

Facial Image based Emotion Recognition System using Neural Network

Parnal Dudul¹, Prof. S. M. Tayade², Prof. Ajay Talele³

M. Tech Student, Department of E & TC Engineering, V. I. T., Pune, India¹

Assistant Professor, Department of E & TC Engineering, V. I. T., Pune, India^{2,3}

Abstract: Facial emotion plays a vital role for human interactive communication and also used in numerous real applications. Facial expression identification from frontal still images has in recent times become a hopeful investigation area. Their applications include human-computer interface, human emotion examination robot control, driver state surveillance and medical fields. This paper aims to develop emotion classification scheme to identify seven dissimilar facial emotions, such as surprise, sad, neutral, happy, fear, disgust and anger by using JAFFE database. Two different approaches of feature selection and extraction have been used for generation of optimal feature vector. LBP and 2D-DCT coefficients are employed in addition to image statistics, texture and entropy parameters. In order to reduce the high dimensionality of the inputs, the principal component analysis has been used and significant reduction in the input-dimensionality has been achieved. The single hidden layer feed forward neural network has been used as a classifier in order to classify different emotions from frontal facial images. Three learning algorithms such as resilient backpropagation, scaled conjugate gradient and gradient descent algorithm with momentum and adaptive learning rate have been compared. It has been observed that our meticulously designed LBP based hybrid feature vector and a single hidden layer neural network containing only 70 neurons in the hidden layer trained with gradient descent algorithm with momentum and adaptive learning rate delivers the maximum average overall classification accuracy of 97.2%, which has not been reported so far in the literature for the said database. The proposed neural network is very compact, as it is comprised of only 5,607 connection weights including biases.

Keywords: Facial Emotions, 2D DCT, LBP, PCA, neural network, Classifier.

I. INTRODUCTION

In humans, emotions play an extremely important role in their lives. They determine how we think, how we behave and how we communicate with others. With this information, it is possible to think, that the new future generation of machines would have some skills to understand human emotions and generate synthetic emotions similar to humans. The new machines could have some kind of functionalities to respond more flexibly, predict and adjusting to what humans want.

Expression detection is useful as a non-invasive method of lie detection and behavior prediction. However, these facial expressions may be difficult to detect to the untrained eye. A facial expression recognition system needs to solve the following problems: detection and location of faces in a cluttered scene, facial feature extraction, and facial expression classification. It is very important to understand the presence of mind using face expression such as the situation of mind can be read by their mouth, eyes, eyebrows etc. No wonder automatic face expression recognition has become an area of immense interest within the computer science, psychology, medicine and human-computer interaction research communities.

Emotion recognition is a promising area of development and research. The voice interactive systems can adapt as per the detected input emotion. This could lead to more realistic interactions between system and the user. Expression is the most important mode of non-verbal communication between people. Recently, the facial expression recognition (FER) technology attracts more and more attention with people's growing interest in expression information. Facial expression carries crucial information about the mental, emotional and even physical states of the conversation. FER has practical significance; it has very broad application prospects, such as user-friendly interface between man and machine, humanistic design of goods, and emotional robot etc. With FER systems, the computer will be able to assess the human expressions depending on their effective state in the same way that human's senses do. The intelligent computers will be able to understand, interpret and respond to human intentions, emotions and moods. The FER system can be applied in different areas of life such security and surveillance, they can predict the offender or criminal's behavior by analyzing the images of their faces that are captured by the control-camcorder. Furthermore, the FER system has been used in communication to make the "answer machine" more interactive with people. FER has attracted



increasing attention in computer vision, pattern recognition, and human-computer interaction research communities. The rest of the paper has been organized as follows. In Section II, we have reported the brief overview of related published research work in the literature. The proposed research methodology has been explained in Section III. Experimental results are demonstrated in Section IV and in the end; Section V highlights the conclusive remarks.

II. LITERATURE SURVEY

A person's face is considered as the mirror of the mind. Facial expressions and the changes in facial expressions provide important information about affective state of the person, his temperament and personality, psychopathological diagnostic information, information related to stress levels, truthfulness etc. [1]. With growing terrorist activities all over the world, detection of potential troublemakers continues to be a major problem. Body language and facial expressions are the best ways to know the personality of a person and the response of a person in various situations. The facial expressions tell us about concealed emotions which can be used to verify if the information provided verbally is true. These expressions representing the emotional state of a person can serve an important role in the field of terrorism control and forensics.

Facial expression analysis can also be used effectively in psychopathology. It may give us information related to the diagnostic information relevant to depression, mania, schizophrenia and other disorders. The information relevant to the patient's response to the treatment could also be monitored with the facial expression analysis [2, 3]. Thus, expression analysis can be effectively used for behavioral studies and medical rehabilitation. Stress detection through facial expression analysis can be useful in caseslike monitoring stress levels in astronauts since other methods may not work in that environment [4]. Thus, there is an increasing need to understand the human facial expressions in a better way and to develop a system to accurately classify the emotions of subjects based on their facial images.

Over the last few years, active research is being done to correlate movements of the face with emotional states. Darwin [5] first published "The Expression of the emotions in Man and Animals" in which he stated the three basic principles related to expressions and gestures in humans, viz., "(a) Principle of Serviceable associated habits, (b) Principle of Antithesis, and (c) Principle of actions due to constitution of nervous system, independently from the will, and independently to a certain extent habit" [5].

Ekman and Friesen [6] developed the Facial Action Coding System (FACS) to measure the facial behavior. In FACS, they used Action Units (AUs) based on the muscular activity that produces momentary changes in the facial expression. The system further classified an expression by correctly identifying the action unit or combination of action units related to a particular expression. A FACS database was created to determine the category or categories in which to fit each facial behavior. This database is available in the form of a FACS manual. Based on the action units, the researcher has to interpret the actual emotion.

Padgett [7], Hara and Kobayashi [8, 9], Zhang [10] and Zhao [11] used neural network approach for expression classification. They classified images into six or seven emotional categories. Padgett et al., [7] trained neural networks from the data of 11 subjects and tested with the data from one subject. The training and testing dataset was interchanged and new networks were trained and tested. A classification accuracy of 86% was achieved in this study. Hara and Kobayashi [8, 9] also used neural networks approach. The training dataset consisted of from data of 15 subjects (90 images) and these networks were tested using data from another 15 subjects. The classification accuracy achieved was 85 %. Zhang et al., [12] used the JAFFE data base which consists of 10 Japanese female subjects. Although an accuracy of 90.1% was achieved; same data was used for training and testing. A 100 % recognition rate was achieved by Zhao et al., [11] who used the Ekman and Friesen database [13], but they used the same data for training and testing. Khan et al [14] used thermal methods to quantify the facial expressions. He could achieve an accuracy of 56 %.

As mentioned above, most of the facial expression analysis systems have been based on FACS and identifying AUs coded from static facial images, and further classifying the expressions based on combinations of these AUs identified. These systems also required sequential video images (from a video) for expression classification. Some of the systems used neural network approach, but these systems either used a very small dataset or the training and testing datasets were not well differentiated. Thus, a method for recognizing expressions needs to be developed for fast, easy, and accurate classification of facial expressions and tested on a larger dataset of facial expressions like the Cohn-Kanade database [15]. A classification system consisting of a committee of neural networks can give better results than a classification system consisting of a single network. Reddy and Buch [16], Das et al., [17] and Reddy et al. [18] observed such an enhanced classification performance in their studies.



Though much progress has been made [19-24], recognizing facial expression with a high accuracy remains difficult due to the subtlety, complexity and variability of facial expression. And it is very difficult to integrate the internal feelings of a person with facial expression; if the person is fraud then his facial expression will be different from his internal feelings.

