Development of Hybrid Learning Machine with Complex Domain Neurocomputing

Swati Srivastava[1]*   Bipin Kumar Tripathi[1]

[1]*Department of Computer Science & Engineering, Harcourt Butler Technical University, Kanpur, India*
* Corresponding author's Email: swati.cs1409@gmail.com

Abstract: This paper presents a hybrid learning machine for human identification which is an amalgamation of eigenface method, fisherface method, genetic fuzzy distribution and complex neural network. This hybrid machine constitutes multiple modules including the image representation in lower dimensional feature space, unsupervised clustering and supervised classification. The classification module employs proposed complex neuron structure ČTROIKA as hidden neurons, complex conventional neuron (ČMLP) as output neurons along with complex resilient propagation (ČRPROP) learning procedure. The efficacy and potency of the proposed classifier based hybrid learning machine demonstrated on two benchmark biometric datasets-- FERET and AR face datasets which shows that our learning machine outperforms over state-of-the-art methods. Further, the classification module in the proposed learning machine is replaced by different classifiers to carry out the comparative analysis of various classifiers. The variations in classifier is based on the type of hidden neuron and learning algorithm. The performance comparisons of different classifiers reflects the superiority of the proposed classifier in terms of accuracy, speedy convergence, better learning with reduced learning cycles.

Keywords: Eigenface, Fisherface, Genetic fuzzy partitioning, Complex resilient propagation, Biometric.

1. Introduction

Over the years extensive attempts have been made in human recognition domain for the purpose of identification and authentication. This paper focuses on neural network (NN) based methods [1, 2]. NN with conventional neurons is widely used classifier but it could not offer outstanding performance in terms of accuracy and learning cycles. To improve the efficiency of NNs broad attempts have been made. One of those attempts is to develop higher order neuron structures which contributes pi-sigma [3, 4], second order neurons [5], generalized neurons [6, 7] and higher order neurons [8, 9]. Among above contributions, higher order neurons have evidenced to be the 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 augmentation of higher order neurons is growing continuously which contributed high performance in

many classification models. In extension to the above progression, this paper presents a higher order neuron *ČTROIKA* in complex domain with the ability to capture higher order correlations among input patterns. We present a network built up on novel neurons and complex conventional neurons. Moreover, 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 given like addition of momentum term [10], modified error function [11] and Quick Prop [12]. But none of these variations was capable to make significant improvement in convergence rate. The fast convergence with effective performance along with less complexity of neural network are the demanding factors for variety of applications. Additional attempts in the direction of performance enhancement shifted the classifier from real domain to complex domain. The complex

back propagation (ČBP) reduces the probability of getting trapped in local minima and improves the convergence rate. ČBP outperforms ŘBP for both real [13] and complex [14] valued problems. With widespread efforts of researchers towards the performance advancement of recognizer, RPROP was developed [15] as an improvement upon BP algorithm. With the employment of higher order neurons RPROP outperforms BP by providing faster training and better prediction accuracy. Additional advantage with RPROP is that it does not require to specify any free parameter values such as the learning rate and an optional momentum term as opposed to BP. In our previous work [16] we have developed a deep model in real domain and ŘBP learning algorithm was used in the classification segment. In this work we use complex domain neurocomputing and ČRPROP learning algorithm instead of ŘBP algorithm which drastically accelerated the performance of the recognition system. This paper presents a hybrid learning machine that comprises of statistical techniques (eigenface with fisherface method) incorporated with genetic fuzzy distribution and complex neural classifier. The new classifier exploit modified ČRPROP algorithm along with novel neuron model *ČTROIKA* in order to boost the recognition accuracy with reduced learning cycles.

Rest of the paper is organized as: Section 2 elaborates the proposed hybrid learning machine together with the formal algorithm. Section 3 is devoted to the performance evaluation of the proposed learning machine. Finally, we conclude the work in Section 4.

2. 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. This unsupervised clustering of input classes act as the input-output mapping for the classification segment. The new classifier incorporates hidden layer which consists of novel 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 the 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 representation of database images in lower dimensional feature space, unsupervised genetic fuzzy distribution and supervised classification.

2.1 Lower dimensional feature space representation of database images

The performance of any image recognition system profoundly depends upon the choice of features of image to be processed. Eigenface paradigm [17, 18] is considered in this paper to extract features which represents the image with reduced data that 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, we use fisherface[19] paradigm which is widely used discrimination criterion that overcomes the limitations of eigenfaces while retaining the idea of projection from high dimensional feature space to extensively lower dimensional feature space.

