International Journal of Machine Intelligence
ISSN: 0975-2927 & E-ISSN: 0975-9166, Volume 7, Issue 2, 2016, pp.-474-479.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000231

Research Article

AMERICAN SIGN LANGUAGE ALPHA-NUMERIC CHARACTER CLASSIFICATION USING NEURAL NETWORK CLASSIFIERS

MAPARI R.B.[1]* AND KHARAT G.U.[2]

[1]Anuradha Engineering College, Chikhli, 443201, MS, India
[2]Sharadchandra Pawar College of Engineering, Otur, 412409, MS, India
*Corresponding Author: Email-rajesh_mapari2001@yahoo.com

Received: May 13, 2016; Revised: May 31, 2016; Accepted: June 01, 2016; Published: June 14, 2016

Abstract- The American Sign Language (ASL) alpha-numeric character classification/recognition without using any aid (embedded sensor, color glove) is really difficult task. This paper describes a novel method to classify static sign by obtaining feature set based on DCT (Discrete Cosine Transform) and Regional properties of hand image. Feature set of size 1860×74 is later trained and tested using different classifiers like MLP, GFFNN, SVM. We have collected dataset (alpha numeric character) from 60 people including students of age 20-22 years and few elders aged between 25-38 who have performed 31 signs resulting in total dataset of 1860 signs. Out of this 90% dataset is used for training and 10% considered for Cross validation. We have got maximum classification accuracy as 86.16 % on CV dataset using GFF Neural Network.

Keywords- ASL, MLP, GFFNN, SVM

Citation: Mapari R.B. and Kharat G.U., (2016) American Sign Language Alpha-Numeric Character Classification Using Neural Network Classifiers. International Journal of Machine Intelligence, ISSN: 0975-2927 & E-ISSN: 0975-9166, Volume 7, Issue 2, pp.-474-479

Introduction

Sign language is the only way for deaf-mutes to have conversation. The figures obtained from World Federation of the Deaf (WFD) is in millions. Some Sign Language recognition Systems need to have some aids with signer like instrumented/Data Glove to perform sign. However using bared hand also signer can perform sign in some cases. So roughly recognition systems are either based on some instrumental gadgets or vision (Camera) based. However a combination of both is also tied by researchers. Due to direct contact of sensors, information is more accurate when considered instrumented glove based system. However vision based system need to first segment the hand from image which may be depend on environment like background, lighting conditions, wearing of signer, skin color of signer.

So to do this, vision based system need to first identify the object from an image based on color space selection may be based on skin color or color glove used in segmentation process. Skin color based segmentation is mainly done with plain background or with cloths of dark color where complete hand is covered and only palm, fingers are uncovered.

However new sensors like Leap Motion Sensor & Kinect have changed the traditional way of processing the information. Using this information is now available in 3D format (2D +Depth information).

Related Work

Fu-Hus Chou et al. [1] have detected & recognized images of numbers from 1 to 5. First forearm and elbow is deleted by adjusting image. In recognition, phase model is first constructed for static hand gesture and then unidentified gesture is recognized by Gaussian Model match. For five numbers recognition, 300 gesture images are used to build the Gaussian Model match model & 200 test samples for each number gesture. M.S. Sinith et al. [2] have recognized few static signs namely A, W, O, H, I and L of ASL. First color image is converted to gray scale image and then filter (Sobel filter) is used to get binary hand image. Longest three connected components of this image is considered as feature vector as a input to Support Vector Machine. In [3], Hee-Deok Yang et al. have worked on recognition of Manual and Non-manual sign of ASL using color data glove. Using hierarchical Conditional random field, Manual signs are recognized. Motion and location used as features. Boost Map methods are used to recognize shape of hand. For Non manual sign recognition, multiclass SVM is used which uses 31 feature point and distance and angle as a measurement facial expressions classification. In [4], Fahad Ullah et al. have worked on recognition of 26 alphabets of ASL using Cartesian Genetic Programming where color image is converted to binary images of resolution 47*27 pixel which later on converted to linear array of size 1*1269 pixels vector. CGP has 1269 inputs and 5 bit output representing exact number of recognized sign. Using queuing technique word is formed by collecting alphabets.