Past works used different methods for feature extraction. G.Guo et al. [25] adopted Gabor and FSLP (Feature Selection via Linear Programming) for feature extraction process. Their appreciation rate is 91.0% using JAFFE database. W.Yuwen et al. [26] tested fuzzy integral for feature extraction. Their recognition rate is 83.2% in using JAFFE database. Z.Wenming et al. [27] used KCCA (Kernel Canonical Correlation Analysis). Their recognition rate is 77.05% in JAFFE database. C.Zhengdong et al. [28] used WMMC (Weighted Maximum Margin Criterion). Their recognition rate is 65.77%. C.Shan et al. [29] adopted Boosted-LBP. Their recognition rate is 81.0% for JAFFE database and 95.1% for CK database. In this paper, patch based features are accurately extracted from the image to find emotions and established to be vigorous, even in the situation of face misalignment to give the correct recognition rate using JAFFE database.

Urvashi Bakshi, Rohit Singhal [30], have introduced a new technique to recognize human face artificially using DCT, PCA and SOM neural network. Principal component analysis (PCA) is a classical and successful method of dimension reduction. Discrete Cosine Transform (DCT) is a well-known compression technique and Self Organize Map (SOM) act as a classifier and has been used for face space representation.

Wang, Xun; Liu, Xingang; Lu, Lingyun; Shen, Zhixin et al. [31], have proposed a new FER system, which uses the active shape mode (ASM) algorithm to align the faces, then extracts local binary patterns (LBP) features and uses support vector machine (SVM) classifier to predict the facial emotion.

III. PROPOSED METHODOLOGY AND DISCUSSION

A. Proposed Research Methodology

The research problem undertaken in this project is to develop a neural network based classifier system for identification of principal emotions (expressions) from human facial images with respect to standard benchmark database, namely, Japanese female expression database (JAFFE). All these images are frontal still images of the human faces, which will be used to develop an optimal neural network based classifier system with a view to correctly identify and recognize the expressed emotion by a person.

A unique image feature selection and generation scheme has been suggested. Features based on 2D DCT and LBP are augmented with features with regard to image statistics, entropy, texture, homogeneity, etc., so that every image is represented by a unique feature vector. Based on the annotations available and the visual inspection of the frontal still facial image of a person, the exact emotion is recognized, which is the target emotion. The facial image database is transformed to a knowledgebase to be used by a neural network, where a feature vector is followed by the expressed emotion. Primarily, two different feature extraction schemes based on 2D DCT and LBP are compared in order to determine the optimal feature vector of an image. Consequently, an optimal feature vector is designed. In a software environment available in MATLAB 2016a, codes are developed for accomplishing the tasks of feature extraction, development of a classifier based on neural network, scatter plots, etc. In addition, we have used XLSTAT 2014 for principal component analysis. It is a projection method as it projects observations from a p-dimensional space with p variables to a k-dimensional space (where $k < p$) so as to conserve the maximum amount of information (information is measured here through the total variance of the dataset) from the initial dimensions. PCA dimensions are also called axes or Factors. If the information associated with the first 2 or 3 factors represents a sufficient percentage of the total variability of the scatter plot, the observations could be represented on a 2 or 3-dimensional chart, thus making interpretation much easier.

In order to have an understanding of the complexity of the classification problem, it is necessary to inspect visually the two-dimensional input feature space with respect to all emotions. We have considered seven different emotions for Japanese female expression database. When a typical two-dimensional scatterplot is drawn between any one of input features versus another input feature, the crux of the emotion identification problem is to estimate appropriate decision boundaries, so that each distinct emotion is clearly separated out. However, because of the fact that emotion expression ability varies significantly from a person to person and in addition, there is a great deal of variation in expression of emotions with regard to male and female subjects. Consequently, estimation of decision boundaries from the input feature space is not a trivial task because of overlapping noticed between different classes (emotions) and highly nonlinear nature of the decision boundaries. Such decision boundaries are required to be estimated as precisely as possible in order to correctly identify or recognize an emotion from an image. As a matter of fact, numerous scatter plots can be drawn and examined. In order to simplify the matter, we have fixed input feature 1 and input feature 2 for examination and understanding of the complexity of the nature of decision boundaries. The neural network based

classifier is developed to estimate the decision boundaries, so that the all emotions are correctly classified (recognized) and there are no misclassifications or incorrect classifications.

For designing the optimal neural network based classifier, we have employed a network growing approach and we also followed a double loop design strategy. For ensuring simplicity of the neural network and also reduced time complexity and space complexity (hardware, number of connection weights including biases), we have adhered to a single hidden layer feed forward neural network, which is known as *patternnet* in MATLAB. The number of neurons (processing elements) in the hidden layer is increased gradually from 10 to 1000 in the increment of 10. For each value of the number of PEs in the hidden layer, ten different neural network candidates are configured and retrained with different random initialization of connection weights and biases as well as different random data partitioning into training, validation (cross validation) and testing subsets. This is necessary to ensure true learning without any prejudices and biases. Moreover, neural network learning has been made almost independent of any specific calculated choice of data partitioning into training, cross validation and testing datasets. Because of random partitioning of data into the training, validation and testing sets, each sample participates, on an average, an equal number of times to all these three sets. We used 90% of the total samples as Training Dataset, different 5 % of the total samples as cross-validation (validation) dataset and remaining 5% of the total samples as testing dataset. Because of the fact that JAFFE database has 213 different images or samples, 191 samples (exemplars) are used for training the neural network, other 11 samples are used for validation of the trained neural network and remaining 11 samples are used for testing the performance of the trained neural network.

Obviously, the performance of the neural network is always the best on the training dataset, because it is this dataset only which has been presented iteratively to the neural network in order to achieve learning and estimate the optimal values of the connection weights and biases of the neural network. Therefore, examination of the performance of the neural network on validation as well as the testing dataset is more important as neural network never sees these datasets during the process of learning. Cross-validation dataset is used for termination of the neural network training and as training proceeds, at the end of every iteration, it is ensured that both the error on training dataset and that on validation dataset go on decreasing with respect to the number of iterations (epochs). Though, the magnitude of these errors might be significantly different. It is generally noticed that the error on the training dataset is always much lower than that on validation dataset. The magnitude of the error gradient and the number of validation checks are used to terminate the training. The gradient will become very small as the training reaches a minimum of the performance. If the magnitude of the gradient is less than, say, $1e-5$, the training will stop. This limit can be adjusted by setting the parameter *net.trainParam.min_grad*. The number of validation checks represents the number of successive iterations that the validation performance fails to decrease. If this number reaches 6 (the default value), the training will stop.

Our image data is stationary, that is, the statistics for each data dimension follow the same distribution, and then we might want to consider subtracting the mean-value of a feature from each value of the feature. In images, this normalization has the property of removing the average brightness (intensity) of the data point. In many cases, we are not interested in the illumination conditions of the image, but more so in the content; removing the average pixel value per data point makes sense here. Feature standardization refers to (independently) setting each dimension of the data to have zero-mean and unit-variance. This is the most common method for normalization and is generally used widely (e.g., when working with neural networks, feature standardization is often recommended as a pre-processing step). In practice, one achieves this by first computing the mean of each dimension (across the dataset) and subtracts this from each dimension. Next, each dimension is divided by its standard deviation. All input features are normalized with zero mean and unit variance, so that the neural network based classifier models can run faster and better. Fig. 1 shows some sample facial images of JAFFE database.

