I.J. Intelligent Systems and Applications, **2019, 1, 55-66**
Published Online January 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2019.01.06

Development of Hybrid Learning Machine in Complex Domain for Human Identification

Swati Srivastava
Department of Computer Science & Engineering, Harcourt Butler Technical University, Kanpur, India
E-mail: swati.cs1409@gmail.com

Bipin K. Tripathi
Department of Computer Science & Engineering, Harcourt Butler Technical University, Kanpur, India
E-mail: abkt.iitk@gmail.com

Received: 10 March 2018; Revised: 01 May 2018; Accepted: 23 May 2018; Published: 08 January 2019

***Abstract*—**This paper presents a hybrid learning machine for human identification. It is a merger of eigenface with fisherface method, genetic fuzzy clustering and complex neural network. The non-linear aggregation based summation and radial basis function neural networks (NLA-SRBF NNs) are proposed as one of the functional component of the novel learning machine. The architecture of NLA-SRBF NNs incorporates hidden neurons, with summation and radial basis aggregation, and output neurons with only summation aggregation, along with complex resilient propagation (ČRPROP) learning procedure. The improved learning and speedy convergence of NLA-SRBF NN enables the hybrid machine to provide better recognition accuracy. The learning machine consists of feature extraction, unsupervised clustering and supervised classification module. The aim of our proposal is to enhance the performance of biometric based recognition system. The efficacy and potency of our hybrid learning machine demonstrated on three benchmark biometric datasets- extended Cohn-Kanade, FERET and AR face datasets to comprehend the motivation. The performance comparisons of different variations of hidden neuron and learning algorithm thoroughly presented the superiority of the proposed NN based hybrid learning machine.

***Index Terms*—**Eigenface, fisherface, genetic fuzzy clustering, complex neural network, complex resilient propagation, biometric.

I. Introduction

In current scenario, when more or less all real life applications are biometric dependent for security reasons, the biometric based recognition system has expanded its immense magnitude. Over the years extensive attempts have been made in human recognition domain for the purpose of identification and authentication. Several recognition techniques in various areas have been proposed in past [1-4]. Among a range of techniques, neural network (NN) based methods hold considerable position due to efficient performance. NN with conventional neurons is widely used classifier in different domains. To improve the efficiency of conventional NNs broad attempts have been made to develop various neuron structures [5-12]. Though among them higher order neurons have evidenced to be most efficient but due to combinatorial outburst of terms they experience the curse of dimensionality specifically when they are implemented in complex domain. Work on the development of higher order neurons is growing continuously which contributed high performance in many classification models. In continuation to the above progression, this paper presents a higher order neuron called *ČTROIKA* with non-linear aggregation of summation and radial basis functions. We present a network built up on novel neurons and conventional neurons. The traditional neural network employs real back propagation (ŘBP) algorithm for learning which has foremost limitations of slow convergence and getting stuck into local minima. To surmount these limitations, some variations in basic error propagation procedure were suggested [13-15]. Although none of the suggested variations were capable to make significant improvement in convergence rate. Additional attempts in the direction of performance enhancement shifted the classifier from real domain to complex domain. The supremacy of complex NN has been observed in the recent past [11]. The complex back propagation (ČBP) reduces the probability of getting trapped in local minima and improves the convergence rate. With widespread efforts of researchers towards the performance advancement of recognizer, resilient propagation(RPROP) algorithm was developed [18] in an attempt to improve upon the BP algorithm. RPROP outperforms BP by providing faster training and better efficiency along with the advantage that RPROP does not require to specify any free parameter values such as the learning rate and an optional momentum term as opposed to BP. RPROP provides adequately faster convergence with employment of higher order neurons. This paper presents NLA-SRBF NNs which are based on modified ČRPROP learning algorithm along with novel neuron model *ČTROIKA.* This NN constitutes a classification module of the proposed hybrid learning machine. Over the years, various researches have projected the significance of exploiting multiple computational intelligent techniques to develop a classification model which provides better recognition accuracy than conventional diminutive models. Various classification techniques in this sequence have brought forward by researchers [19, 20]. This paper presents a hybrid learning machine that comprises of statistical techniques incorporated with genetic fuzzy partitioning and proposed neural classifier.

Rest of the paper is organized as: Section II presents related work. Section III elaborates the proposed hybrid learning machine. Section IV is devoted to the performance evaluation of the new learning machine along with the performance comparisons with other classifiers. Finally, we conclude the paper in Section V.

II. Related Work

Over the years various recognition methods have been proposed such as geometric feature based methods [1,2], template based methods [1], correlation based methods [1], support vector machine approach [1,3], NN based methods [1,4]. NN based recognition and classification techniques are widely applicable and therefore became the preferred area of researchers. NN with conventional neurons are not efficient to recognize the images acquired in real situations. To improve the efficiency of NNs broad attempts have been made to develop higher order neuron structures. These attempts contributes pi-sigma [5, 6], second order neurons [7], generalized neurons [8, 9] and other higher order neurons [10-12]. Among them higher order neurons have verified to be the most efficient Moreover, the conventional NN employs BP algorithm for training of the network. BP is less accurate with slow convergence rate and therefore not cope up with the images possessing wide variations in facial expressions, illumination conditions, emotions and pose. To improve the performance of BP, some variations in basic error propagation procedure were suggested like addition of momentum term [13], modified error function [14], and Quick Prop [15]. But none of these variations contribute significant improvement in BP. Further, BP provides improved results in complex domain [16, 17] but still needs an improvement. RPROP learning algorithm [18] has developed to additionally improve the training of the neural network which in turn provides better recognition accuracy. This local gradient based adaptation technique is dependent on the sign of partial derivative of error function on weight update rather than the size of the derivative. A variety of recognition methods with multiple machine intelligent techniques have been developed such as probabilistic neural network with adjustable fuzzy clustering [19], PCA-LDA with polynomial Radial Basis Function (RBF) neural network [20]. The aim of all such fusions is to enhance the performance of recognition system. This work presents a hybrid model which is an amalgamation of eigenface method, fisherface method, fuzzy distribution with evolutionary computation and proposed neural classifier. The proposed neural classifier is built up on proposed novel neurons, conventional neurons and ČRPROP learning algorithm.

III. Hybrid Learning Machine (HLM): The Proposed Recognizer

This paper presents a hybrid learning machine which can perform face recognition with the proposed classifier. First, the eigenface method in combination with fisherfaces extract the features of database images. Each class is then represented by the average image which is obtained by calculating the mean of training images of the respective class. Input classes are clustered optimally by means of genetic fuzzy partitioning. The outcome of unsupervised clustering act as the input-output mapping for the classification segment. The novel classifier incorporates hidden layer which consists of proposed complex neurons *ČTROIKA* and output layer which consists of complex conventional neurons. The ČRPROP learning algorithm is employed to train the network. Then the trained network perform classification of test patterns. The aim of our learning machine is to improve the performance of biometric based identification and recognition system. The proposed learning machine includes following steps:

- Feature extraction
- Unsupervised Genetic Fuzzy Clustering
- Supervised Classification through NLA-SRBF NNs

A. Feature extraction

The performance of any image recognition system profoundly depends upon the choice of features of image to be processed. The first and foremost step of a recognition system is to pre-process the input images by extracting noteworthy features which are able to define the original image with all significant features. Eigenface paradigm [21,22] is one of the popular technique to mine such features which represents the image with reduced data which contains only representative information. Eigenfaces does not consider the classification aspect as it is based on optimal representation criterion. To improve the standalone classification performance of eigenfaces, it is further needed to combine with some discrimination criterion. Discriminant eigenface paradigm [22] is widely used discrimination criterion which overcomes the limitations of eigenfaces while retaining the idea of projection from high dimensional feature space to extensively lower dimensional feature space. To extract the desired features of input dataset, eigenface pursued by fisherface method is used in our proposal. Let an $a \times a$ image is represented as a linear vector I_i of size a^2. Let P be the total number of images in the dataset and Q be the number of input classes with s images in each class.