In 2015, Asha Thalange et al. [5] proposed a method to detect static images of numbers 0-9 in ASL. Number of open fingers and distance between them is considered as features. On similar platform, Priyanka Mekala et al. [6] have recognized all the alphabets of ASL using combinational neural networks architecture. The feature vector contains 55 features which includes finger tip elements, motion vector elements, MV sequence and wavelet transform of the Fourier transformed image of a gesture. In [7], Taehwan Kim et al. have worked on finger spelling sequences which form words in American Sign Language (ASL) from a video where outputs of multilayer perceptron classifiers are used as observations in a hidden Markov model based recognizer. SIFT is used for feature extraction followed by PCA. Dominique Uebersax et al. [8] have worked on.

Recognition system for recognizing letters and finger-spelled words in real-time of ASL. System Test data was composed of 7 subjects using TOF camera and two depth sensors. The depth data is used for hand localization and segmentation. Combinations of three methods are used for letter recognition namely average neighborhood margin maximization, depth difference and hand rotation.

Using Cyber Glove, Jerome M. Allen et al. [9] have designed a system to recognize 24 static finger spelling letter of ASL and translate it to corresponding alphabet in printed and spoken English letters. Pattern recognition technique with perceptron network was used. Vasiliki E. Kosmidou et al. [10] proposed a analysis of the surface electromyogram signal for ASL gesture recognition. Sixteen features are extracted from the user's forearm and evaluated by the Mahalanobis distance criterion. Feature dimension reduction is achieved using discriminant analysis for classification of sign. Using Kinect depth sensor C. S. Weerasekera et al. [11] have proposed a vigorous method for recognition of bare-handed static ASL. Local Binary Patterns histogram based on color and depth information, and also geometric features of the hand are used as features. Linear binary SVM classifiers are used for recognition. Similarly using same senor L. Nanni et al. [12] have proposed a system based on distance and curvature as a Features to recognize ASL . A combination of SVM classifiers and rotation boosting is used for recognition. Lucas Rioux-Maldague et al. [13] present a technique for extraction features of ASL using depth and intensity images. Classification of Finger spelling carried using a Deep Belief Network.

Using Leap Motion Sensor, A.S.Elons et al. [14] haves captured hands and fingers movements in 3D digital format. These temporal and spatial features are fed into a MLP Neural Network. Similarly Cao Dong et al. [15] recognized 24 static ASL alphabets with accuracy of 92% using localize hand joint positions under kinematic constraints. 13 key angles of the hand skeleton were used as the features for Random Forest (RF) classifier to describe hand gestures.

Giulio Marin et al. [16] proposed a novel ASL static hand gesture recognition scheme using Leap motion and Kinect. Feature set of leap Motion consists of Fingertips distances, Fingertips angles and Fingertips elevations. Feature set of Kinect consists of Curvature, Correlation. A Multi-class SVM classifier is used to recognize the performed gestures.

Experimental Setup

We have kept Black background using black cloth and Signers have wear black T-shirt while performing sign. This has helped to segment the hand easily from uniform and fixed background. For acquiring image, we have used SONY camera 16.1 mega pixel plus 5x Optical zoom. In first phase we have read original image as shown in [Fig-1 (a)] and cropped it by maintaining height width ratio of hand portion using bounding box technique with L*a*b color space as shown in [Fig-1 (b)]. This way hand is exactly at the center of image as shown in [Fig-1 (c)]. Hand image is then converted to 256×256 size RGB image.

Later on image is converted to gray scale image. The gray scale image is divided in to 32×32 block using block-processing operation. 2-D DCT of each 32-by-32 block is calculated which results in 64 values.

Filtering operation is carried out by testing various filters but the best result is obtained using Gaussian Filter. Followed by smoothing operation image is converted to black and white image using gray threshold as shown in [Fig-1(d)]. However to get proper black and white image to extract regional properties, it must be smooth. So series of morphological operations as shown in [Fig-1(e-i)] are performed to get best result. It can be observed from [Fig-1(e)] & [Fig-1(i)] that jagged edges have been removed.