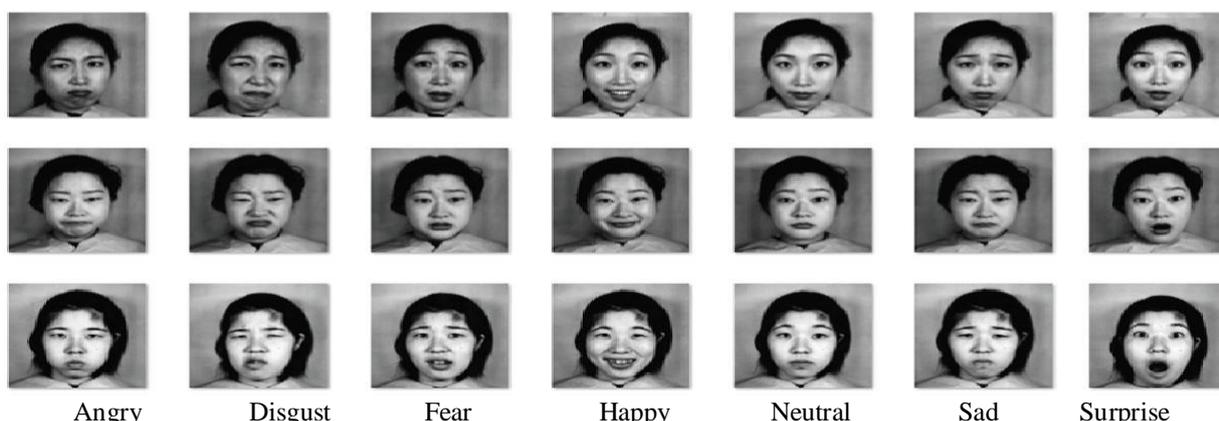


Fig. 1 Samples of images from the JAFFE benchmark database. Seven different expressions in the columns for 3 different subjects in rows



For every setting of the number of neurons in the hidden layer, we have ten different trials of the neural network and during every trial, the results of the neural network are found to be drastically different most of the time. This is because of randomness in initialization of connection weights and biases and randomness in data partitioning into training, CV and testing subsets at the beginning of every trial run. In addition, as this problem represents learning from data, the problem always has multiple non-optimal solutions (multiple local minima) and one optimal solution (global minimum). When training of the neural network proceeds, the learning algorithm often gets trapped to any one of the local minima and during every trial it can go on finding different local minima and thus further worsens the matter of finding a global optimal solution. Because of these complex situations, it is impossible to maintain exact reproducibility in the results produced by the same configuration of the neural network.

For training of the neural network, three different learning algorithms are used.

- Gradient descent with momentum and adaptive learning rate backpropagation (GDX)
- Resilient backpropagation (RP)
- Scaled Conjugate Gradient backpropagation (SCG)

Because of the fact that input feature values are bipolar, we used ‘tansig’ activation function of neurons in the hidden as well as output layer. All inputs are normalized by maintaining zero mean and unity variance. We insist on the simplicity and the minimum time and space complexity of the neural network based classifier. The network with minimum number of free parameters (connection weights and biases) should be employed. Possibility of using Principal Component Analysis is also explored for reduction of high input dimensionality and yielding reasonable classification performance with minimum complexity of the neural network model.

Cross entropy error criterion is appropriate to classification where the goal is to minimize the number of mis-classified training samples by imposing an exponentially increasing error the closer an output comes to being "1" when it should be "0", and vice versa. It calculates a network performance given targets (t) and outputs (y), with optional performance weights and other parameters. The function returns a result that heavily penalizes outputs that are extremely inaccurate (y near 1-t), with very little penalty for fairly correct classifications (y near t), where t denotes the target and y denotes the output of the neural network classifier. Minimizing cross-entropy leads to good classifiers.

The performance of the neural network based classifier is recorded for all ten different trials (runs) with respect to Cross-entropy Error on training dataset, on CV dataset and on Testing dataset; Average Classification Accuracy on Training dataset, Average Classification Accuracy on CV dataset, Average Classification Accuracy on Testing dataset and Overall Average Classification Accuracy. Finally, the best performance measures are highlighted with an emphasis to the Average Classification Accuracy in comparison with cross-entropy error, because we are solving a classification problem. As overall average classification accuracy reflects the average classification accuracy on training, validation and testing datasets; the neural network configuration with the highest average overall classification accuracy is chosen.

B. Principal Component Analysis for 2D – DCT based features

The feature vector based on 2D – DCT for Japanese Female Facial Expression (JAFFE) Database is

$FV_{DCT} = [DCT1, DCT2, \dots, DCT64, \text{Average}, SD, \text{Entropy}, \text{moment2}, \text{moment3}, \text{median}, \text{variance}, \text{contrast}, \text{correlation}, \text{energy}, \text{homogeneity}]$

Here, as total number of features in FV = 75, number of dimensionality of the input is equal to 75. Now principal component analysis is used to reduce 75 dimensional space with 75 variables to a k-dimensional space, where, $k < 75$.

A **scree plot** depicted in Fig. 2 displays the eigenvalues associated with a principal component or factor in descending order versus the number of the component or factor. It is used in principal components analysis to visually assess which components or factors explain most of the cumulative variability in the input data.



Fig. 2 Scree plot(Japan DCT PCA)

The Fig. 2 shows the scree plot which depicts number of principal components versus cumulative variability (%) and this plot enables us to determine the appropriate number of principal components for the desired cumulative variability. Cumulative variability of 99.9%, 99.5%, 99.1% and at least 95% are considered enough to represent the original input data into principal components transformation in order to achieve the reduction in the dimensions of the input space. After such a transformation, we can determine the number of factors (number of new input variables) needed to maintain the desired cumulative variability.

TABLE I JAPAN DCT PCA

No. of Factors	Cumulative Variability in %
21 (F1, ..., F21)	99.9
9 (F1, ..., F9)	99.5
6 (F1, ..., F6)	99.1
3 (F1, ..., F3)	97

Table I portrays the number of principal components and corresponding cumulative variability. So we can see that the maximum cumulative variability (99.9%) is obtained for 21 factors (F1, F2, F3, ..., F21) and here, the original 75 dimensional input space is converted to 21 dimensional input space with 21 input variables.

C. Principal Component Analysis for LBP based features

The feature vector based on LBP for Japanese Female Facial Expression (JAFFE) Database is as follows

$FV_{LBP} = [LBP1, LBP2, \dots, LBP59, \text{Average}, \text{SD}, \text{Entropy}, \text{moment2}, \text{moment3}, \text{kurtosis}, \text{skewness}, \text{median}, \text{variance}, \text{contrast}, \text{correlation}, \text{energy}, \text{homogeneity}]$

Here, as total number of features in $FV = 72$, number of dimensionality of the input is equal to 72. Now principal component analysis is used to reduce 72 dimensional input space with 72 input variables to a k -dimensional input space, where, $k < 72$.

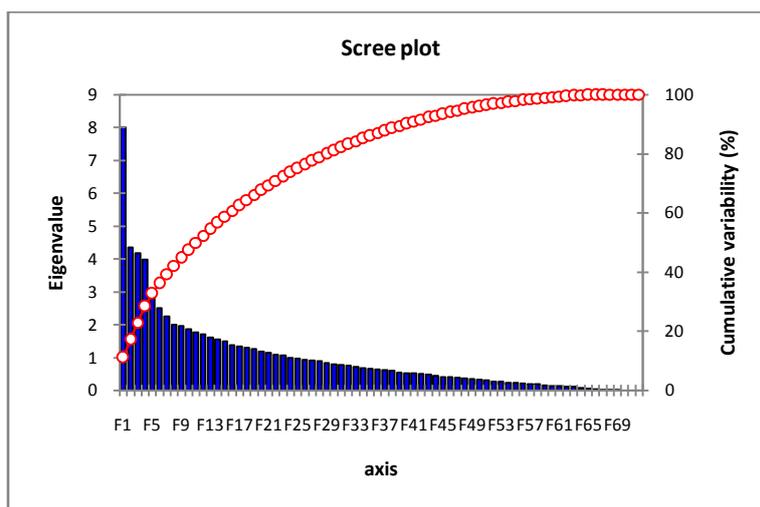


Fig. 3 Scree plot (Japan LBP PCA)

From Fig. 3, it is noticed that 99.9% cumulative variability can be achieved with only 67 factors. As number of principal components is decreased, cumulative variability also drops and for 42 factors, the cumulative variability is equal to 91%.