B. Unsupervised Genetic Fuzzy Clustering

This is the intermediate step of the proposed learning machine in order to get the desired output for supervised classification. Once dataset images in lower dimensional feature space are obtained, average image of each class is calculated for training by considering q ($q<s$) images of each class. Let X be the feature vector of average images of Q input classes and U be the fuzzy distribution of Q classes among C clusters which is obtained by fuzzy-c-means clustering algorithm. The impartial product of fuzzy algorithm is not sufficient to obtain the optimal distribution.

Algorithm for feature extraction

1. Accumulate the database images in a matrix I $(a^2 \times P)$ and find the adjusted data matrix, $\theta = I - I_{mean}$
2. Determine the covariance matrix, $\chi = \theta \times \theta^T$ where θ^T is the transpose of θ . The dimensions of χ will be $a^2 \times a^2$.
3. Compute the eigenvectors and eigenvalues of χ .
4.Select v eigenvectors $F = \{F_1, F_2, ..., F_v\}$ corresponding to v largest eigenvalues.
5. Eigenface based features G can be obtained by projecting θ onto Eigenface space F, $G = F^T \times \theta$.
6. Since G does not contain any class discrimination information but only reduces dimensions, apply the Fisherface method to $G = \{G_1, G_2, ..., G_P\}$ in order to find an optimal subspace for classification.
7. Determine between-class scatter matrix and within-class scatter matrix[22] respectively as:
$$S_{bet} = \sum_{i=1}^{Q} s \times \left[\left(\sum_{j=1}^{s} G_j^i \Big/ s \right) - G_{mean} \right] \times \left[\left(\sum_{j=1}^{s} G_j^i \Big/ s \right) - G_{mean} \right]^T \quad \text{and}$$
$$S_{wthn} = \sum_{i=1}^{Q} \sum_{j=1}^{s} \left[G_j - \left(G_j^i / s \right) \right] \times \left[G_j - \left(G_j^i / s \right) \right]^T$$
8. Determine the optimal subspace H by selecting the generalized eigenvectors corresponding to the $Q-1$ largest generalized eigenvalues β_i by solving $S_{bet} G_i = \beta_i S_{wthn} G_i$ where $i = 1, 2, ..., Q-1$.
9.Discriminant eigenface based features D can be obtained by projecting eigenfaces G onto optimal subspace H, $D = H^T \times G$. Matrix D (size $P \times (Q-1)$) represents the database images in the desired lower dimensional feature space.

The reason behind is for the same dataset it gives dissimilar partitions in different runs. In order to obtain the optimized partition, the above obtained partitions are processed based on the criteria of survival of the fittest [23], and the partition with highest fitness value [24] return the optimized one. This distribution will act as initial fuzzy partition for the next generation. The process continues till the difference of two consecutive generations is less than or equal to pre-defined threshold ρ. The required optimal distribution is the partition of the last generation which further go through defuzzification. Let Ω be the maximum number of classes that are allowed to be in a cluster. The cluster allocation matrix (size $C \times \Omega$) is the desired output for the classifier which is trained accordingly to classify the test images.

Algorithm for unsupervised clustering

1. Calculate the mean vector for all Q classes using q training images of each class from matrix D as: $\overline{x_j} = \frac{1}{q}\left(\sum_{k=1}^{q} x_{kj} \right)$, where $1 \le j \le Q$ and the array of average image of input classes is given as $X = \left\{ \overline{x_1}, \overline{x_2}, ..., \overline{x_Q} \right\}$.

2. Initialize $g=1$; repeat for g generations until $\left\| U^{(g+1)} - U^g \right\| \le \rho$ (for $g>1$)
   i. repeat for *run=1:r*
      (a) For $g=1$, initialize fuzzy partition matrix U of size $C \times Q$. The initial population $U = \left[u_{ij} \right]$ is randomly initialized such that $u_{ij} \in [0,1]$, where $1 \le i \le C$ and $1 \le j \le Q$.
      (b) Update U by recursively minimizing the objective function $J = \sum_{i=1}^{C} \sum_{j=1}^{Q} u_{ij}^{\tau} \delta_{ij}$ which includes the update process of membership function and cluster centers respectively as $u_{ij} = \delta_{ij}^{\frac{-2}{\tau-1}} \Big/ \sum_{k=1}^{C} \delta_{kj}$ and $c_i = \sum_{j=1}^{Q} u_{ij}^{\tau} \overline{x_j} \Big/ \sum_{j=1}^{Q} u_{ij}^{\tau}$
      where $\delta_{ij} = \left(\sum_{p=1}^{Q-1} \left\| x_{jp} - c_{ip} \right\|^{\gamma} \right)^{\gamma^{-1}}$ is the distance of j^{th} class from i^{th} cluster, $\gamma \in [1, \infty)$ is the *g*eneralization parameter, $1 \le i \le C$, $1 \le j \le Q$ and $\tau \in (1, \infty)$ is fuzzifier.
      (c)Terminate the loop if $\left\| J(u,c)^{(t+1)} - J(u,c)^t \right\| < \vartheta$ for t > 1, where t reflects iteration and ϑ is pre-defined threshold.
      (d) Save U for the current run.
      (e) *run=run+1;*
   ii. Find the fitness value for r obtained partitions $\left\{ U_{(0)}, U_{(1)}, ..., U_{(r)} \right\}$ by using following fitness function[24]:
   $$£ = \frac{Q \times d_{\min}}{\sum_{i=1}^{C} \sum_{j=1}^{Q} u_{ij}^{\tau} d} \quad \text{where} \quad d = \sum_{i=1}^{C} \sum_{j=1}^{Q} \left\| \overline{x_j} - c_i \right\|^2$$
   iii. Select the population with the highest fitness value from step 2(*ii*).
   iv. Now the above obtained partition will act as initial population for the next generation, *g=g+1.*

3. The fuzzy distribution obtained for the last generation is referred to as optimal fuzzy distribution $U^{(opt)}$.

4. Now defuzzify the $U^{(opt)}$ to obtain the cluster allocation matrix (ς) of size $C \times \Omega$. In $U^{(opt)}$, we arrange all the classes in a cluster according to the degree of membership in descending order and obtain ς by selecting top membership rank elements equal to Ω, $\varsigma \leftarrow U^{(opt)}(1:C, 1:\Omega)$ where $\Omega < Q$.