From the [Fig-1(i)], Regional properties like Area, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Convex Area, Equiv Diameter, Solidity, Extent & Perimeter are calculated. So feature set consists of 64 DCT values and 10 values of regional properties resulting in feature set of total 74 values.

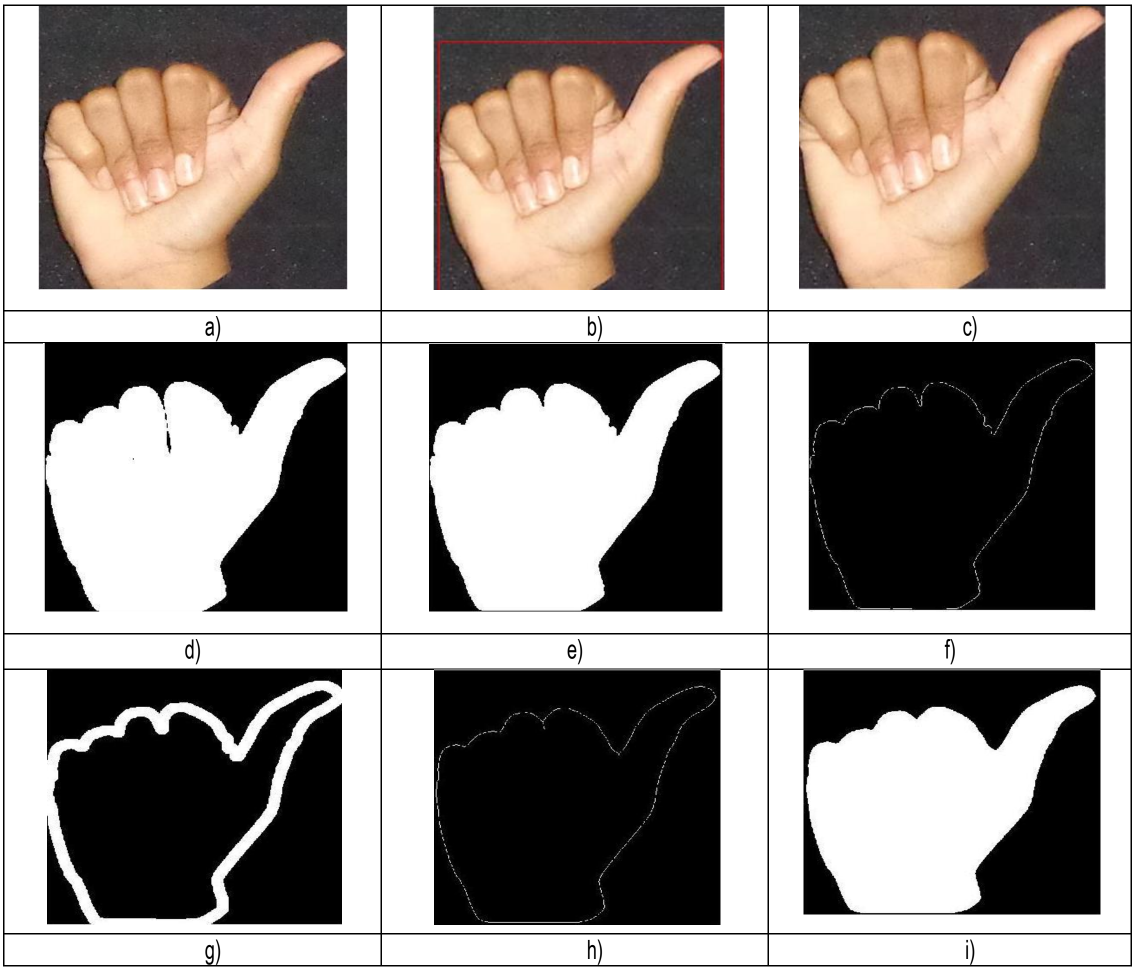

Fig-1- a) original RGB image b) bounding box c) hand at the center of image d) black and white image e) morphological closing & filling operation f) morphological remove operation g) dialation operation h) thining operation i) filling of holes

Generalized Feed Forward Neural Network

Generalized feed forward networks are a generalization of the MLP such that connections can jump over one or more layers. Following trials have been performed to get optimal parameters for minimum MSE and maximum percentage Average Classification Accuracy. Feature vectors are divided into two part as 90 % for training (TR) and 10% for Cross validation (CV). By keeping only one hidden layer, first network is tested to search number of Processing Element (PE) required in Hidden Layer, which gives minimum Mean Square Error (MSE) on training dataset. [Fig-2] shows that minimum MSE is given by processing element (PE) number 18.