TABLE II JAPAN LBPPCA

No. of Factors	Cumulative Variability in %
67 (F1, ..., F67)	99.9
63 (F1, ..., F63)	99.5
60 (F1, ..., F60)	99.1
49 (F1, ..., F49)	95
42 (F1, ..., F42)	91



The above Table II displays the number of factors and corresponding cumulative variability. So, we can see that the maximum cumulative variability (99.9%) is obtained for 67 factors.

IV. EXPERIMENTAL RESULTS

A. Scatter Plots for Visualization of Decision Regions

Scatter plots are plots of sample input feature vectors in input features space. They are excellent visualization tools for determining feature vector distribution in \mathbf{R}^d , where $d \leq 3$. They often facilitate natural or obvious clustering of class-specific feature data and the partitioning of \mathbf{R}^d into decision regions for classification.

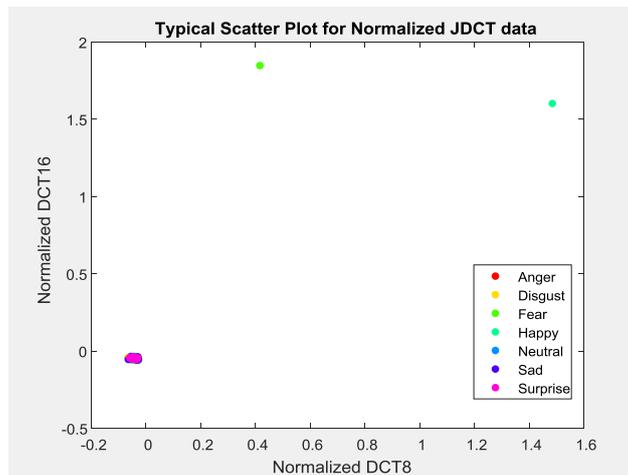


Fig. 4 Scatter plot showing Normalized DCT16 versus Normalized DCT8 for 7 different emotions

The Fig. 4 shows a typical scatter plot using features as Normalized DCT16 and Normalized DCT8 for 7 different emotions in relation to JAFFE database. It is noticed that there exists overlapping between emotions, complete absence of many emotion-classes and also there does not exist sufficient clustering of emotion-class specific feature data. Therefore, decision regions cannot be estimated for classification into seven different classes. This seems to be an issue with 2D – DCT based features. In view of this, it is certain that these features are not able to provide reasonable classification accuracy. We can see that distribution contains very few points. So it is impossible to classify seven emotions from only few points. This situation might be improved if we transform raw input features into principal components domain. After carrying out principal component analysis, we again sketch a typical scatter plot as shown in Fig. 5 using factor 16 (PC16) and factor 8 (PC8). It is observed that now at least the decision regions can be visualized in order to separate the feature space into seven different classes. Though, it is seen that the decision regions are highly nonlinear and complex in view of some overlapping between different emotions and therefore is very hard to distinguish one emotion from the another one. However, as compared to raw input DCT based features, PC factors can yield better classification leading to better recognition. Results are the best if two classes have little or no overlap in feature space. Accordingly, it makes sense to include (combination of) features that separate the classes as well as possible.

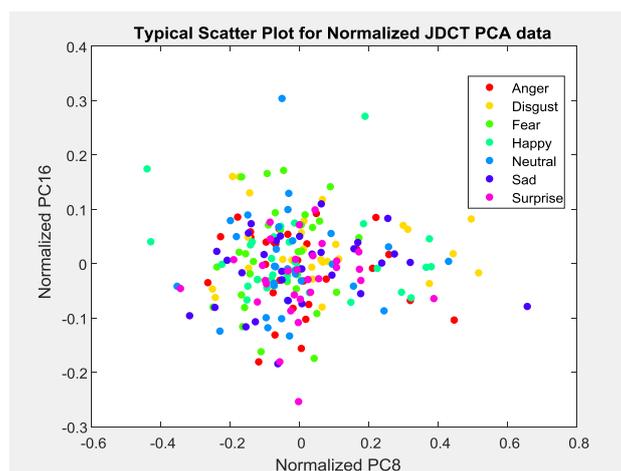


Fig. 5 Scatter plot showing Normalized PC16 versus Normalized PC8 for 7 different emotions

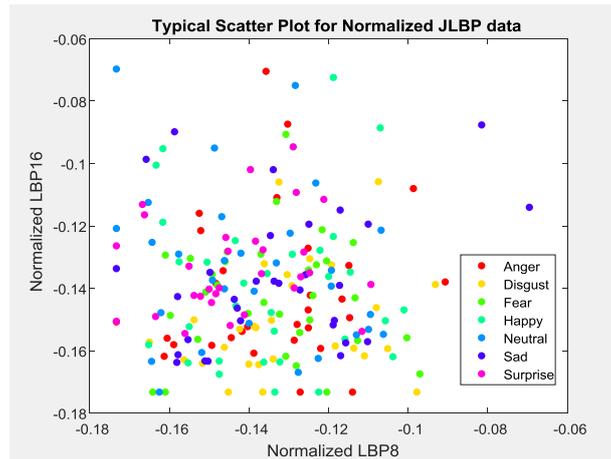


Fig. 6 Scatter plot showing Normalized LBP16 versus Normalized LBP8 for 7 different emotions

The Fig. 6 shows the scatter plot using features as Normalized LBP16 and Normalized LBP8 for 7 different emotions with respect to JAFFE database. It is noticed that there exists sufficient clustering of emotion-class specific feature data with very little or no overlapping between different emotion classes. The emotion-classes are sufficiently disjoint. The decision regions are noticed as nonlinear; however, they can be estimated by an optimal neural network model for classification into seven different classes. As compared to raw input features based on 2D – DCT, the quality of raw input features based on LBP is much better.

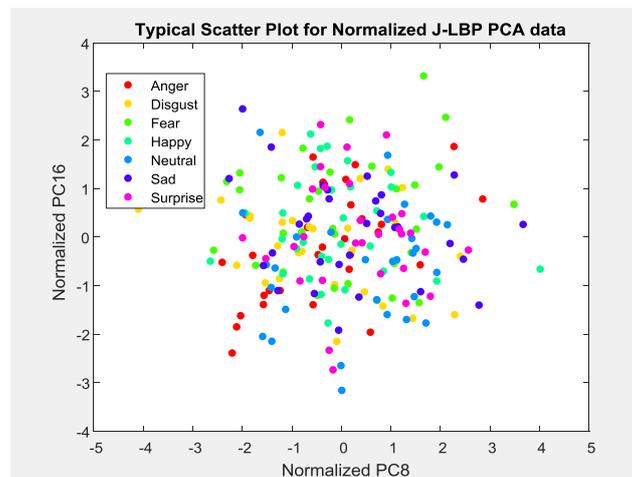


Fig. 7 Scatter plot showing Normalized PC16 versus Normalized PC8 for 7 different emotions

After carrying out principal components analysis on the features based on LBP, a typical scatter plot is depicted in Fig. 7 using factor 16 (PC16) and factor 8 (PC8) for seven different emotion-classes. It is seen that there exists sufficient clustering of emotion-class specific feature data with very little or no overlapping between different emotion classes. Although, the decision regions are noticed as nonlinear, they can be estimated by an optimal neural network model for classification into seven different classes. However, so far as the quality of raw input features based on LBP and transformed features using PCA is concerned, both are more or less equivalent, thus expecting reasonable classification and much better recognition in comparison with the features based on 2D – DCT.

B. Design of Neural Network based Classifier using Principal Components of Inputs

For the development of a classifier based on neural network, as reported earlier, two different types of features, such as, FV_{DCT} and FV_{LBP} have been employed. In addition, with a view to reduce the dimensions of raw input features, principal component analysis has been carried out for both types of features in order to determine the number of factors needed for maintaining a specific percentage of cumulative variability. Based on the cumulative variability, different number of PCA factors is applied to the neural network. As already mentioned in Section III, the optimal neural network for each value of the cumulative variability is determined. It is noticed that DCT based features are unable to produce an acceptable classification accuracy on training, validation and testing datasets. Their results are reported in Table III.