C. Supervised Classification through NLA-SRBF NNs

The proposed classifier learn the inputs according to the cluster allocation matrix ς obtained from the unsupervised genetic fuzzy partitioning. In this paper, conventional neuron means the neuron with only summation aggregation and is referred to as Multi Layer Perceptron (MLP). MLP based network refers to the network which consists of conventional neurons both in its hidden as well as output layer. The proposed NLA-SRBF NN refers to the network which consists of novel *ČTROIKA* neurons in the hidden layer and complex conventional neurons in the output layer with the employment of ČRPROP learning algorithm. *ČTROIKA* is proposed as a neuron with summation and radial basis aggregation functions along with their compensatory product. The novelty in the aggregation function of the proposed neuron is to take advantage of the virtues of perceptron and radial basis processing. The sub-functions summation and radial basis are integrated non linearly ($\odot$) in the desired proportion ($\lambda : \psi$) in the proposed neuron model. The imprecision involved is taken care by the compensatory parameters λ and ψ which lay down the summation and radial basis contributions respectively.

a. Learning rules for ČTROIKA based classifier (NLA-SRBF NN)

This subsection presents the weight update equations of the proposed classifier for both ČBP and ČRPROP learning algorithms. A multilayer network NLA-SRBF is build with new neuron model and conventional neurons. Such a network is associated to each cluster. Consider a frequently used three layer structure *{L-M-N}(C)* where first layer has *L=Q-1* inputs, second layer has *M* proposed complex neurons *ČTROIKA*, third layer consists of *N=Ω* complex conventional neurons and *C* is number of clusters which reflect the number of associated networks. Complex proposed neuron *ČTROIKA* is computational efficient which ensures improved convergence speed and prediction precision of the proposed complex neural classifier. Here, all inputs and weights are considered to be complex numbers. An imaginary component of complex is added to the mean feature vector (*X*) of input classes which is considered to be the training set for this classifier. Thus, the training patterns are complex-valued where the imaginary part is negligible in comparison to the real counterpart. The complex input for the complex-valued network is $X = X + i \times .001$, where i is an imaginary unity. Conventionally, w_{lm} represents the weight from l^{th} neuron to m^{th} neuron. Let input vector be $Z = \{z_1, z_2, ... z_L\}$, $W^S{}_m = \{w^S{}_{1m}, w^S{}_{2m}, ..., w^S{}_{Lm}\}$ be the weights from inputs to the summation part of m^{th} proposed complex neuron and $W^{rb}{}_m = \{w^{rb}{}_{1m}, w^{rb}{}_{2m}, ..., w^{rb}{}_{Lm}\}$ be the weights from inputs to the radial basis part of m^{th} neuron. Let $W_0 = \{w_{01}, w_{02}, ..., w_{0M}\}$ be the bias weight vector and $z_{m0} = 1 + i \times .001$ be the bias input for M complex *ČTROIKA* neurons in the hidden layer. Let $W_n = \{w_{1n}, w_{2n}, ..., w_{Mn}\}$ be the weight vector of hidden neurons to n^{th} output neuron, $B_0 = \{b_{01}, b_{02}, ..., b_{0N}\}$ be the bias weight vector and $z_{n0} = 1 + i \times .001$ be the bias input for N complex conventional neurons in the output layer. If $h \odot y = 1 + h + y + h \times y$, then the proposed *ČTROIKA* neuron is defined as $Y = f_\Phi\left(\lambda \times W^S \times Z^T \odot \psi \times \exp\left(-\left\|Z - W^{rb}\right\|^2\right)\right)$ where f_Φ is complex-valued non-linear activation function whose derivative is represented as $f_\Phi{}'$ and $\left\|Z - W^{rb}\right\|^2 = \left(Z - W^{rb}\right) \times \left(Z - W^{rb}\right)^\dagger$. Here, superscripts † and T represents the complex conjugate and transpose respectively, $\Re$ and ξ represents the real and imaginary components of complex correspondingly. Let $V_m = \Re(V_m) + i\xi(V_m)$ be the net potential of m^{th} *ČTROIKA* neuron in hidden layer, then by the definition of $\odot$ operation:

$$V_m = w_{0m} z_{m0} + V_{m1} + V_{m2} + V_{m1} V_{m2}$$

$$\text{where } V_{m1} = \lambda_m W^S{}_m Z^T$$

$$\text{and } V_{m2} = \psi_m \exp\left(-\left\|Z - W^{rb}{}_m\right\|^2\right)$$

The output of m^{th} *ČTROIKA* neuron can be expressed as $Y_n = f_\Phi(V_n) = \Re(Y_n) + i\xi(Y_n)$. The net potential and output of n^{th} complex conventional neuron in output layer respectively can be given by:

$$V_n = \Re(V_n) + i\xi(V_n)$$

$$V_n = \sum_{m=1}^{M} w_{mn} Y_m + b_{0n} z_{n0}$$

$$\text{and } Y_n = f_\Phi(V_n) = \Re(Y_n) + i\xi(Y_n)$$

Let Y_d be the desired output, then the error at n^{th} output neuron can be computed as: $e_n = \Re\left(e_n\right) + i\xi\left(e_n\right) = Y_{dn} - Y_n$. The complex-valued cost function (MSE) can be given by: $E = \frac{1}{2} \sum_{n=1}^{N} e_n e_n^\dagger = \frac{1}{2} \sum_{n=1}^{N} \left[(\Re(e_n))^2 + (\xi(e_n))^2 \right]$

The updated weights can be obtained as: $w^{(t+1)} = w^{(t)} + \Delta w^{(t)}$, where t is iteration.

For ČBP algorithm

$$\Delta w^{(t)} = -\eta\left(\frac{\partial E}{\partial \Re(w)} + i \times \frac{\partial E}{\partial \xi(w)}\right)$$

where $0 < \eta < 1$

For ČRPROP Algorithm

Let the increase(μ^+) and decrease(μ^-) factors are defined as $0 < \mu^- < \mu^+ < 1.2$, the step size is initialized as $\Delta_0 = 0.1$.

Let $\frac{\partial E(t-1)}{\partial \Re(w)} = \frac{\partial E(t-1)}{\partial \xi(w)} = 0$ for $t = 1$, $\Re(\Delta_{\max}) = \xi(\Delta_{\max}) = \Delta_{\max}$, $\Re(\Delta_{\min}) = \xi(\Delta_{\min}) = \Delta_{\min}$ where $\Delta_{\max}$ and $\Delta_{\min}$ are maximum and minimum step size respectively.