Different transfer function like Tanh, Linear Tanh, Sigmoid, Linear Sigmoid, Softmax and Learning rules like Step, Momentum, Conjugate Gradient, Quick Propagation, Delta Bar Delta are varied in hidden Layer to get maximum percentage classification accuracy as shown in [Fig-3].

GFFNN with the following parameter setting gives maximum Percentage classification accuracy of 97.08 % on training and 86.15 % on CV dataset.

Tagging of Data: 90% for Training & 10% Cross validation

Input Layer: Input Processing Element - 74 Exemplars - 1674

Hidden Layer: Processing Elements - 18 Transfer Function - Tanh Learning Rule - Conjugate Gradient

Output Layer: Output PE's - 31 Transfer Function - Tanh Learning Rule - Conjugate Gradient.

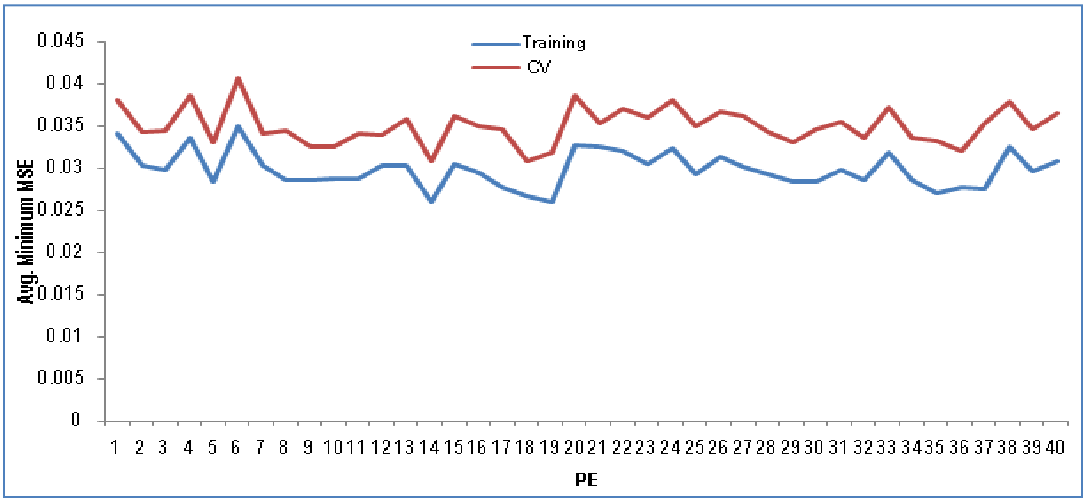

Fig-2 Processing Element (PE) Vs Minimum MSE

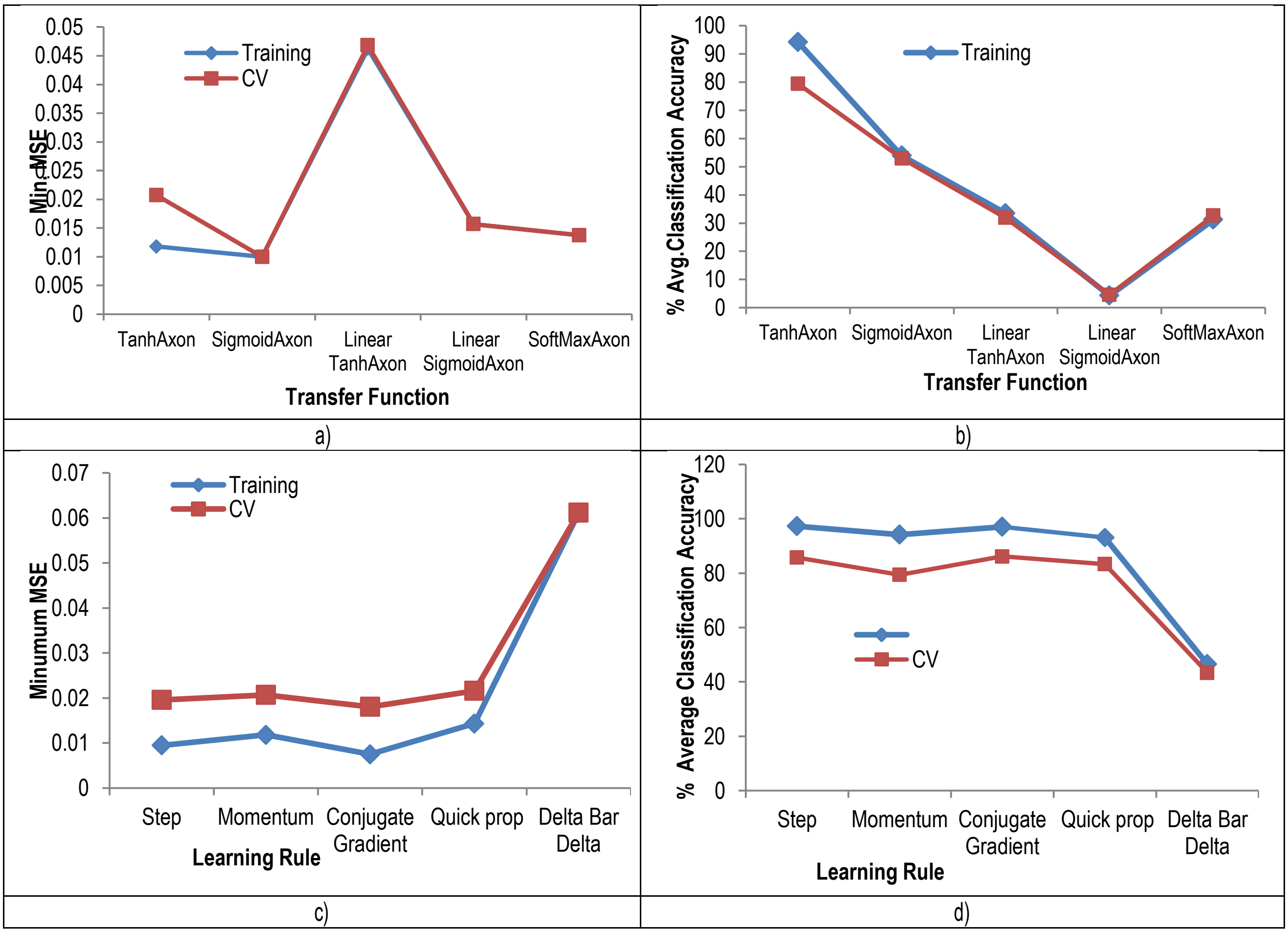

Fig-3- a) Minimum MSE Vs Transfer Function b) Percentage of Average classification Accuracy Vs Transfer Function c) Minimum MSE Vs Learning Rule d) Percentage of Average classification Accuracy Vs Learning Rule

Table-1 *Confusion Matrix for Cross Validation (CV) data set using GFF Neural Network*

Desired \ O/P	1	3	4	5	7	8	9	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	**5**	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	0	**4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	**3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
5	0	0	0	**11**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	**3**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	1	**4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	**6**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	1	0	0	0	0	0	0	0	**9**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	1	1	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**5**	1	0	0	0	0	2	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	1	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0
U	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	1	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**2**	1	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**57**

Table-2 *Performance Matrix for Cross Validation (CV) data set using GFF Neural Network*

Sign	1	3	4	5	7	8	9	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
% Correct classification	83	67	75	100	75	80	100	83	100	100	71	100	100	100	100	83	75	100	100	83	100	80	80	83	63	100	100	50	50	89	100

Multilayer Perceptron Neural Network

Multilayer perceptrons (MLPs) are layered feed forward networks typically trained with static back propagation. These networks have found their way into applications requiring static pattern classification.

Like GFF Neural Network, we have performed similar trials using MLP Neural Network. It is observed from graph as shown in [Fig-4] that Minimum MSE (Mean Square Error) is for CV data is at 29 numbered PE.