PCA counterparts of feature vector, FV_{DCT} has been presented to the neural network. Different factors are applied to the neural network corresponding to the cumulative variability of 95%, 99.1%, 99.5% and 99.9%, respectively and the best optimal neural network delivering maximum average classification accuracy for each input feature vector is obtained by following a network growing approach and a double loop design strategy as discussed in Section III.

Application of 21 factors (principal components needed for 99.9% cumulative variability) to a neural network with a single hidden layer comprising of 250 processing elements (neurons) results into the best classifier with average overall classification accuracy of 91.54929%, when trained with resilient backpropagation algorithm.

The results are shown in Table III. This network consists of 7,257 connection weights inclusive of all biases in hidden as well as output layer. In all other cases, the classification performance of the neural network is not satisfactory despite the use of large number of connection weights and biases. In Table, various acronyms are used, where FV denotes the feature vector, PE denotes the processing elements or neurons, CV denotes the cross validation, NCW denotes the total number of connection weights and biases present in the neural network.

TABLE III PERFORMANCE MEASURES OF NEURAL NETWORK FOR FV BASED ON DCTPCA

FV	PE	Cross-entropy Error			Average Classification Accuracy in %				Learning Algorithm	NCW
		Train	CV	Test	Train	CV	Test	Overall		
DCTPCA3	1500	0.2560	0.2677	0.2754	31.937	18.181	27.27	30.985	GDX	16507
	210	0.2549	0.2747	0.3286	29.3193	27.27	9.09	28.16901	RP	2317
	1140	0.2631	0.2400	0.4190	34.554	18.18	27.27	33.333	SCG	12547
DCTPCA6	510	0.1864	0.3382	0.3454	57.591	18.18	9.09	53.051	GDX	7147
	860	0.1483	0.4617	0.7016	75.392	27.272	0.00	69.014	RP	12047
	960	0.1468	0.4756	0.5118	72.774	45.45	9.09	68.075	SCG	13447
DCTPCA9	980	0.1782	0.3582	0.5695	67.015	45.45	18.18	63.380	GDX	16667
	910	0.1430	0.6367	0.6984	85.863	27.27	36.36	80.28169	RP	15477
	910	0.1635	0.5976	0.5223	79.581	18.181	9.09	72.769	SCG	15477
DCTPCA21	490	0.1054	0.3654	0.3401	89.528	36.36	27.27	83.568	GDX	14217
	250	0.0655	0.2916	0.5412	97.905	36.36	36.363	91.54929	RP	7257
	970	0.0695	0.3672	0.5760	98.429	27.272	9.09	90.140	SCG	28137

Now, principal components (factors) of FV_{LBP} are presented to the neural network with a view to reduce the input dimensions. Different factors are applied to the neural network corresponding to the cumulative variability of 91%, 95%, 99.1%, 99.5% and 99.9%, respectively and the best optimal neural network entailing maximum average classification accuracy for each input FV is obtained by following a network growing approach and a double loop design strategy. The performance measures of the best classifier for each FV for different learning algorithms are shown in Table IV.

As the recognition results for 60 PCA factors (99.1% cumulative variability) are found to be the best with Gradient descent with momentum and adaptive learning rate backpropagation algorithm, the computer simulation results are demonstrated as below.

In the following Table IV, we can see that for LBPPCA data with 60 PCA coefficients and PE=110, we got accuracy of 94.83568% for GDX training algorithm and the number of connection weights required are also very less i.e. 7487 as compared to other two algorithms for same number of PCA coefficients. Though the accuracies for other two algorithms for PCA = 60 are greater than that for GDX, the connection weights required are more. This, in turn, increases the complexity of Neural Network. So we always prefer optimal design of Neural Network with minimum number of connection weights and biases. So we have considered PCA = 60 with minimum number of connection weights and that we got using GDX training algorithm.

Application of 60 factors (principal components needed for 99.1% cumulative variability) to a neural network with a single hidden layer comprising of 110 processing elements (neurons) results into the best classifier with average overall classification accuracy of 94.83568% when trained with Gradient descent with momentum and adaptive learning rate backpropagation algorithm. The results are shown in Table IV. This network consists of 7,487 connection weights inclusive of all biases in hidden as well as output layer.



TABLE IV PERFORMANCE MEASURES OF NEURAL NETWORK FOR FV BASED ON LBPPCA

FV	PE	Cross-entropy Error			Average Classification Accuracy in %				Learning Algorithm	NCW
		Train	CV	Test	Train	CV	Test	Overall		
LBPPCA42	330	0.047835	0.203284	0.417653	100	54.54545	45.45455	94.83568	GDX	16507
	170	0.05795	0.592754	0.505196	100	54.54545	63.63636	95.77465	RP	8507
	180	0.033049	0.209439	0.259094	100	54.54545	45.45455	94.83568	SCG	9007
LBPPCA49	490	0.044524	0.454447	0.098758	100	27.27273	81.81818	95.30516	GDX	27937
	130	0.051886	0.453659	0.549054	100	45.45455	72.72727	95.77465	RP	7417
	60	0.049512	0.19879	0.400422	100	63.63636	36.36364	94.83568	SCG	3427
LBPPCA60	110	0.0595	0.16176	0.3884	100	81.818182	18.181818	94.83568	GDX	7487
	190	0.0446	0.4271	0.4152	100	54.54545	63.63636	95.774648	RP	12927
	220	0.026004	0.125794	0.326341	100	72.72727	63.63636	96.71362	SCG	14967
LBPPCA63	170	0.038091	0.217154	0.363387	100	54.54545	54.54545	95.30516	GDX	12077
	100	0.039593	0.141652	0.623631	100	63.63636	45.45455	95.30516	RP	7107
	170	0.037318	0.210147	0.367018	100	63.63636	54.54545	95.77465	SCG	12077
LBPPCA67	280	0.034096	0.211101	0.308751	100	63.63636	54.54545	95.77465	GDX	21007
	200	0.064415	0.446574	0.797732	100	54.54545	54.54545	95.30516	RP	15007
	470	0.049387	0.283246	0.486565	100	72.72727	36.36364	95.30516	SCG	35257

The parameters of the neural network are as follows:

Input Layer: 60 Neurons

No. of PEs in Hidden Layer = 110 (optimal number decided after rigorous experimentation and re-training)

Output Layer: 7 Neurons as Seven emotion-classes are 1. Surprise, 2.Sad, 3.Neutral, 4.Happy, 5.Fear, 6.Disgust and 7.Anger

Architecture of patternnet: 60-110-7

Learning Algorithm: Gradient descent with momentum and adaptive learning rate backpropagation algorithm

For 110 neurons in the hidden layer, the neural network is re-trained ten times with different random initialization of connection weights including biases and random partitioning of the data into training, cross-validation and testing datasets (Training: 90%, Cross-validation: 5 % and Testing 5 %). Among ten different trials, during Trial 6, the maximum average overall classification accuracy is observed.

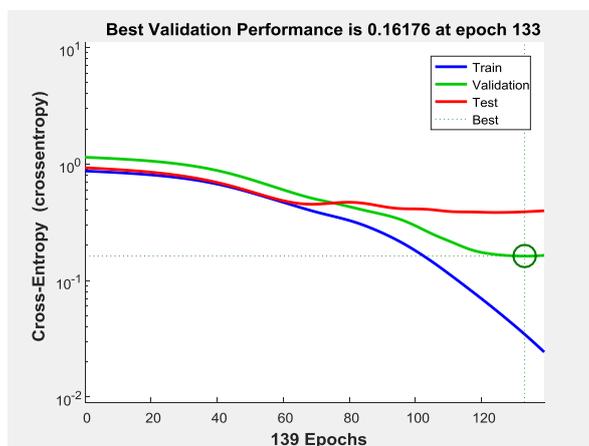


Fig. 8 Plot showing Cross-Entropy Error versus number of Epochs for training, cross validation and testing data

In above graph shown in Fig. 8, we can see that the best validation cross-entropy error performance after six consecutive validation checks is 0.16176, which is obtained at epoch number 133. So at this instant, the training of the Neural Network is terminated.