There are 3 cases for weight update:

case1: if $\left(\frac{\partial E(t-1)}{\partial \Re(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)}\right) \times \left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) > 0$ then

$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) \times \left(\Re(\Delta(t)) + i \times \xi(\Delta(t))\right)$ where $\Re(\Delta(t)) = \min(\Re(\Delta(t-1)) \times \mu^+, \Re(\Delta_{\max}))$ and $\xi(\Delta(t)) = \min(\xi(\Delta(t-1)) \times \mu^+, \xi(\Delta_{\max}))$

case2: if $\left(\frac{\partial E(t-1)}{\partial \Re(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)}\right) \times \left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) < 0$ then

$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) \times \left(\Re(\Delta(t)) + i \times \xi(\Delta(t))\right)$ where $\Re(\Delta(t)) = \max(\Re(\Delta(t-1)) \times \mu^-, \Re(\Delta_{\min}))$ and $\xi(\Delta(t)) = \max(\xi(\Delta(t-1)) \times \mu^-, \xi(\Delta_{\min}))$ if $(E(t) > E(t-1))$ then $w^{(t+1)} = w^{(t)} - \Delta w^{(t-1)}$ and $\frac{\partial E(t)}{\partial w} = 0$

case3: if $\left(\frac{\partial E(t-1)}{\partial \Re(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)}\right) \times \left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) = 0$ then

$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)}\right) \times \left(\Re(\Delta(t)) + i \times \xi(\Delta(t))\right)$

where $\Re(\Delta(t)) = \xi(\Delta(t)) = \Delta_0$.

For the above learning algorithms viz. ČBP and ČRPROP, let $\frac{\partial E(t)}{\partial \Re(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} = A$. The value of A for various learning parameters is as follows:

For $w = w_{mn}$,

$$A = Y_m^{\dagger}\left(\Re\left(e_n\right) f_\Phi{}'(\Re(V_n)) + i\xi\left(e_n\right) f_\Phi{}'(\xi(V_n))\right)$$

For $w = b_{0n}$,

$$A = z_{n0}^{\dagger}\left(\Re\left(e_n\right) f_\Phi{}'(\Re(V_n)) + i\xi\left(e_n\right) f_\Phi{}'(\xi(V_n))\right)$$

For $w = w^S_{lm}$, $A = z_l^{\dagger} \lambda_m^{\dagger}\left(1 + V_{m2}^{\dagger}\right) K_m$

For $w = w^{rb}_{lm}$,

$$A = \begin{bmatrix} 2 \times \exp\left(-Z - W^{rb}_m{}^2\right)\left(z_l - w^{rb}_{lm}\right) \times \\ \begin{bmatrix} \Re\left(K_m\right)\left\{\Re(\psi_m)\left(1 + \Re\left(V_{m1}\right)\right) - \xi\left(\psi_m\right)\xi(V_{m1})\right\} \\ + \xi\left(K_m\right)\left\{\xi(\psi_m)\left(1 + \Re\left(V_{m1}\right)\right) + \Re\left(\psi_m\right)\xi(V_{m1})\right\} \end{bmatrix} \end{bmatrix}$$

For $w = \lambda_m$, $A = \left(W^S_m Z^T\right)^{\dagger}\left(1 + V_{m2}^{\dagger}\right) K_m$

For $w = \psi_m$, $A = \exp\left(-\left\|Z - W^{rb}_m\right\|^2\right)\left(1 + V_{m1}^{\dagger}\right) K_m$

For $w = w_{0m}$, $A = Z_{m0}^{\dagger} K_m$

where

$$\Re\left(K_m\right) = f_\Phi{}'\left(\Re\left(V_m\right)\right) \times \sum_{n=1}^{N} \left\{ \begin{array}{l} \Re\left(e_n\right) f_\Phi{}'\left(\Re\left(V_n\right)\right)\Re\left(w_{mn}\right) \\ + \xi\left(e_n\right) f_\Phi{}'\left(\xi\left(V_n\right)\right)\xi\left(w_{mn}\right) \end{array} \right\}$$

and

$$\xi\left(K_m\right) = f_\Phi{}'\left(\xi\left(V_m\right)\right) \times \sum_{n=1}^{N} \left\{ \begin{array}{l} \xi\left(e_n\right) f_\Phi{}'\left(\xi\left(V_n\right)\right)\Re\left(w_{mn}\right) \\ - \Re\left(e_n\right) f_\Phi{}'\left(\Re\left(V_n\right)\right)\xi\left(w_{mn}\right) \end{array} \right\}$$

Weights update take place recursively till the network is trained according to desired output and the optimized cost is conquered. This trained NN is further used to classify the test patterns.

b. Recognition/Classification

Trained networks classify the test patterns by using the function *Max of Max* of the outputs of each associated neural NN. For every input test pattern, we have obtained the $C \times \Omega$ output matrix. Let $O_k(NN_i)$ be the maximum output of i^{th} associated network for k^{th} test image and α be the column index corresponding to $O_k(NN_i)$. Cluster C is acquired corresponding to $Max_{i=1}^{C}\left(O_k(NN_i)\right)$. The resulting class R of a test pattern can be identified from cluster allocation matrix by using the obtained values of C and α . For every input pattern, the truthfulness of $\left[mod\left(k,Q\right) > 0 \,\&\&\left(R = mod\left(k,Q\right)\right)\right]$ condition confirms the acceptance of k^{th} test pattern, where Q is the number of classes to be tested.

Algorithm for supervised classification

> 1. Train the network for $X = X + i \times .001$ inputs, for which input-output mapping is known in the form of ς, according to the learning rules presented in section III(C(a)).
> 2. Test set is fetched from $D = D + i \times .001$, which is then classified by the trained network as explained in section III(C(b)).

IV. Evaluation of New Learning Machine

In order to assess the potency and efficacy of the proposed hybrid learning machine, the performance evaluation has been carried out on three benchmark face datasets viz. extended Cohn-Kanade dataset [25], FERET face dataset [26] and AR face dataset [27]. Standard biometric measures, prediction accuracy, FAR(False acceptance rate) and FRR (False rejection rate), number of learning cycles (with the assumption that time taken by one learning cycle is unity) are considered for performance estimation. The cross-validation is performed on the considered datasets and the results are averaged over. For testing of a class, an input image of same class is considered as a positive case whereas images of other classes are considered as negative case. The performance of the proposed machine is accounted for the following eight combinations of hidden neuron and learning procedure: (i) ŘMLP with ŘBP (ii) ŘTROIKA with ŘBP (iii) ŘMLP with ŘRPROP (iv) ŘTROIKA with ŘRPROP (v) ČMLP with ČBP (vi) *ČTROIKA* with ČBP (vii) ČMLP with ČRPROP (viii) *ČTROIKA* with ČRPROP. It is worth mentioning here that output neurons in all the classifiers are complex conventional neurons. The computational supremacy and generalization skill for all the above mentioned classifiers have been compared in terms of number of learning cycles, training and testing errors, recognition rates, FAR, FRR and network topology with respect to number of hidden neurons.

A. Performance with Extended Cohn-Kanade Dataset

The extended version of Cohn-Kanade (CK+) dataset is used to evaluate the performance of the proposed learning machine where we considered 1230 face images of 123 subjects. All images are of same size with bright homogenous background along with wide variations in expression and emotion. Few example images are shown in Fig.1. Here, 4 images per subject are considered for training and 6 images per subject are used for testing. Thus, 492 images constitutes the training set and 738 images make up the test set.