2.2 Unsupervised genetic fuzzy distribution

This is the intermediate step of the proposed learning machine in order to get the desired output for supervised classification. Let P be the total number of images in the dataset, Q be the total number of classes and S be the number of images in each class. Once dataset images in lower dimensional feature space are obtained, average image of each class is calculated by considering training set containing 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 independent product of fuzzy algorithm is not sufficient to obtain the optimal distribution. 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, and the partition with highest fitness value [20] 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 optimized partition matrix (size $C \times \Omega$) is the desired output for the classifier which is trained accordingly to classify the test images.

2.3 Supervised classification

The proposed classifier learn the inputs according to the optimized partition obtained from the unsupervised genetic fuzzy partitioning. In this paper, conventional neuron (MLP) means the neuron with only summation aggregation. MLP based network refers to the network which consists of conventional neurons both in its hidden as well as output layer. The proposed $\check{C}TROIKA$ based network refers to the network which consists of novel $\check{C}TROIKA$ neurons in the hidden layer and complex conventional neurons in the output layer. $\check{C}TROIKA$ is proposed as a complex 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 (Θ) in the desired proportion (λ:Ψ) in the complex 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.

2.3.1. Weight update regulations for $\check{C}TROIKA$ based classifier both for ČBP and ČRPROP learning algorithms

A multilayer network is built 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 the first layer has $L = Q - 1$ inputs, second layer has M proposed complex neurons $\check{C}TROIKA$, third layer consists of $N = \Omega$ complex conventional neurons and C is the number of clusters which reflect the number of associated networks. Complex proposed neuron $\check{C}TROIKA$ 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 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 its real counterpart. Thus, 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 $Z = \{z_1, z_2, \ldots, z_L\}$, $W^s{}_m = \{w^s{}_{1m}, w^s{}_{2m}, \ldots, w^s{}_{Lm}\}$ be the weights from inputs to the summation part of mth proposed complex neuron and $W^{rb}{}_m = \{w^{rb}{}_{1m}, w^{rb}{}_{2m}, \ldots, w^{rb}{}_{Lm}\}$ be the weights from inputs to the radial basis part of mth neuron. Let $W_0 = \{w_{01}, w_{02}, \ldots, w_{0M}\}$ be the bias weight vector and $zm_0 = 1 + i \times .001$ be the bias input for M complex $\check{C}TROIKA$ neurons in the hidden layer. Let $W_n = \{w_{1n}, w_{2n}, \ldots, w_{Mn}\}$ be the weight vector of hidden neurons to n^{th} output neuron, $B_0 = \{b_{01}, b_{02}, \ldots, b_{0N}\}$ be the bias weight vector and $zn_0 = 1 + i \times .001$ be the bias input for N complex conventional neurons in the output layer. If $A \odot G = 1 + A + G + AG$, then the considered $\check{C}TROIKA$ neuron is defined as:

$$Y = f_\phi \left((\lambda \times W^s \times Z^T) \odot \left(\Psi \times \exp \left(-\|Z - W^{rb}\|^2 \right) \right) \right) \quad (1)$$

Here, f_ϕ is complex-valued non-linear activation function whose derivative is represented as $f_\phi{}'$ and $\|Z - W^{rb}\|^2 = (Z - W^{rb}) \times (Z - W^{rb})^\dagger$. The superscript $(.)^\dagger$ represents the conjugate of a complex number and $(.)^T$ represents the transpose of a matrix. Ʀ and ξ represents the real and imaginary components of complex correspondingly. Let $V_m = ƦV_m + i\xi V_m$ be the net potential of mth $\check{C}TROIKA$ neuron in hidden layer, then by the definition of Θ operation:

$$V_m = w_{0m} zm_0 + V_{m1} + V_{m2} + V_{m1}V_{m2} \quad (2)$$
$$\text{where } V_{m1} = \lambda_m \times W^s{}_m \times Z^T$$
$$\text{and } V_{m2} = \Psi_m \times \exp(-\|Z - W^{rb}{}_m\|^2)$$

The output of mth $\check{C}TROIKA$ neuron can be expressed as:

$$Y_m = f_\phi(V_m) = ƦY_m + i\xi Y_m \quad (3)$$

The net potential and output of nth complex conventional neuron in output layer respectively can be given by:

$$V_n = ƦV_n + i\xi V_n$$
$$V_n = \sum_{m=1}^{M} w_{mn} Y_m + b_{0n} zn_0 \quad (4)$$

$$\text{and } Y_n = f_\phi(V_n) = Ʀ(Y_n) + i\xi(Y_n) \quad (5)$$

Let Y_d be the desired output, then the error at n^{th} output neuron can be computed as:

$$e_n = Y_{dn} - Y_n = R(e_n) + i\xi(e_n) \qquad (6)$$

The complex-valued cost function (MSE) can be given by:

$$E = \frac{1}{2} \sum_{n=1}^{N} e_n\, e_n{}^{\dagger}$$
$$= \frac{1}{2} \sum_{n=1}^{N} \left[\left(R(e_n)\right)^2 + \left(\xi(e_n)\right)^2 \right] \quad (7)$$

The updated weights can be obtained as:

$$w^{(t+1)} = w^{(t)} + \Delta w^{(t)} \qquad (8)$$

where t is iteration.