The Fig. 9 shows the overall confusion matrix which combines classification accuracies on training, cross-validation and testing datasets for various emotions like surprise, sad, neutral, happy, fear, disgust and anger. So we can see that the average overall correct classification accuracy is 94.8% for all the emotions. It is seen from the Table V that all 30 emotions are correctly recognized as surprise, resulting into the classification accuracy for surprise as 100%. It is also



observed that 27 emotions are correctly recognized as sad, but one emotion is incorrectly recognized as surprise, one emotion is incorrectly recognized as neutral, one emotion is incorrectly recognized as happy and one emotion is incorrectly recognized as fear.

Output Class	1	2	3	4	5	6	7	
1	30 14.1%	1 0.5%	1 0.5%	0 0.0%	0 0.0%	0 0.0%	1 0.5%	90.9% 9.1%
2	0 0.0%	27 12.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	1 0.5%	29 13.6%	0 0.0%	0 0.0%	0 0.0%	2 0.9%	90.6% 9.4%
4	0 0.0%	1 0.5%	0 0.0%	29 13.6%	0 0.0%	0 0.0%	0 0.0%	96.7% 3.3%
5	0 0.0%	1 0.5%	0 0.0%	2 0.9%	32 15.0%	0 0.0%	0 0.0%	91.4% 8.6%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 13.1%	0 0.0%	100% 0.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.5%	27 12.7%	96.4% 3.6%
	100% 0.0%	87.1% 12.9%	96.7% 3.3%	93.5% 6.5%	100% 0.0%	96.6% 3.4%	90.0% 10.0%	94.8% 5.2%
	1	2	3	4	5	6	7	
	Target Class							

Fig. 9 Overall Confusion Matrix for 60-110-7 Neural Network with GDx learning algorithm

Thus, out of total 31 emotions, only 27 emotions are correctly recognized as sad, resulting into the classification accuracy for sad of 27/31, i.e., 87.1%. It is also noticed that 29 emotions are correctly recognized as neutral and one emotion is incorrectly recognized as surprise, thus, leading to the classification accuracy for neutral as 29/30, i.e., 96.7%. In addition, it is also seen that 29 emotions are correctly recognized as happy and 2 emotions are incorrectly recognized as fear, thus, resulting into the classification accuracy with respect to happy as 29/31, i.e., 93.5%. Furthermore, all 32 emotions are correctly recognized as fear, entailing the classification accuracy for fear as 100%. It is also obvious that 28 emotions are correctly recognized as disgust and one emotion is incorrectly recognized as anger, resulting into the classification accuracy in relation to disgust as 28/29, i.e., 96.6%. Moreover, it is also evident that 27 emotions are correctly recognized as anger, one emotion is incorrectly recognized as surprise and two emotions are incorrectly recognized as neutral, resulting into the classification accuracy for anger as 27/30, i.e., 90%. The average overall classification accuracy for all emotions is computed as 94.8%. It is also confirmed that the total number of facial images in JAFFE database is equal to 213 (30 facial images expressing Surprise + 31 facial images expressing Sad + 30 facial images expressing Neutral + 31 facial images expressing Happy + 32 facial images expressing Fear + 29 facial images expressing Disgust + 30 facial images expressing Anger = 213, i.e., Total number of facial images).

TABLE V NO OF CORRECT CLASSIFICATIONS BY THE TRAINED NEURAL NETWORK (60-110-7, GDx)

No. of Recognized Emotion(Output)	No. of Desired Emotions (Target)						
	Surprise	Sad	Neutral	Happy	Fear	Disgust	Anger
Surprise	30	1	1	0	0	0	1
Sad	0	27	0	0	0	0	0
Neutral	0	1	29	0	0	0	2
Happy	0	1	0	29	0	0	0
Fear	0	1	0	2	32	0	0
Disgust	0	0	0	0	0	28	0
Anger	0	0	0	0	0	1	27
Correct Recognition	30/30	27/31	29/30	29/31	32/32	28/29	27/30
Classification Accuracy	100%	87.1%	96.7%	93.5%	100%	96.6%	90%
Overall Average Classification Accuracy	94.8%						



C. Design of Neural Network based Classifier using raw inputs

For the development of a classifier based on neural network, as reported earlier, two different types of raw input features, such as, FV_{DCT} and FV_{LBP} have been employed. First, we have applied raw input features based on 2D – DCT. As interpreted from the Scatter plot shown in Fig. 4, the quality of FV_{DCT} is very poor. Therefore, we cannot expect acceptable classification accuracy. In case of FV_{DCT} , the maximum average overall classification accuracy is obtained for a neural network with a single hidden layer containing 520 neurons and trained with scaled conjugate gradient backpropagation algorithm. The best average overall classification accuracy is observed as only 42.253521% at the cost of formidable total number of connection weights and biases amounting to 43,167. In view of this, we shall train our neural network based classifier using FV_{LBP} in order to achieve maximum average classification accuracy.

The parameters of the neural network are as follows:

Input Layer: 72 Neurons

No. of PEs in Hidden Layer = 70 (optimal number decided after rigorous experimentation and re-training)

Output Layer: 7 Neurons as Seven emotion-classes are 1. Surprise, 2.Sad, 3.Neutral, 4.Happy, 5.Fear, 6.Disgust and 7. Anger

Architecture of patternnet: 72-70-7

Learning Algorithm: Gradient descent with momentum and adaptive learning rate backpropagation algorithm

For 70 neurons in the hidden layer, the neural network is re-trained ten times with different random initialization of connection weights including biases and random partitioning of the data into training, cross-validation and testing datasets (Training: 90%, Cross-validation: 5 % and Testing 5 %). Among ten different trials, during Trial 1, the maximum average overall classification accuracy is observed. Table VI enlists the different performance measures of the optimal neural network delivering the maximum average overall classification accuracy for input features based on 2D – DCT as well as LBP for three different learning algorithms, namely, scaled conjugate gradient descent backpropagation algorithm (SCG), resilient backpropagation algorithm (RP) and Gradient descent with momentum and adaptive learning rate backpropagation algorithm (GDX).

TABLE VI PERFORMANCE MEASURES OF NEURAL NETWORK FOR RAW INPUT DATA, FV_{DCT} AND FV_{LBP}

FV	PE	Cross-entropy Error			Average Classification Accuracy in %				Learning Algorithm	NCW
		Train	CV	Test	Train	CV	Test	Overall		
DC	680	0.2612	0.2600	0.2826	28.272251	18.181818	18.181818	27.230047	GDX	56447
T	80	0.2336	0.2742	0.2747	35.602094	18.181818	9.090909	33.333333	RP	6647
(75)	520	0.2333	0.3347	0.2733	43.455497	45.454545	18.181818	42.253521	SCG	43167
LBP	70	0.032	0.11119	0.3427	100	72.727273	72.727273	97.183099	GDX	5607
(72)	120	0.040972	0.393554	0.257441	100	54.54545	54.54545	95.30516	RP	9607
	380	0.034832	0.212824	0.248576	100	63.63636	54.54545	95.77465	SCG	30407

In the above Table VI, we can see that for feature vector, FV_{LBP} , a neural network with PE=70, we have obtained maximum average overall classification accuracy of 97.183099% for GDX training algorithm and the number of connection weights required are also very less i.e., only 5,607 as compared to other two algorithms. In case of the feature vector based on DCT data, i.e., FV_{DCT} , the classification accuracies are very poor and the number of connection weights inclusive of biases required is enormous.