Fig.1. Example images from CK+ face dataset

The learning graphs of considered dataset with different hidden neuron and learning algorithm are shown in Fig.2. It is clearly visible from the plots that *ČTROIKA* with ČRPROP gives the fastest convergence among all classifiers. The local minima crisis in real domain can also be observed from the corresponding graphs in Fig.2 whereas with complex-valued neural network this problem get trounced. The simulation results for this database with different classifiers are presented in Table1 which reflects the best plausible results obtained for various classifiers. From extensive simulations on this dataset following inferences can be made:

- Proposed neuron *ČTROIKA* with ČRPROP gives the best results in terms of convergence, number of learning cycles, recognition rate, FAR, FRR, training and testing errors.
- *ČTROIKA* with ČRPROP outperforms ČMLP with ČRPROP significantly in terms of network topology and the similar outcome is well-founded for real domain as well.
- *ČTROIKA* with ČBP perform considerably well than ČMLP with ČBP in terms of network topology and learning cycles and the identical upshot is applicable for real domain.

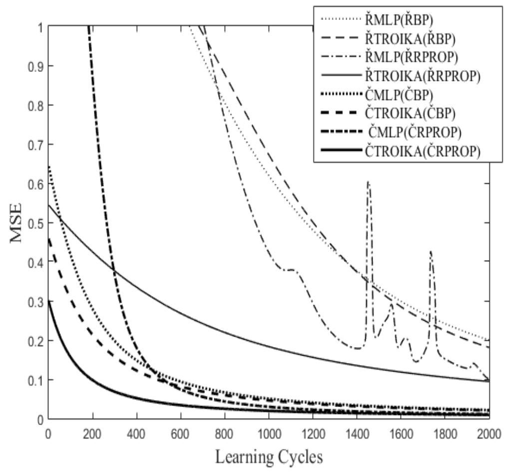

Fig.2. Learning Graphs in CK+ face dataset with different hidden neuron and learning algorithm

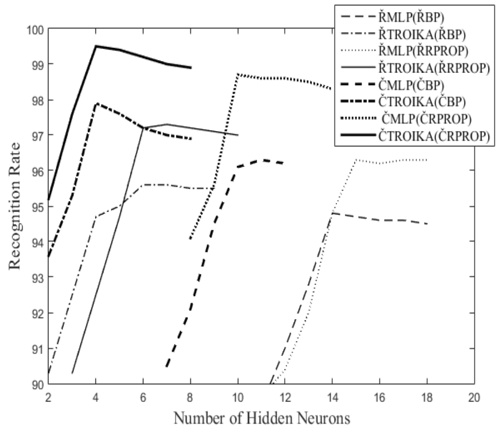

Fig.3. Recognition rate vs number of hidden neurons for different hidden neuron and learning algorithm in CK+ face dataset

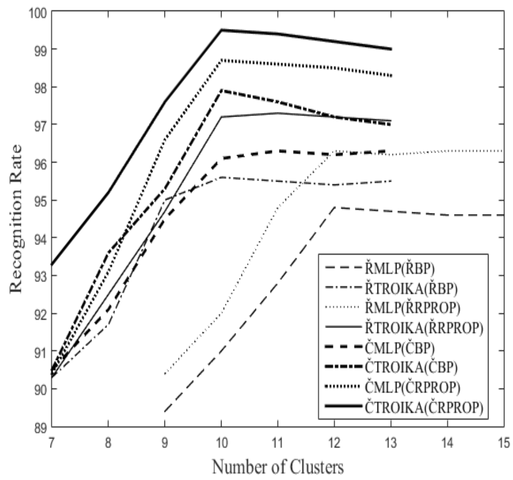

Fig.4. Recognition rate vs number of clusters for different hidden neuron and learning algorithm in CK+ face dataset

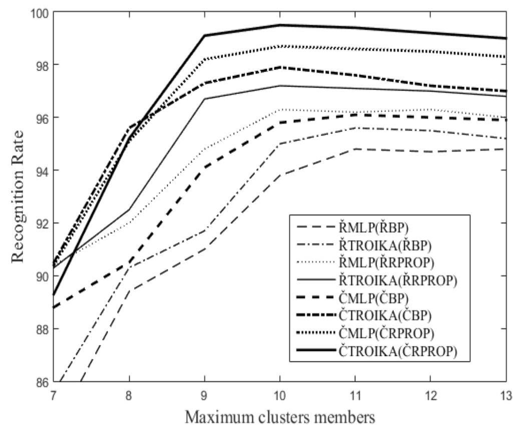

Fig.5. Recognition rate vs maximum cluster members for different hidden neuron and learning algorithm in CK+ face dataset

From Table1 it is evidenced that *ČTROIKA* is superior than ČMLP and ŘTROIKA is an improvement over ŘMLP. It is also evidenced that RPROP gives enhanced performance over BP both in real and complex domain. Complex domain results are superior than that of real domain both for RPROP and BP. In order to observe the effect of number of hidden neurons on the recognition accuracy for different classifiers, the comparative performance is given in Fig.3. The curves shows the similar nature for all the considered classifiers but with different number of hidden neurons. Fig.3 shows the plots for different classifiers where it is observed that accuracy increases with increase in number of hidden neurons but up to a certain number beyond which there is no further improvement in the recognition rate. Least number of hidden neurons are required for the proposed neuron *ČTROIKA* with ČRPROP learning based classifier which results in reduced computations and compact network topology. It is observed that for the same learning algorithm, *ČTROIKA* based classifier perform well with lesser number of hidden neurons while ČMLP based classifier requires adequately larger number of hidden neurons to achieve the nearly same accuracy. The above statement is factual for real domain also. The variation effect of number of clusters on the recognition rate is depicted in Fig.4 which again demonstrate the outperformance of *ČTROIKA* with ČRPROP over other classifiers. The behaviour of the curve for different neuron and learning algorithm imitates that recognition precision increases with increase in number of clusters up to some extent however no enhancement in accuracy is observed afterwards. Fig.5 presents the accuracy variations captured for maximum cluster members where it is observed that performance is almost directly proportional to the maximum cluster members up to some degree but no improvement is recorded later. The proposed classifier with *ČTROIKA* and ČRPROP again gives the best performance with respect to other combinations of neuron and learning algorithm.

Table 1. Comparisons of training and testing performance for CK+ Dataset with different hidden neuron and learning algorithm

Hidden Neuron	ŘMLP		ŘTROIKA		ČMLP		*ČTROIKA*	
Algorithm	ŘBP	ŘRPROP	ŘBP	ŘRPROP	ČBP	ČRPROP	ČBP	ČRPROP
Network	{122-15-11} (12)	{122-15-10} (12)	{122-6-11} (10)	{122-6-10} (10)	{122-10-11} (10)	{122-10-10} (10)	{122-4-10} (10)	{122-4-10} (10)
MSE	.087762	.03335	.01368	.00566	.00862	.00099	.00186	.00062
Learning cycles	23000	9000	14000	6000	8000	3500	5500	2000
FAR	0.058	0.050	0.047	0.040	0.047	0.036	0.026	0.021
FRR	0.22	0.19	0.21	0.17	0.19	0.17	0.18	0.16
Accuracy	94.9%	96.5%	95.6%	97.2%	96.1%	98.7%	97.9%	99.6%

B. Performance with FERET Face Dataset

FERET is one of the largest publicly available database which contains face images of 1208 subjects. We consider a subset of 120 subjects with 1200 images to evaluate the performance of the proposed learning machine. 480 images are considered for training and 720 images are engaged in testing. Images in the database possesses pose variations. As an example some random images are shown in Fig.6. The proposed learning machine copes well with the pose variation problem and gives accuracy of 99.98 % in the presence of pose variations. Fig.7 presents the plots of learning for this dataset with different combination of hidden neuron and learning algorithm.