For ČBP algorithm

$$\Delta w^{(t)} = -\eta \left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) \qquad (9)$$

where η is learning parameter and is defined as $0 < \eta < 1$.

For ČRPROP Algorithm

Let the increase factor (μ^+) and decrease factor (μ^-) are defined as $0 < \mu^- < \mu^+ < 1.2$, the step size is initialized as $\Delta_0 = 0.1$.
For t=1, $\frac{\partial E(t-1)}{\partial R(w)} = \frac{\partial E(t-1)}{\partial \xi(w)} = 0$
Let $R(\Delta_{max}) = \xi(\Delta_{max}) = \Delta_{max}$,
$R(\Delta_{min}) = \xi(\Delta_{min}) = \Delta_{min}$
where Δ_{max} and Δ_{min} are maximum and minimum step size respectively.
There are 3 cases for weight update:

case1: if

$$\left(\frac{\partial E(t-1)}{\partial R(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)} \right) \times \left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) > 0$$

then

$$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) \times \left(R(\Delta(t)) + i \times \xi(\Delta(t)) \right) \qquad (10)$$

where $R(\Delta(t)) = \min(R(\Delta(t-1)) \times \mu^+, R(\Delta_{max}))$
and $\xi(\Delta(t)) = \min(\xi(\Delta(t-1)) \times \mu^+, \xi(\Delta_{max}))$

case2: if

$$\left(\frac{\partial E(t-1)}{\partial R(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)} \right) \times \left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) < 0$$

Then

$$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) \times (R(\Delta(t)) + i \times \xi(\Delta(t))) \qquad (11)$$

where $R(\Delta(t)) = \max(R(\Delta(t-1)) \times \mu^-, R(\Delta_{min}))$

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

case3: if

$$\left(\frac{\partial E(t-1)}{\partial R(w)} + i \times \frac{\partial E(t-1)}{\partial \xi(w)} \right) \times \left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) = 0$$

Then

$$\Delta w^{(t)} = -sign\left(\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} \right) \times \left(R(\Delta(t)) + i \times \xi(\Delta(t)) \right) \qquad (12)$$

where $R(\Delta(t)) = \xi(\Delta(t)) = \Delta_0$
For the above learning algorithms viz. ČBP and ČRPROP , let $\frac{\partial E(t)}{\partial R(w)} + i \times \frac{\partial E(t)}{\partial \xi(w)} = A$. The value of A for various learning parameters is as follows:
For $w = w_{mn}$,

$$A = Y_m{}^{\dagger}\left(R(e_n)f_\phi{}'\left(R(V_n)\right) + i\xi(e_n)f_\phi{}'\left(\xi(V_n)\right) \right) \qquad (13)$$

For $w = b_{0n}$,

$$A = zn_0{}^{\dagger}\left(R(e_n)f_\phi{}'\left(R(V_n)\right) + i\xi(e_n)f_\phi{}'\left(\xi(V_n)\right) \right) \qquad (14)$$

For $w = w^s{}_{lm}$, $A = z_l{}^{\dagger}\lambda_m{}^{\dagger}\left(1 + V_{m2}{}^{\dagger}\right)K_m \qquad (15)$

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

$$A = 2 \times exp\left(-\|Z - W^{rb}{}_m\|^2\right)\left(z_l - w^{rb}{}_{lm}\right) \times$$
$$[R(K_m)\{R(\Psi_m)\left(1 + R(V_{m1})\right) - \xi(\Psi_m)\xi(V_{m1})\} +$$
$$\xi(K_m)\{\xi(\Psi_m)\left(1 + R(V_{m1})\right) + R(\Psi_m)\xi(V_{m1})\}] \qquad (16)$$

For $w = \lambda_m$, $A = (W^s{}_m Z^T)^{\dagger}\left(1 + V_{m2}{}^{\dagger}\right)K_m \qquad (17)$

For w $= \Psi_m$,
$$A = exp\left(-\|Z - W^{rb}{}_m\|^2\right)\left(1 + V_{m1}{}^{\dagger}\right)K_m \qquad (18)$$

For $w = w_{0m}$, $\quad A = zm_0{}^{\dagger}K_m$ $\qquad$ (19)

where $\qquad\qquad\qquad\qquad$ $R(K_m) =$
$$f_\phi{}'\big(R(V_m)\big) \sum_{n=1}^N \begin{Bmatrix} R(e_n)f_\phi{}'\big(R(V_n)\big)R(w_{mn}) \\ +\xi(e_n)f_\phi{}'\big(\xi(V_n)\big)\xi(w_{mn}) \end{Bmatrix} \quad (20)$$

and $\quad \xi(K_m) =$
$$f_\phi{}'\big(\xi(V_m)\big) \sum_{n=1}^N \begin{Bmatrix} \xi(e_n)f_\phi{}'\big(\xi(V_n)\big)R(w_{mn}) \\ -R(e_n)f_\phi{}'\big(R(V_n)\big)\xi(w_{mn}) \end{Bmatrix} \quad (21)$$

