systems*, Vol. 39, No. 3, Article No. e12842, March, 2022.

[35] T. Ojala, M. Pietikainen, D. Harwood, A Comparative Study of Texture Measures with Classification Based on Featured Distribution, *Pattern Recognition*, Vol. 29, No. 1, pp. 51-59, January, 1996.

[36] P. K. Mall, P. K. Singh, D. Yadav, GLCM Based Feature Extraction and Medical X-RAY Image Classification using Machine Learning Techniques, *IEEE Conference on Information and Communication Technology*, Allahabad, India, 2019, pp. 1-6.

[37] C. Zhou, S. Zhou, J. Xing, J. Song, Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network, *IEEE Access*, Vol. 9, pp. 28822-28831, February, 2021.

[38] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention Residual Learning for Skin Lesion Classification, *IEEE Transactions on Medical Imaging*, Vol. 38, No. 9, pp. 2092-2103, September, 2019.

[39] Y. Djenouri, G. Srivastava, J. C. W. Lin, Fast and Accurate Convolution Neural Network for Detecting Manufacturing Data, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 4, pp. 2947-2955, April, 2021.

[40] F. Nan, W. Jing, F. Tian, J. Zhang, K.-M. Chao, Z. Hong, Q. Zheng, Feature super-resolution based Facial Expression Recognition for multi-scale low-resolution images, *Knowledge-Based Systems*, Vol. 236, Article No. 107678, January, 2022.

[41] Keras, https://keras.io/

[42] C. Pramerdorfer, M. Kampel, Facial expression recognition using convolutional neural networks: State of the art, *Computer Vision and Pattern Recognition*, December, 2016. https://arxiv.org/abs/1612.02903

[43] H. Yang, U. Ciftci, L. Yin, Facial expression recognition by de-expression residue learning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 2168-2177.

[44] G. Lu, J. He, J. Yan, H. Li, Convolutional neural network for facial expression recognition, *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)*, Vol. 36, No. 1, pp. 16-22, February, 2016.

[45] X. J. Fan, T. Tjahjadi, A dynamic framework based on local Zernike moment and motion history image for facial expression recognition, *Pattern Recognition*, Vol. 64, pp. 399-406, April, 2017.

[46] B.-F. Wu, C.-H. Lin, Adaptive feature mapping for customizing deep learning based facial expression recognition model, *IEEE Access*, Vol. 6, pp. 12451-12461, February, 2018.

Biographies

C. R. Dhivyaa received her Ph.D. degree from Anna University, Chennai in 2019. She is working as Assistant Professor in the department of Computer Science and Engineering at Kongu Engineering College, Erode. Her current research interests include computer vision, image processing, medical imaging, machine learning, and pattern recognition. She has published many research papers in various national, international journals and conferences. She has very depth knowledge of her research areas.

K. Nithya is currently working as Assistant Professor in the department of Computer Science and Engineering at Kongu Engineering College, Erode. Her current research interests include sentiment analysis, image processing and pattern recognition. She is pursuing her Ph.D. degree in Information and Communication Engineering, Anna University, Chennai.

K. Karthika is working as Assistant Professor in the Department of Computer Science and Engineering, in Kongu Engineering College, Erode. Her research interests include Sentiment Analysis, Machine learning, social media analytics, Predictive Analysis.

S. Mythili is working as an Assistant Professor in Bannari Amman Institute of Technology. Her area of interest is wireless communication it motivates her to pursue PhD in wireless networks under Anna University, Chennai. Her keen area of interest in research thrives her to publish many articles in various journals.