Fig.6. Example images from FERET face dataset

Among different combinations, the superiority of *ČTROIKA* with ČRPROP again demonstrated in this experiment as this classifier optimize the error faster than all other considered networks. The best obtained test results for FERET dataset with different classifiers are revealed in Table 2.

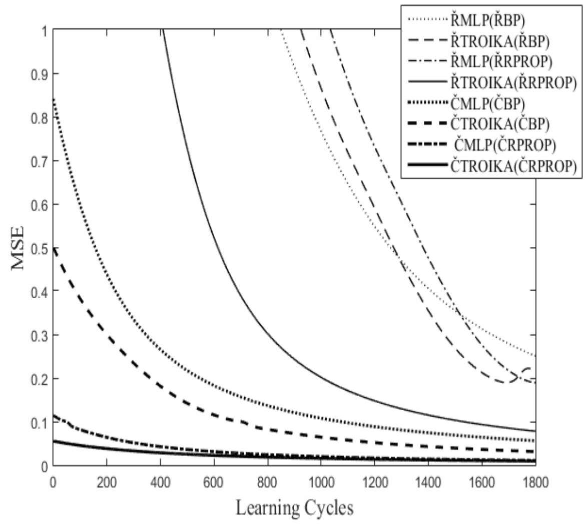

Fig.7. Learning Graphs in FERET face dataset with different hidden neuron and learning algorithm

Table 2. Comparisons Of Training And Testing Performance For FERET Face Dataset With Different Hidden Neuron And Learning Algorithm

Hidden Neuron	ŘMLP		ŘTROIKA		ČMLP		*ČTROIKA*	
Algorithm	ŘBP	ŘRPROP	ŘBP	ŘRPROP	ČBP	ČRPROP	ČBP	ČRPROP
Network	{119-14-18} (12)	{119-14-18} (12)	{119-5-15} (12)	{119-5-15} (12)	{119-10-16} (12)	{119-10-16} (12)	{119-3-14} (12)	{119-3-14} (12)
MSE	.07853	.04102	.01525	.00598	.00792	.00101	.00169	.00058
Learning cycles	20000	8500	12000	6000	7500	4500	5000	1800
FAR	0.47	0.33	0.31	0.19	0.27	0.25	0.26	0.18
FRR	0.32	0.29	0.23	0.27	0.29	0.27	0.25	0.20
Accuracy	94.4%	95.3%	95.92%	97.9%	96.7%	99.2%	97.5%	99.98%

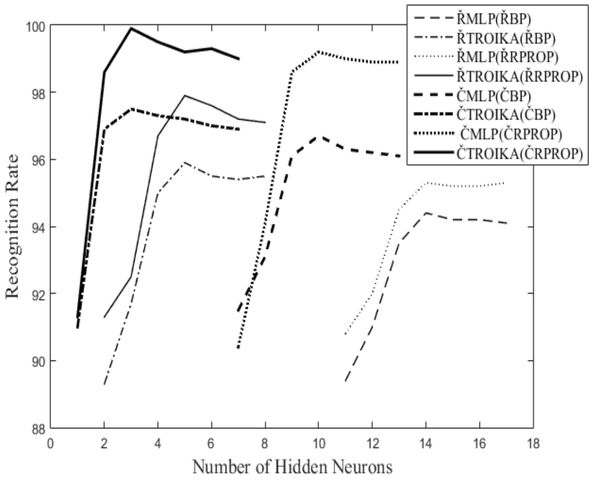

Fig.8. Recognition rate vs number of hidden neurons for different hidden neuron and learning algorithm in FERET face dataset

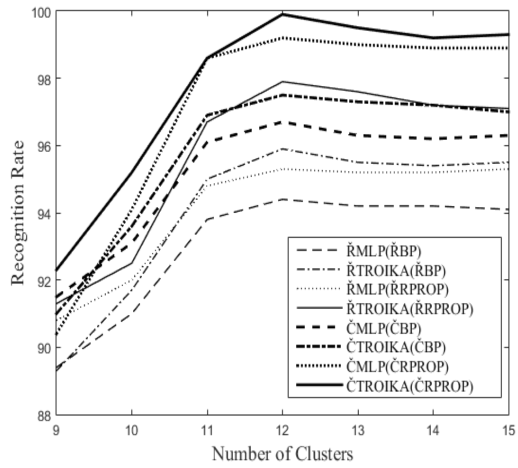

Fig.9. Recognition rate vs number of clusters for different hidden neuron and learning algorithm in FERET face dataset

Different performance measures evidently exhibit that the projected hybrid learning machine in complex domain offers enhanced performance in all respects. Classifiers based on *ČTROIKA* and ŘTROIKA neuron gives effectively improved prediction accuracy with compact network topology in much less number of training cycles as compared to ČMLP and ŘMLP respectively.

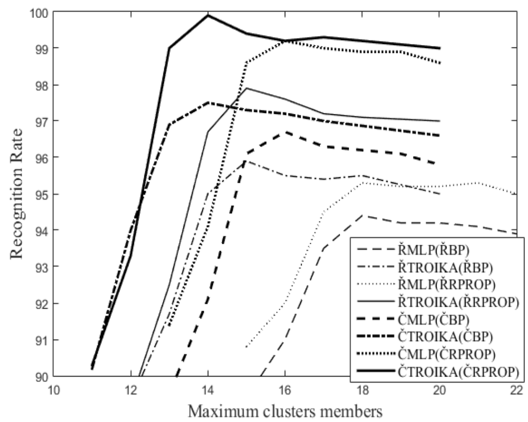

Fig.10. Recognition rate vs maximum cluster members for different hidden neuron and learning algorithm in FERET face dataset

The recognition accuracy with respect to number of hidden neurons for different classifiers are given in Fig.8. For every permutation, the maximum rate occurs at different number of hidden neurons and beyond that value the rate becomes nearly constant with no performance enhancement. The proposed classifier achieves best accuracy among all with sufficiently fewer number of hidden neurons than others which again demonstrate the effectiveness of *ČTROIKA* based network with ČRPROP training algorithm. It can be observed from Fig.8 that *ČTROIKA* and ŘTROIKA based classifiers gives enhanced performance with much less number of hidden neurons as compared to ČMLP and ŘMLP respectively for the same training procedure. Fig.9 presents the impact of number of clusters on the recognition accuracy where it can be examined that rate goes on increasing with number of clusters up to some level but degrades later. In this dataset, with the same number of clusters, *ČTROIKA* with ČRPROP gives the best performance among other classifiers. Fig.10 shows the accuracy precision with regard to maximum cluster members for dissimilar neuron and learning algorithm. The behavior of the curves disclose that up to some extent the value of maximum number of classes allowed in a cluster enhances the performance, but degrades subsequently. *ČTROIKA* and ŘTROIKA based classifier performs better than their MLP counterparts respectively.