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

2.3.2. Recognition/Classification

The trained networks are then used to recognize the test patterns. Recognition of test images is carried out by using the function *Max of Max* of the outputs of each associated trained neural network. For every input test pattern, we have obtained the $C \times \Omega$ output matrix. Let F_{ki} be the maximum output of ith associated network for k^{th} test image and α be the column index corresponding to F_{ki}. Cluster C is acquired corresponding to $Max_{i=1}^{C}(F_{ki})$. 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 $[(mod(k, Q) > 0 \,\&\&\, (R = mod(k, Q))]$ condition confirms the acceptance of kth test pattern.

2.4 Hybrid learning machine (HLM) algorithm

Input: Database Images
Steps:
//feature extraction
1. Represent the database images in lower dimensional feature space using eigenfaces with fisherfaces [19]. Store it in a matrix D (size $P \times (Q-1)$) after normalization.
//unsupervised genetic fuzzy distribution
2. Calculate the mean vector for all Q classes using q training images of each class as:
$\overline{x_J} = \frac{1}{q}(\sum_{k=1}^{q} x_{kj})$, where $1 \le j \le Q$ and the array of average image of given input classes is given as $X = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_Q}\}$.
3. Initialize $g=1$; repeat for g generations until $\|U^{(g+1)} - U^{(g)}\| \le \rho$ (for $g>1$)
$\quad$ A. repeat for *run=1 to r*

(a) For $g=1$, initialize fuzzy partition matrix U of size $C \times Q$. The initial population $U = [u_{ij}]$ 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 $H = \sum_{i=1}^{C} \sum_{j=1}^{Q} u_{ij}{}^\tau \delta_{ij}$ 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 $center_i = \sum_{j=1}^{Q} u_{ij}{}^\tau \overline{x_J} / \sum_{j=1}^{Q} u_{ij}{}^\tau$ where
$\delta_{ij} = \left(\sum_{p=1}^{Q-1} \| x_{jp} - center_{ip} \|^\gamma \right)^{\gamma^{-1}}$ is the distance of jth class from i^{th} cluster, $\gamma \in [1, \infty)$ is generalization parameter which is able to generalize the shape of the clusters and $\tau \in (1, \infty)$ is fuzzifier which determines the degree of fuzziness.
(c) Terminate the loop, if $\|H(u, center)^{t+1} - H(u, center)^t\| < \vartheta$ for $t > 1$, where t reflects iteration and ϑ is pre-defined threshold.
(d) Save U for the current run.
(e) *run=run+1*
$\quad$ B. Find the fitness value for r obtained partitions $\{U_{(0)}, U_{(1)}, \dots, U_{(r)}\}$ by using following fitness function [20]:
$$\pounds = \frac{Q \times d_{min}}{\sum_{i=1}^{C} \sum_{j=1}^{Q} u_{ij}{}^\tau d}$$

where $d = \sum_{i=1}^{C} \sum_{j=1}^{Q} \|\overline{x_J} - center_i\|^2$
$\quad$ C. Select the population with highest fitness value from step 3(B).
$\quad$ D. Now the above obtained partition will act as initial population for the next generation, $g=g+1$.
4. The fuzzy distribution obtained for the last generation is referred to as optimal fuzzy distribution $U^{(opt)}$.
5. Now defuzzify the $U^{(opt)}$, to obtain the cluster allocation matrix (ζ) 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 ζ by selecting top membership rank elements equal to Ω.
$\quad \zeta \leftarrow U^{(opt)}(1:C, 1:\Omega)$ where $\Omega < Q$
//supervised classification

6. Train the network for $X = X + i \times .001$ inputs for which input-output mapping is known in the form of ζ .

7. Test set is fetched from $D = D + i \times .001$, which is classified by the trained network obtained in the previous step.

Output: Recognition Results

3. Evaluation of new learning machine

In order to assess the effectiveness of the proposed hybrid learning machine, the performance evaluation has been carried out on two benchmark datasets viz. FERET[21] and AR[22] face datasets. The cross-validation is performed on the 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 hybrid machine is compared with other methods. Further, extensive comparative analysis is carried out for various considered classifiers. The classifiers considered in this work include following combinations of hidden neuron and learning algorithm; (1)ŘMLP with ŘBP, (2) ŘTROIKA (proposed neuron structure in real domain) with ŘBP, (3)ŘMLP with ŘRPROP, (4)ŘTROIKA with ŘRPROP, (5)ČMLP with ČBP, (6)*ČTROIKA* with ČBP, (7)ČMLP with ČRPROP and (8) the proposed *ČTROIKA* with ČRPROP. It is worth mentioning here that the output neurons in all the considered classifier variations are conventional neurons. Let the time taken by one learning cycle is unity. The computational supremacy and generalization skill of various classifiers are 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.