Here also different transfer function like Tanh, Linear Tanh, Sigmoid, Linear Sigmoid, Softmax and Learning rules like Step, Momentum, Conjugate Gradient, Quick Propagation, Delta Bar Delta are varied in hidden Layer to get maximum percentage classification accuracy as shown in [Fig-5].

MLP with the following parameter setting gives maximum Percentage classification accuracy of 97.25 % on training and 85.58 % on CV dataset.

Tagging of Data: 90% for Training & 10% Cross validation

Input Layer: Input-Processing Element - 74 Exemplars - 1674

Hidden Layer: Processing Elements - 29 Transfer Function - Tanh

Learning Rule - Conjugate Gradient

Output Layer: Output PE's :31 Transfer Function - Tanh

Learning Rule:- Conjugate Gradient.

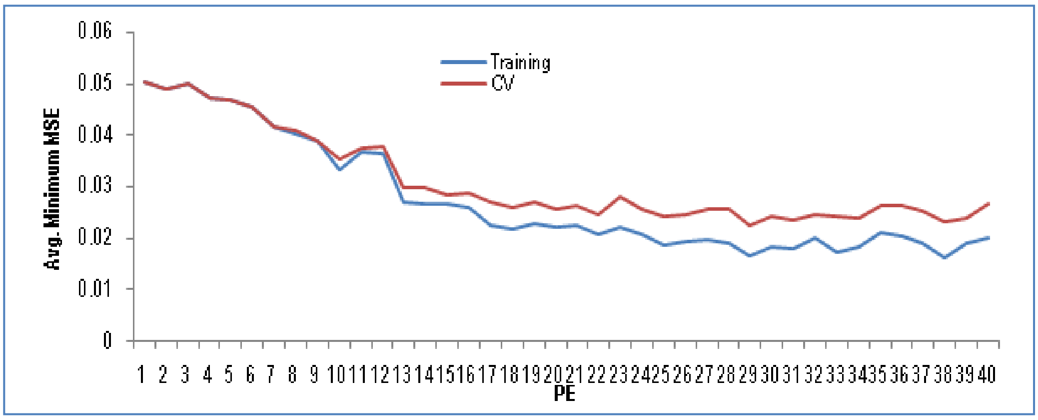

Fig-4 Processing Element (PE) Vs Minimum MSE

Support Vector Machine

We have varied epoch & number of runs by fixing the step size at 0.1. It is observed that from epoch 25 onwards, there is very little change is MSE for CV data as shown in [Fig-6]. It can be observed from [Fig-7] that maximum Percentage classification accuracy is obtained at step size 0.1.

After experimentation we have observed that the best result is i.e. : 99.11% on training and 84.62% on CV data set with optimal parameter setting as below