C. Performance with AR Face Dataset

This dataset consists of face images of 126 individuals with more than 4000 color images. The images are acquired in two sessions separated by two weeks. In this face dataset, images are with major distinctions including illumination conditions, expressions and facial concealment. A subset of 1400 images of 100 individuals with different facial expressions and illumination is considered for measuring the performance of the proposed learning machine. 700 images from session 1 are used to train the system while 700 images from session 2 are used for testing. Fig.11 shows the example images having variations in expressions and illumination.

Fig.11. Example images from AR face dataset

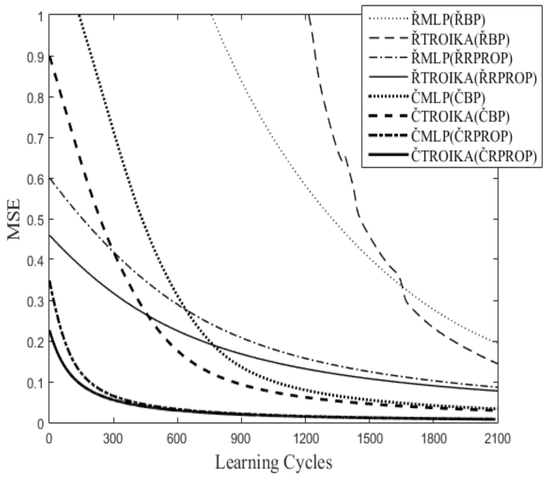

Fig.12. Learning Graphs in AR face dataset with different hidden neuron and learning algorithm

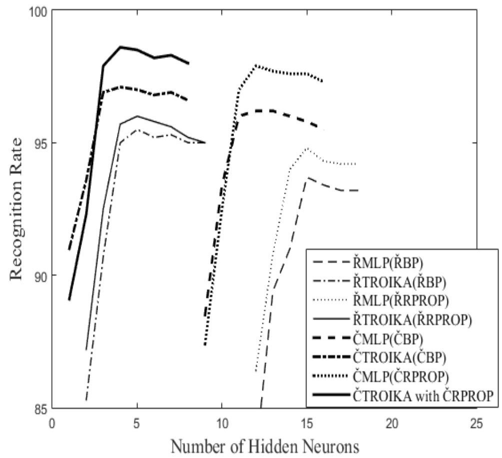

Fig.13. Recognition rate vs number of hidden neurons for different hidden neuron and learning algorithm in AR face dataset

Fig.12 illustrates the comparative performance on the basis of learning graphs for AR dataset with different permutations of hidden neuron and learning procedure. The dominance of the proposed combination of *ČTROIKA* with ČRPROP once again evidenced on the basis of fastest convergence among all. The best performance results of all considered variations of classifiers for this dataset are presented in Table 3. The supremacy of the proposed *ČTROIKA* neuron again exposed as the corresponding classifier realized the better accuracy with smaller network topology in adequately less learning cycles as compared to ČMLP for the common learning algorithm.

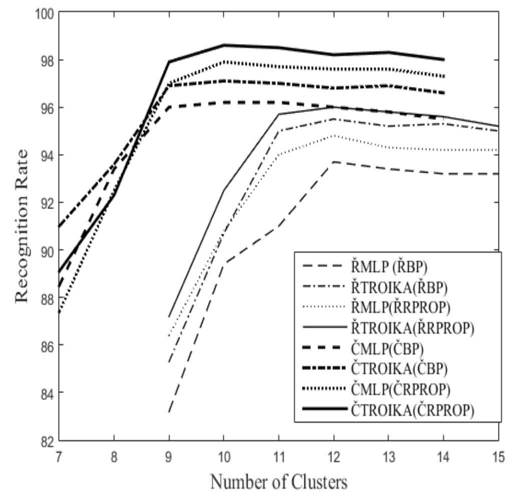

Fig.14. Recognition rate vs number of clusters for different hidden neuron and learning algorithm in AR face dataset

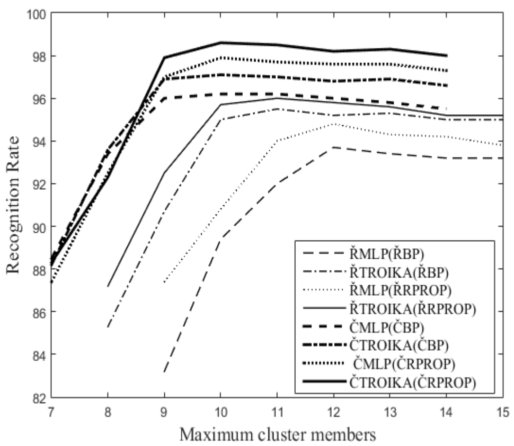

Fig.15.Recognition rate vs maximum cluster members for different hidden neuron and learning algorithm in AR face dataset

All together, the aforesaid assertion is over and above applicable to real domain. The noteworthy feature of Table 3 is that the proposed classifier requires least training cycles with superior recognition rate. The value of FAR is observed quite low in *ČTROIKA* based classifier which makes the system more effective specifically in real life situations. The analysis of different classifiers as a function of number of hidden neurons is presented in Fig.13. It is obtained from the respective curves of different classifiers that *ČTROIKA* with ČRPROP is the best performing classifier which outperforms over all other methods. *ČTROIKA* and ŘTROIKA based classifier performs significantly better when we consider lesser number of hidden neurons than ČMLP and ŘMLP based classifier respectively. The accuracy goes on increasing with number of hidden neurons but up to some extent, moreover it starts degrading later with further increase in number of hidden neurons. The performance comparisons with different hidden neuron and learning algorithm are presented in Fig.14 in order to monitor the recognition rate with respect to number of clusters. It is again observed that recognition rate with the proposed combination *ČTROIKA* with ČRPROP is better than any other considered classifier. With the increase in clusters, performance raises initially although it becomes nearly constant with no further improvement with more increase in clusters after some extent. One can examine the effect of variation in maximum cluster members on the accuracy precision of the proposed algorithm from Fig.15. The curves follow more or less similar manner for all the classifiers. *ČTROIKA* with ČRPROP gained best accuracy among all which again illustrate the supremacy of the proposed classifier. Up to some extent recognition rate increases but no enhancement is noticed later with further increase in number.