3.1 Performance with FERET face dataset

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. Fig. 1 shows few example images of this dataset. The proposed learning machine copes well with the pose variation problem and gives accuracy of 99.98% in presence of pose variations.

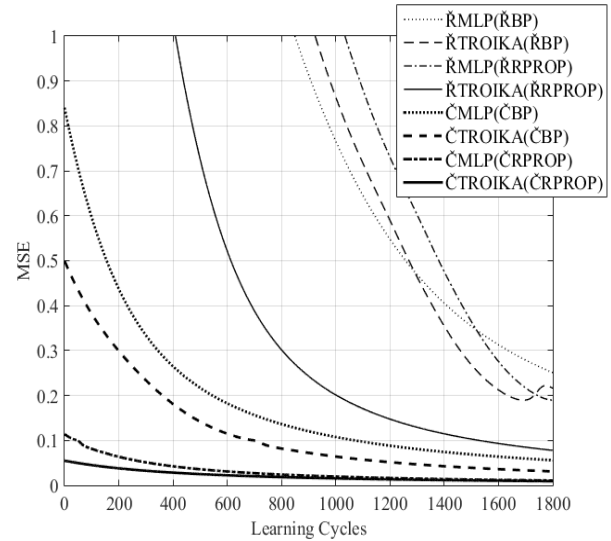Figure. 1 Example images from FERET face dataset

Figure. 2 Learning graphs in FERET face dataset with different hidden neuron and learning algorithm
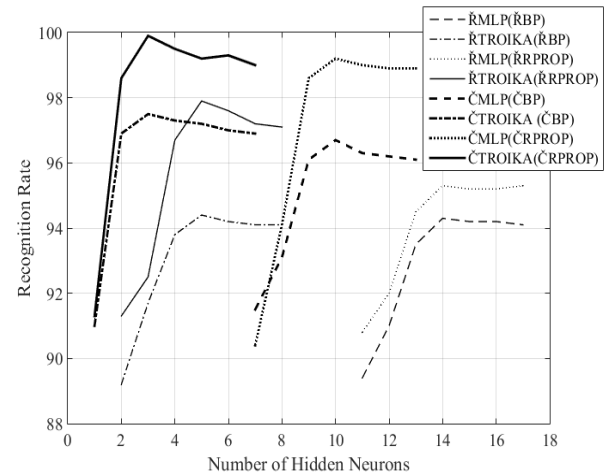

Figure. 3 Recognition rate vs number of hidden neurons for different hidden neuron and learning algorithm in FERET Face Dataset

Fig. 2 presents the learning graphs for this dataset with different combination of hidden neuron and learning algorithm. Among different combinations, the superiority of *ČTROIKA* with ČRPROP is observed in this experiment as this classifier optimize the error faster than all other considered networks. The best plausible test results for FERET dataset with different classifiers are revealed in Table 1. Different performance measures evidently exhibit that the projected hybrid learning machine with novel classifier offers enhanced performance in all respects than with other classifiers.

Classifiers based on *ČTROIKA* and ŘTROIKA neurons gives improved prediction accuracy with compact network topology in much less number of training cycles as compared to ČMLP and ŘMLP

Table 1. 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	94.3%	95.3%	94.4%	97.9%	96.7%	99.2%	97.5%	99.98%
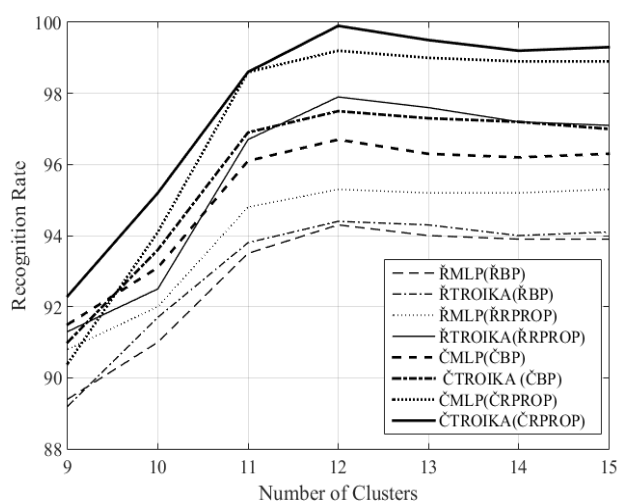
Figure. 4 Recognition rate vs number of clusters for different hidden neuron and learning algorithm in FERET Face Dataset
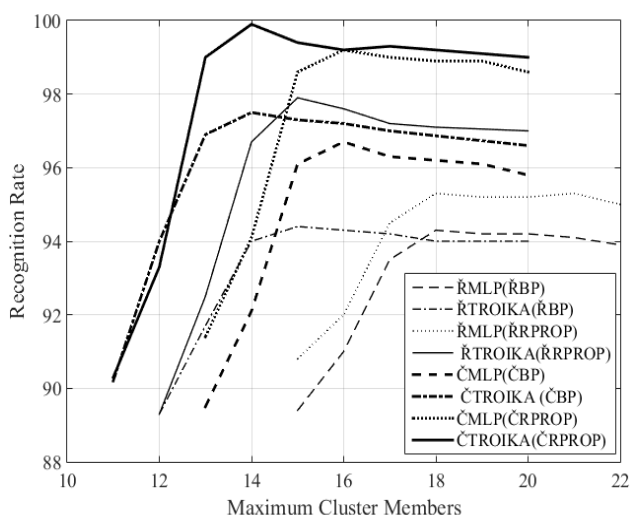