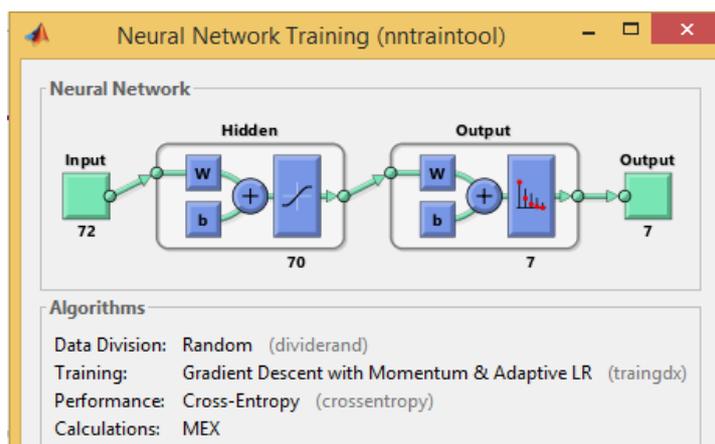


Fig. 10 Neural Network architecture (72-70-7), GDX Learning Algorithm



The Fig. 10 shows Neural Network architecture containing single hidden layer with 72 inputs of FV_{LBP} applied for training. Optimal number of hidden neurons are determined as 70 after following the design strategy as explained in Section III. Output layer contains 7 outputs representing 7 different emotions. The training algorithm used is GDX and the transfer function used is tanh for both hidden and output layers.

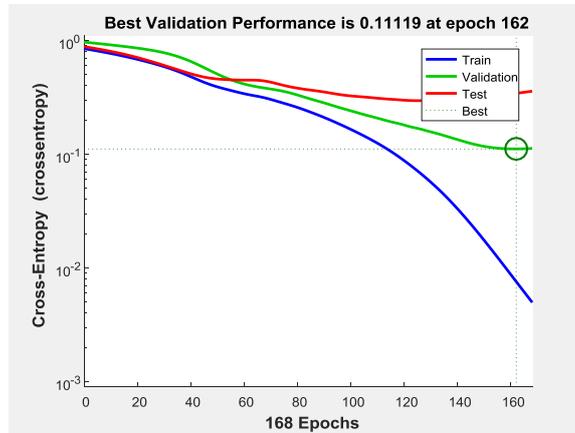


Fig. 11Plot showing Cross-Entropy Error versus number of Epochs for training, cross validation and testing data

In above graph shown in Fig. 11, we can see that the best validation cross-entropy error performance after six consecutive validation checks is 0.11119, which is obtained at epoch number 162. So at this instant, the training of the Neural Network is terminated.

Overall Confusion Matrix

	1	2	3	4	5	6	7	
1	28 13.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	1 0.5%	29 13.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.7% 3.3%
3	0 0.0%	1 0.5%	29 13.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.7% 3.3%
4	1 0.5%	0 0.0%	0 0.0%	31 14.6%	1 0.5%	0 0.0%	0 0.0%	93.9% 6.1%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	31 14.6%	0 0.0%	0 0.0%	100% 0.0%
6	0 0.0%	1 0.5%	0 0.0%	0 0.0%	0 0.0%	29 13.6%	0 0.0%	96.7% 3.3%
7	0 0.0%	0 0.0%	1 0.5%	0 0.0%	0 0.0%	0 0.0%	30 14.1%	96.8% 3.2%
	93.3% 6.7%	93.5% 6.5%	96.7% 3.3%	100% 0.0%	96.9% 3.1%	100% 0.0%	100% 0.0%	97.2% 2.8%
	1	2	3	4	5	6	7	

Target Class

Fig. 12OverallConfusion Matrixfor 72-70-7 Neural Network with GDX learning algorithm

The Fig. 12 shows the overall confusion matrix which combines classification accuracies of training, cross-validation and testing datasets for various emotions like surprise, sad, neutral, happy, fear, disgust and anger. So we can see that the average overall correct classification accuracy is 97.2% for all the emotions.

It is seen from the Table VII that 28 emotions are correctly recognized as surprise, one emotion is incorrectly recognized as sad and one emotion is incorrectly recognized as happy, resulting into the classification accuracy for surprise as 28/30, i.e., 93.3%. It is also observed that 29 emotions are correctly recognized as sad, one emotion is incorrectly recognized as neutral and one emotion is incorrectly recognized as disgust. Thus, out of total 31 emotions, only 29 emotions are correctly recognized as sad, resulting into the classification accuracy for sad of 29/31, i.e., 93.5%. It is also noticed that 29 emotions are correctly recognized as neutral and one emotion is incorrectly recognized as anger, thus, leading to the classification accuracy for neutral as 29/30, i.e., 96.7%.



TABLE VII NO OF CORRECT CLASSIFICATIONS BY THE TRAINED NEURAL NETWORK (72-70-7, GDX)

No. of Recognized Emotion(Output)	No. of Desired Emotions (Target)						
	Surprise	Sad	Neutral	Happy	Fear	Disgust	Anger
Surprise	28	0	0	0	0	0	0
Sad	1	29	0	0	0	0	0
Neutral	0	1	29	0	0	0	0
Happy	1	0	0	31	1	0	0
Fear	0	0	0	0	31	0	0
Disgust	0	1	0	0	0	29	0
Anger	0	0	1	0	0	0	30
Correct Recognition	28/30	29/31	29/30	31/31	31/32	29/29	30/30
Classification Accuracy	93.3%	93.5%	96.7%	100%	96.9%	100%	100%
Overall Average Classification Accuracy	97.2%						

In addition, it is also seen that all 31 emotions are correctly recognized as happy, thus, resulting into the classification accuracy with respect to happy as 100%. Furthermore, 31 emotions are correctly recognized as fear and one emotion is incorrectly recognized as happy, entailing the classification accuracy for fear as 31/32, i.e., 96.9%. It is also obvious that all 29 emotions are correctly recognized as disgust, resulting into the classification accuracy in relation to disgust as 100%. Moreover, it is also evident that all 30 emotions are correctly recognized as anger, resulting into the classification accuracy for anger as 100%. The average overall classification accuracy for all emotions is computed as 97.2%.

V. CONCLUSION

The problem of recognition of facial expressions (emotions) from JAFFE database using neural network is investigated in this research work. Between 2D- DCT and LBP based feature extraction schemes, FV_{LBP} delivers the best classification performance of the neural network as compared to the other one. Input features based on principal components (factors) of FV_{DCT} presented to the single hidden layer neural network with variation in number of processing elements in the hidden layer, retraining the network ten times (ten different trials for each setting of the number of neurons in the hidden layer) with different random initialization of connection weights and biases, randomly chosen data partitioning during every trial and different learning rules do not seem to provide reasonable classification performance. This is because, overall classification accuracy of 91.54929% is achieved for DCTPCA21 principal components (21 factors chosen to maintain 99.9% cumulative variability) using a neural network with 250 neurons in the hidden layer trained with resilient backpropagation algorithm and the total number of connection weights inclusive of biases is 7,257 as shown in Table III. On the contrary, LBP based feature extraction scheme seems to be more promising and fruitful, which is again obvious from the scatter plots drawn in original raw input feature space and PCA transformed feature space. Maximum average overall classification accuracy of 97.183099% is achieved for raw input features based on LBP, i.e., FV_{LBP} (Total number of features = 72) using a neural network with 70 neurons in the hidden layer and trained with Gradient descent with momentum and adaptive learning rate backpropagation algorithm and the total number of connection weights inclusive of biases is only 5,607, which is very optimal as demonstrated in Table VI. However, when LBPPCA 60 (60 factors chosen to maintain 99.1% cumulative variability) features are presented to the single hidden layer neural network containing 110 processing elements in the hidden layer and trained with Gradient descent with momentum and adaptive learning rate backpropagation algorithm, the maximum average overall classification accuracy achieved is noticed as 94.83568% at the cost of 7,487 connection weights inclusive of biases as shown in Table IV. Similarly, when LBPPCA 60 (60 PC factors chosen to maintain 99.1% cumulative variability) features are presented to the single hidden layer neural network constituting 190 processing elements in the hidden layer and trained with resilient backpropagation algorithm, the maximum overall classification accuracy achieved is noticed as 95.774648% at the cost of 12,927 connection weights inclusive of biases as exhibited in Table IV. Thus, in the case of LBP PCA based features, it is noticed that even though the input dimensionality is reduced, the overall classification accuracy has not improved much and this is only possible with the help of a neural network based classifier comprising of more number of connection weights and biases. However, raw input features based on LBP, i.e., FV_{LBP} delivers the best classification performance as maximum average overall classification accuracy of 97.183099% is achieved for just 70 processing elements in the hidden layer and only 5,607 connection weights inclusive of biases. No other researcher so far has reported such a high value of average classification accuracy with very compact neural network with reduced time as well as space-complexity on JAFFE database. The major contribution of this research work is in designing not only an optimal neural network but also an optimal feature vector by fusion of image LBP coefficients along with some other carefully chosen features describing image statistics,

texture, entropy, etc. In view of this, it is recommended to use raw input features based on LBP combined with image statistics and other features such as texture, entropy, etc. for training a neural network with a single hidden layer with only 70 neurons, Gradient descent with momentum and adaptive learning rate backpropagation algorithm.