Table 3. Comparisons of training and testing performance for AR Face Dataset with different hidden neuron and learning algorithm

Hidden Neuron	ŘMLP		ŘTROIKA		ČMLP		*ČTROIKA*	
Algorithm	ŘBP	ŘRPROP	ŘBP	ŘRPROP	ČBP	ČRPROP	ČBP	ČRPROP
Network	{99-14-20} (12)	{99-14-20} (12)	{99-5-14} (11)	{99-5-14} (11)	{99-12-16} (10)	{99-12-16} (10)	{99-4-12} (10)	{99-4-12} (10)
MSE	.09335	.03996	.01689	.00642	.00935	.000114	.00278	.00077
Learning cycles	24000	10000	15000	7000	8000	5500	6500	2100
FAR	0.48	0.38	0.33	0.30	0.29	0.22	0.23	0.21
FRR	0.66	0.42	0.46	0.41	0.38	0.27	0.26	0.25
Accuracy	93.7%	94.8%	95.5%	96.0%	96.2%	97.9%	97.1%	98.6%

V. Conclusion

In this paper, a hybrid learning machine is presented which is a novel synergism of statistical techniques (eigenface with fisherface method), genetic fuzzy distribution and complex neural classifier. The classifier is build with proposed hidden neurons *ČTROIKA* and output MLP neurons along with ČRPROP learning algorithm. A detailed comparative analysis has been carried out with ČMLP based classifiers as well as with real domain classifiers. The remarkable achievement of the proposed recognizer is its better recognition accuracy with speedy convergence and compact topology with respect to hidden neurons. The proposed hybrid learning machine is robust for unauthorized cases but also stringent for authorized cases which has been demonstrated by performance evaluation at very low FAR and FRR presented in this paper. Eight variations of classifiers are compared on three benchmark face datasets. It has been examined from the results that the relative performance of classifiers never changes. As we shifted from real domain to complex domain, the performance is significantly improved with the proposed classifier although the computations increases with complex numbers. *ČTROIKA* and ŘTROIKA always perform better in all respects than ČMLP and ŘMLP respectively with the same learning algorithm. ČRPROP always perform better than ČBP in terms of convergence, learning cycles and recognition rate which hold for real domain simultaneously.

From extensive experiments and analysis it has been established that our learning machine is robust enough to recognize the human identity even with images possesses wide variations in poses, facial expressions and illumination conditions. Thus, it can perform efficiently in real environment. Finally, we wind up with the conclusion that the proposed learning machine is well-organized recognition system where the novel neuron *ČTROIKA* with ČRPROP learning algorithm based classifier offers best recognition accuracy with speedy convergence and compact topology among other considered classifiers.

References

[1] Vijayakumari V. Face recognition techniques: A survey. World journal of computer application and technology, 2013, 1(2):41-50.

[2] Islam MN, Loo CK. Geometric feature-based facial emotion recognition using two-stage fuzzy reasoning model. In International Conference on Neural Information Processing, 2014, pp. 344-351, Springer, Cham.

[3] Li B, Meng MQ. Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. IEEE Transactions on Information Technology in Biomedicine, 2012, 16(3):323-329.

[4] Shi B, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(11):2298-2304.

[5] Li CK. A sigma-pi-sigma neural network (SPSNN). Neural Processing Letters, 2003, 17(1):1-9.

[6] Epitropakis MG, Plagianakos VP, Vrahatis MN. Hardware-friendly higher-order neural network training using distributed evolutionary algorithms. Applied Soft Computing, 2010, 10(2):398-408.

[7] HOMMA N, GUPTA MM. A General Second-order Neural Unit. Bulletin of College of Medical Sciences, Tohoku University, 2002, 11(1):1-6.

[8] Chaturvedi DK, Mohan M, Singh RK, Kalra PK. Improved generalized neuron model for short-term load forecasting. Soft Computing, 2004, 8(5):370-379.

[9] Tripathi BK, Chandra B, Kalra PK. The generalized product neuron model in complex domain. In International Conference on Neural Information Processing, 2008, pp. 867-876, Springer, Berlin, Heidelberg.

[10] Zhang M, Xu S, Fulcher J. Neuron-adaptive higher order neural-network models for automated financial data modeling. IEEE Transactions on Neural Networks, 2002, 13(1):188-204.

[11] Tripathi BK, Kalra PK. On efficient learning machine with root-power mean neuron in complex domain. IEEE transactions on neural networks, 2011, 22(5):727-738.

[12] Kumar S, Tripathi BK. Root-Power Mean Aggregation-Based Neuron in Quaternionic Domain. IETE Journal of Research, 2018, pp.1-19.

[13] Zeiler MD. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701. 2012 .

[14] Chen X, Tang Z, Li S. An Modified Error Function for the Complex-value Backpropagation. Neural Information Processing-Letters and Reviews, 2005 , 8(1).

[15] Luo X, Patton AD, Singh C. Real power transfer capability calculations using multi-layer feed-forward neural networks. IEEE Transactions on Power Systems, 2000, 15(2):903-908.

[16] Aizenberg I, Paliy DV, Zurada JM, Astola JT. Blur identification by multilayer neural network based on multivalued neurons. IEEE Transactions on Neural Networks, 2008, 19(5):883-898.

[17] Hirose A. Complex-valued neural networks. Springer Science & Business Media, 2012.

[18] Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In IEEE International Conference on Neural Networks, 1993, pp. 586-591.

[19] Wang Z, Jiang M, Hu Y, Li H. An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. IEEE Transactions on Information Technology in Biomedicine, 2012, 16(4):691-699.

[20] Oh SK, Yoo SH, Pedrycz W. Design of face recognition algorithm using PCA-LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks: Design and its application. Expert Systems with Applications, 2013, 40(5):1451-1466.

[21] Liu C, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image processing, 2002, 11(4):467-476.

[22] Er MJ, Wu S, Lu J, Toh HL. Face recognition with radial basis function (RBF) neural networks. IEEE transactions on neural networks, 2002, 13(3):697-710.

[23] Khan SS, Quadri SM, Peer MA. Genetic Algorithm for Biomarker Search Problem and Class Prediction. International Journal of Intelligent Systems and Applications, 2016, 8(9):47.

[24] Sheng W, Howells G, Fairhurst M, Deravi F. Template-free biometric-key generation by means of fuzzy genetic clustering. IEEE Transactions on Information Forensics and Security, 2008, 3(2):183-191.

[25] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94-101.

[26] Feret face database. https://www.nist.gov/itl/iad/image-group/color-feret-database.

[27] Martinez AM. The AR face database. CVC Technical Report24. 1998.

Authors' Profiles

Swati Srivastava is pursuing her Ph.D in Computational Intelligence from HBTU Kanpur, India and completed M.Tech in Computer Science from NIT Allahabad, India. She is currently research scholar in the Department of Computer Science and Engineering of HBTU Kanpur, India. Her areas of research include high-dimensional neurocomputing, computational intelligence, machine learning and computer vision focused on biometrics.

Bipin K. Tripathi completed his Ph.D in Computational Intelligence from IIT Kanpur, India and M. Tech in Computer Science and Engineering from IIT Delhi, India. He is currently serving as a Professor in the Department of Computer Science and Engineering of HBTU Kanpur, India. He is also leading the Nature-inspired Computational Intelligence Research Group (NCIRG) at HBTU. His areas of research include high-dimensional neurocomputing, computational neuroscience, intelligent system design, machine learning and computer vision focused on biometrics and 3D Imaging. He has published several research papers in these areas in many peer reviewed journals including IEEE Transaction, Elsevier, Springer and other international conferences. He has also contributed book chapters in different international publications and patent in his area. He is continuously serving as PC for many international conferences and as a reviewer of several international journals.

How to cite this paper: Swati Srivastava, Bipin K. Tripathi, "Development of Hybrid Learning Machine in Complex Domain for Human Identification", International Journal of Intelligent Systems and Applications(IJISA), Vol.11, No.1, pp.55-66, 2019. DOI: 10.5815/ijisa.2019.01.06