Figure. 5 Recognition rate vs maximum cluster members for different hidden neuron and learning algorithm in FERET face dataset

respectively. The recognition accuracy with respect to number of hidden neurons for different classifiers is given in Fig.3. The proposed classifier achieves best accuracy among all with sufficiently fewer number of hidden neurons than others which demonstrate the reduced complexity of ČTROIKA based network with ČRPROP training algorithm. It can also be observed 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. 4 presents the impact of number of clusters on the recognition accuracy. In this dataset, with the same number of clusters, ČTROIKA with ČRPROP gives the best performance among other classifiers. Fig. 5 shows the accuracy precision with regard to maximum cluster members for dissimilar neuron and learning algorithm. In Figs. 3, 4 and 5 the behaviour of the curves disclose that up to some extent the number of hidden neurons, clusters and maximum number of classes allowed in a cluster, respectively enhances the performance, but degrades subsequently. It is observed from results that ČTROIKA and ŘTROIKA based classifier always performs better than their MLP counterparts. The authors in [23] addressed the facial expression variations and contiguous occlusion whereas the pose and illumination variations were not considered. The proposed work copes well with pose and illumination variations. The novel learning machine realized the best recognition accuracy of 99.98% among accuracies obtained by the methods in [23] and [24] as shown in Table 2. The comparison of proposed method with existing methods again demonstrates the superior position of our work among other methods.

3.2 Performance with AR face dataset

This dataset consists of face images of 126 individuals with more than 4000 color images. A subset of 1400 images of 100 individuals with different facial expressions and illumination is

Table 2. Comparison of proposed HLM with other methodologies for FERET face dataset

Method	Accuracy(%)
Linear Regression classification model [23]	87.11(EP1), 84.50(EP2)
Dual Cross Patterns face image descriptor [24]	92.80
HLM (Proposed)	99.98

Figure. 6 Example images from AR face dataset

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. Figure 6 shows the example images having variations in expressions and illumination.

Fig. 7 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 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. 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 applications. It is obtained from the respective curves of different classifiers in Fig. 8 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 which results in compact topology. The performance of various combinations of hidden neuron and learning algorithm are presented in Fig.9 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. One can examine the effect of variation in maximum cluster members on the accuracy precision of the proposed algorithm from Fig.10. 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. It can be observed from Figs. 8, 9 and 10 that up to some extent recognition rate increases but no enhancement is noticed later with further increase in number of hidden neurons, clusters and maximum cluster members respectively. The methods suggested in [25] and [26] offers 95.0% and 95.04% recognition rates respectively. In [25] the authors considered only frontal face images while the real scenario variations are completely unaddressed. Our work considered pose, facial expression and illumination variations. As presented in Table 4 the proposed work outperforms other methods with 98.6% recognition rate.

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%

Table 4. Comparison of proposed HLM with other methodologies for AR face dataset

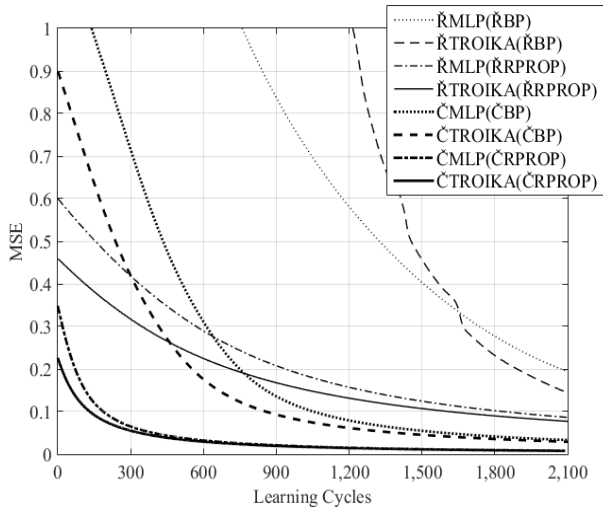| Method	Accuracy(%)
Modified Robust Sparse Coding [25]	95.0
Multi-Channel Multi-Model [26]	95.04
HLM (Proposed)	98.6