ACKNOWLEDGMENT

The principal author wishes to express her deep sense of gratitude towards **Prof. Dr. Shripad Bhatlawande**, Professor and Head of the Department of Electronics and Telecommunication Engineering and **Prof. Dr. R. M. Jalnekar**, Director of Vishwakarma Institute of Technology, Pune for providing all support and facilities in order to pursue the research work.

REFERENCES

- [1] Ekman P., Huang T. S., Sejnowski T. J., Hager J. C., "Final Report to NSF of the Planning Workshop on Facial Expression Understanding, 1992.
- [2] Alvino C., Kohler C., Barrett F., Gur R. E., Gur R. C., Verma R., "Computerized measurement of facial expression of emotions in schizophrenia", *Journal of Neuroscience Methods*, 2007, 163, 350-361.
- [3] Wang P., Barrett F., Martin E., Milonova M., Gur R. E., Gur R. C., Kohler C., Verma R., "Automated video-based facial expression analysis of neuropsychiatric disorders" *Journal of Neuroscience Methods*, 2008, 168, 224-238
- [4] Dinges D. F., Venkataraman S., McGlinchey E. L., Metaxas D. N., "Monitoring of facial stress during space flight: Optical computer recognition combining discriminative and generative methods", *Acta Astronautica*, 2007, 60, 341 - 350
- [5] Darwin C., "The Expression of Emotions in Man and Animals," Murray J., reprinted by University of Chicago Press, 1965, 1872. Online Edition: <http://human-nature.com/darwin/emotion/contents.html>
- [6] Ekman P., Friesen W.V., "The Facial Action Coding System: A Technique for the Measurement of Facial Movement," San Francisco, Consulting Psychologists Press, 1978., Online Edition: <http://face-and-emotion.com/dataface/facs/manual/TitlePage.html>
- [7] Padgett C., Cottrell G.W., "Representing Face Images for Emotion Classification," *Proceedings of Conference on Advances in Neural Information Processing Systems*, 1996, 894-900.
- [8] Kobayashi H., Hara F., "Facial Interaction between Animated 3D Face Robot and Human Beings," *Proceedings of International Conference on Systems, Man, Cybernetics*, 1997, 3, 732-737.
- [9] Kobayashi, H., Hara, F., "The Recognition of Basic Facial Expressions by Neural Network," *Proceedings of International Joint Conference on Neural Network*, 1991, 460-466.
- [10] Zhang Z., Lyons M., Schuster M., Akamatsu S., "Comparison between Geometry- Based and Gabor Wavelets-Based Facial Expression Recognition Using MultiLayer Perceptron," *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 1998, 454-459.
- [11] Zhao J., Kearney G., "Classifying Facial Emotions by Backpropagation Neural Networks with Fuzzy Inputs," *Proceedings of Conference on Neural Information Processing*, 1996, 1, 454-457.
- [12] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 1999, 13, 893-911.
- [13] Ekman P., Friesen W.V., "Unmasking the Face," Upper Saddle River, New Jersey: Prentice Hall, 1975.
- [14] Khan M. M., Ingleby M., Ward R. D., "Automated Facial Expression Classification and Affect Interpretation Using Infrared Measurement of Facial Skin Temperature Variations", *ACM Transactions on Autonomous and Adaptive Systems*, 2006, 1, 91-113.
- [15] Kanade T., Cohn J., Tian Y., "Comprehensive Database for Facial Expression Analysis," 4th IEEE International Conference on Automatic Face and Gesture Recognition, France, March 2000.
- [16] Reddy N. P., Buch O. "Committee Neural Networks for Speaker Verification," *Computer Programs and Methods in Biomedicine*, 2003, 72, 109-115.
- [17] Das A., Reddy N. P., Narayanan J., "Hybrid Fuzzy Logic Committee Neural Networks for Recognition of Swallow Acceleration Signals," *Computer Methods and Programs in Biomedicine*, 2001, 64, 87-99.
- [18] Reddy N. P., Prabhu D., Palreddy S., Gupta V., Suryanarayanan S., Canilang E.P., "Redundant Neural Networks for Medical Diagnosis: Diagnosis of Dysphagia," *Intelligent Systems through Artificial Neural Networks*, 1995, 5, 699-704
- [19] Ekman, P., & Rosenberg, E. L., "what the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)", New York: Oxford University Press, 1997.
- [20] B. Fasel, J. Luetttin, "Automatic facial expression analysis: a survey", *Pattern Recognition*, Vol. 36, pp 259275, 2003.
- [21] Fernando De La Torre & Jeffrey F. Cohn, 'Facial Expression Analysis', Springer, VAH, pp 377-409, 2011.
- [22] ShishirBashyal, Ganesh K. Venayagamoorthy, "Recognition of Facial Expression using Gabor wavelets and learning vector quantization", Elsevier, EAAI 21, pp 1056-1064, 2009
- [23] Fei Long et al., "Learning Spatiotemporal feature by using Independent component analysis with application to facial expression recognition", Elsevier Trans., *Neurocomputing*, 126-132, 2012.
- [24] WenfeiGu, et al., "Facial expression recognition using radial encoding of local Gabor Feature and Classifier Synthesis", Elsevier Ltd Trans., 2012.
- [25] G. Guo and C.R. Dyer, "Learning from Examples in the Small Sample Case: Face Expression Recognition," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 3, pp. 477-488, June 2005.
- [26] J.-J. Wong and S.-Y. Cho, "A Face Emotion Tree Structure Representation with Probabilistic Recursive Neural Network Modeling," *Neural Computing and Applications*, vol. 19, pp. 33-54, 2010.
- [27] Z. Wenming, Z. Xiaoyan, Z. Cairong, and Z. Li, "Facial Expression Recognition Using Kernel Canonical Correlation Analysis (KCCA)," *IEEE Trans. Neural Networks*, vol. 17, no. 1, pp. 233-238, Jan. 2006.
- [28] C. Zhengdong, S. Bin, F. Xiang, and Z. Yu-Jin, "Automatic Coefficient Selection in Weighted Maximum Margin Criterion," *Proc. 19th Int'l Conf. Pattern Recognition*, pp. 1-4, 2008.
- [29] C. Shan, S. Gong and P.W. McOwan, "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Computing*, vol. 27, pp. 803-816, 2009.
- [30] UrvashiBakshi, R. Singhal, "A New approach of face Recognition using DCT, PCA, and Neural network in Matlab", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Volume 3, Issue 3, May-June 2014.
- [31] Xun Wang, Xingang Liu, Lingyun Lu, ZhixinShen, "A New Facial Expression Recognition Method Based on Geometric Alignment and LBP Features," 2014 IEEE 17th International Conference on Computational Science and Engineering (CSE), pp. 1734-1737, 19-21 Dec.2014.