Figure. 7 Learning graphs in AR face dataset with different hidden neuron and learning algorithm
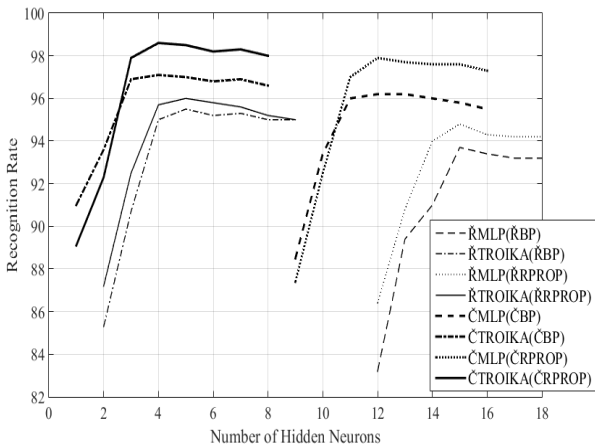

Figure. 8 Recognition rate vs number of hidden neurons for different hidden neuron and learning algorithm in AR face dataset
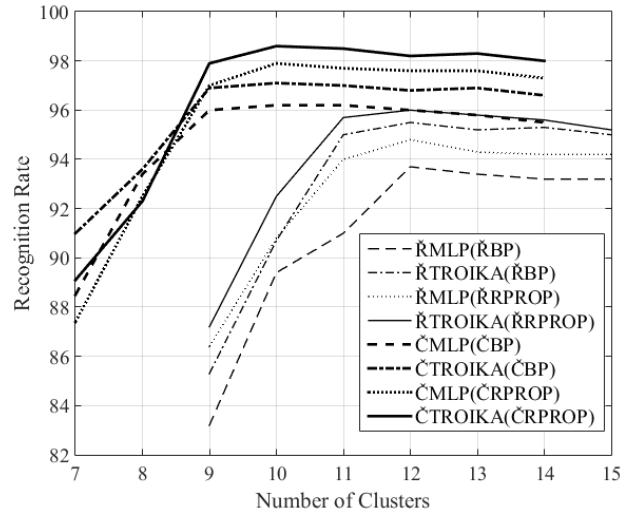

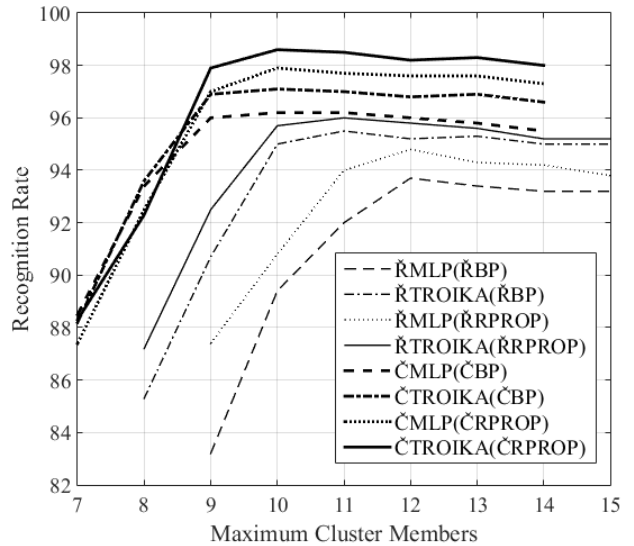Figure. 9 Recognition rate vs number of clusters for different hidden neuron and learning algorithm in AR face dataset

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

4. Conclusion

This paper presents a hybrid learning machine with the novel synergism of statistical techniques, genetic fuzzy distribution and complex neural classifier. The structure of the proposed classifier ensemble novel neurons *ČTROIKA* and complex conventional neurons with ČRPROP learning algorithm. The accuracy results of the proposed hybrid machine have been compared with other methods which exemplified the superiority of the new machine. Moreover, the performance of the proposed machine has been accounted for the various classifiers based on the type of hidden neuron and learning procedure. 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 compact topology, speedy convergence, reduced learning cycles and better recognition accuracy. 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. It is significantly examined 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. From extensive research it can be established that our learning machine is vigorous enough to recognize the human identity even with images possesses wide variations in poses, facial expressions and illumination conditions.

Finally, we wind up the paper with the conclusion that the proposed learning machine is an efficient recognition system which outperforms over other methods. Further, the proposed classifier outperforms the considered classifiers in all respects. From comparative analysis it can be concluded that ČTROIKA and ŘTROIKA always perform better 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 holds for real domain simultaneously.

References

[1] V. Vijayakumari, "Face recognition techniques: A survey", *World journal of computer application and technology*, Vol.1, No. 2, pp. 41-50, 2013.

[2] B. Shi, X. Bai, and C. Yao, "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*, Vol. 39, No. 11, pp. 2298-2304, 2017.

[3] C. K. Li, "A sigma-pi-sigma neural network (SPSNN)", *Neural Processing Letters*, Vol. 17, No. 1, pp. 1-19, 2003.

[4] M. G. Epitropakis, V. P. Plagianakos, and M. N. Vrahatis, "Hardware-friendly higher-order neural network training using distributed evolutionary algorithms", *Applied Soft Computing*, Vol. 10, No. 2, pp.398-408, 2010.

[5] N. Homma and M. M. Gupta, "A General Second-order Neural Unit", *Bulletin of College of Medical Sciences, Tohoku University*, Vol. 11, No. 1, pp.1-6, 2002.

[6] D. K. Chaturvedi, M. Mohan, R. K. Singh, and P. K. Kalra, "Improved generalized neuron model for short-term load forecasting", *Soft Computing*, Vol. 8, No. 5, pp. 370-379, 2004.

[7] B. K. Tripathi, B. Chandra, and P.K. Kalra, "The generalized product neuron model in complex domain", In: *Proc. of International Conf. on Neural Information Processing*, pp.867-876, 2008.

[8] M. Zhang, S. Xu, and J. Fulcher, " Neuron-adaptive higher order neural-network models for automated financial data modeling", *IEEE Transactions on Neural Networks*, Vol. 13, No. 1, pp. 188-204, 2002.

[9] B. K. Tripathi and P. K. Kalra, "On efficient learning machine with root-power mean neuron in complex domain", *IEEE Transactions on Neural Networks,* Vol. 22, No. 5, pp.727-738, 2011.

[10] M. D. Zeiler, "ADADELTA: an adaptive learning rate method", *arXiv preprint arXiv:1212.5701*, 2012.

[11] X. Chen, Z. Tang and S. Li, "An Modified Error Function for the Complex-value Backpropagation", *Neural Information Processing-Letters and Reviews*, Vol. 8, No. 1, 2005.

[12] X. Luo, A. D. Patton, and C. Singh, "Real power transfer capability calculations using multi-layer feed-forward neural networks", *IEEE Transactions on Power Systems,* Vol. 15, No. 2, pp. 903-908, 2000.

[13] I. Aizenberg, D. V. Paliy, J. M. Zurada, and J. T. Astola, "Blur identification by multilayer neural network based on multivalued neurons", *IEEE Transactions on Neural Networks,* Vol. 19, No. 5, pp. 883-898, 2008.

[14] A. Hirose, "Complex-valued neural networks", *Springer Science & Business Media*, Vol. 400, 2012.

[15] C. Igel and M. Hüsken "Empirical evaluation of the improved Rprop learning algorithms", *Neurocomputing,* Vol. 50, pp.105-123, 2003.

[16] S. Srivastava and B. K. Tripathi, "On the Deep Hybrid Computational Model for Face Recognition", *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 1, pp. 121-130, 2018.

[17] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition", *IEEE Transactions on Image processing,* Vol. 11, No. 4, pp. 467-476, 2002.

[18] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks", *IEEE Transactions on Neural Networks,* Vol. 13, No. 3, pp.697-710, 2002.

[19] X. He, S. Yan, Y. Hu, P. Niyogi, and H.J. Zhang, "Face recognition using laplacianfaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp.328-340, 2005.

[20] W. Sheng, G. Howells, M. Fairhurst, and F. Deravi, "Template-free biometric-key

generation by means of fuzzy genetic clustering", *IEEE Transactions on Information Forensics and Security*, Vol.3, No. 2, pp.183-191, 2008.

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

[22] A. Martinez and R. Benavente, "The AR face database, CVC Technical Report 24", http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.

[23] I. Naseem, R. Togneri, and M. Bennamoun, "Linear Regression for Face Recognition", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 32, No. 11, pp.2106-2112, 2010.

[24] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 3, pp.518-531, 2016.

[25] X. Fontaine, R. Achanta, and S. Süsstrunk, "Face Recognition in Real-world Images", In: *Proc. of International Conf. on Acoustics, Speech and Signal Processing,* pp. 1482-1486, 2017.

[26] M. S. Aslan, Z. Hailat, T. K. Alafif, and X. W. Chen, "Multi-channel multi-model feature learning for face recognition", *Pattern Recognition Letter,* Vol. 85, pp.79-83, 2017.