Tagging of Data: 90% for Training & 10% Cross validation

No. of Epoch: 25 No. of Runs: 1

Input Processing Elements: 74

Output Processing Elements 31

Exemplars: 1674 Step Size: 0.1 Kernel Algorithm: Adatron

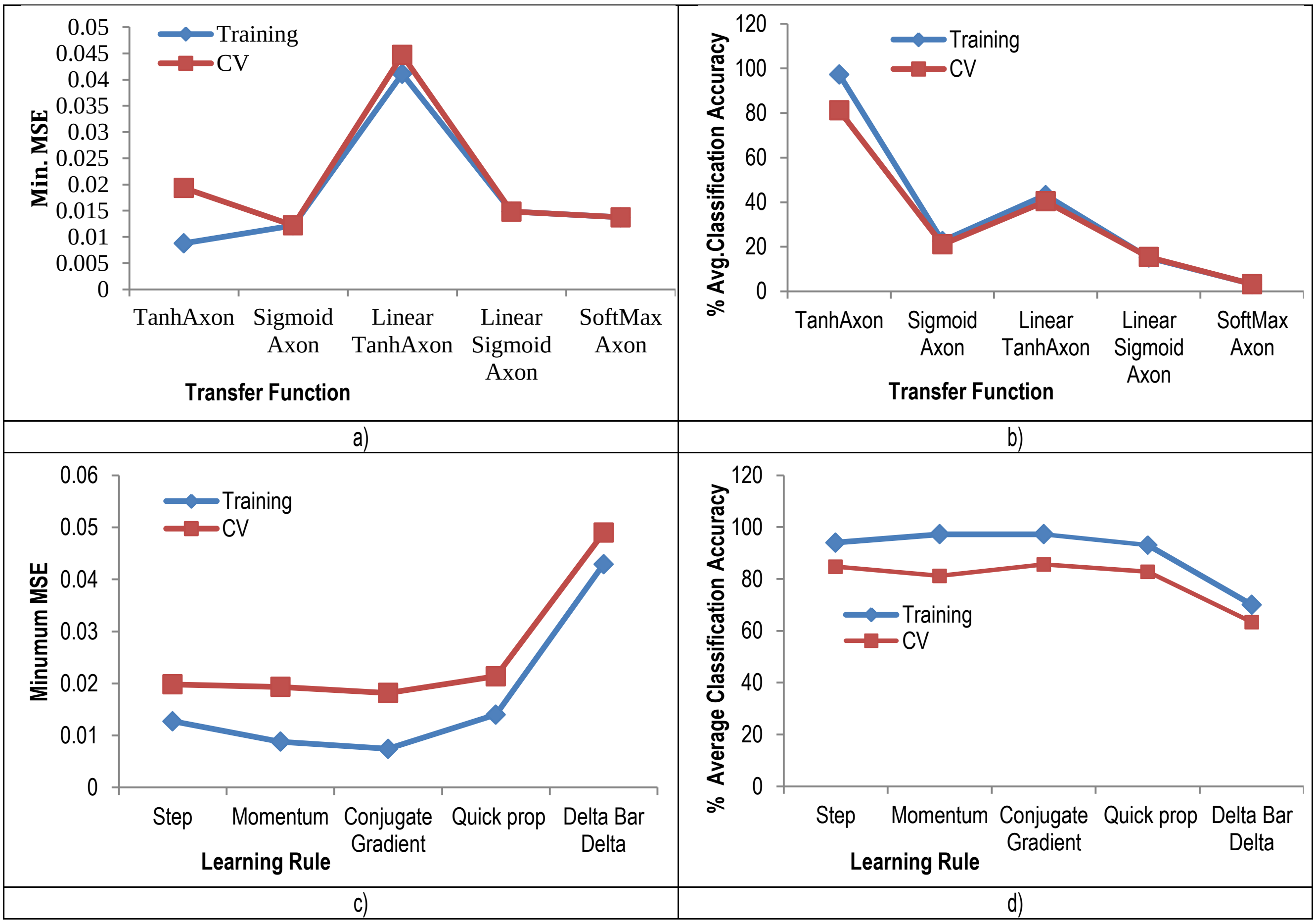

Fig-5- a) Minimum MSE Vs Transfer Function b) Percentage of Average classification Accuracy Vs Transfer Function c) Minimum MSE Vs Learning Rule d) Percentage of Average classification Accuracy Vs Learning Rule

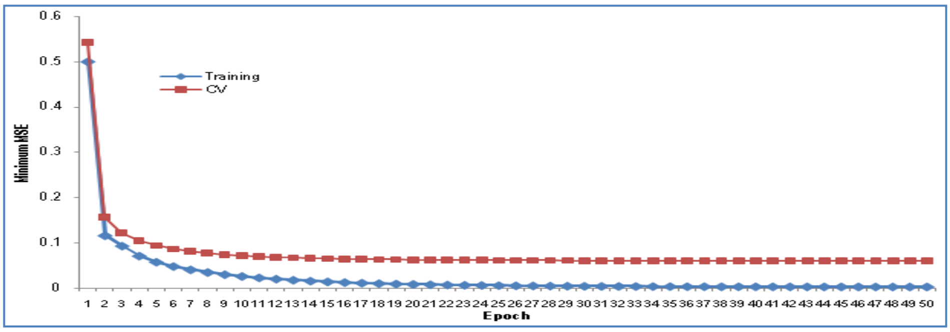

Fig-6 Epochs Vs minimum MSE

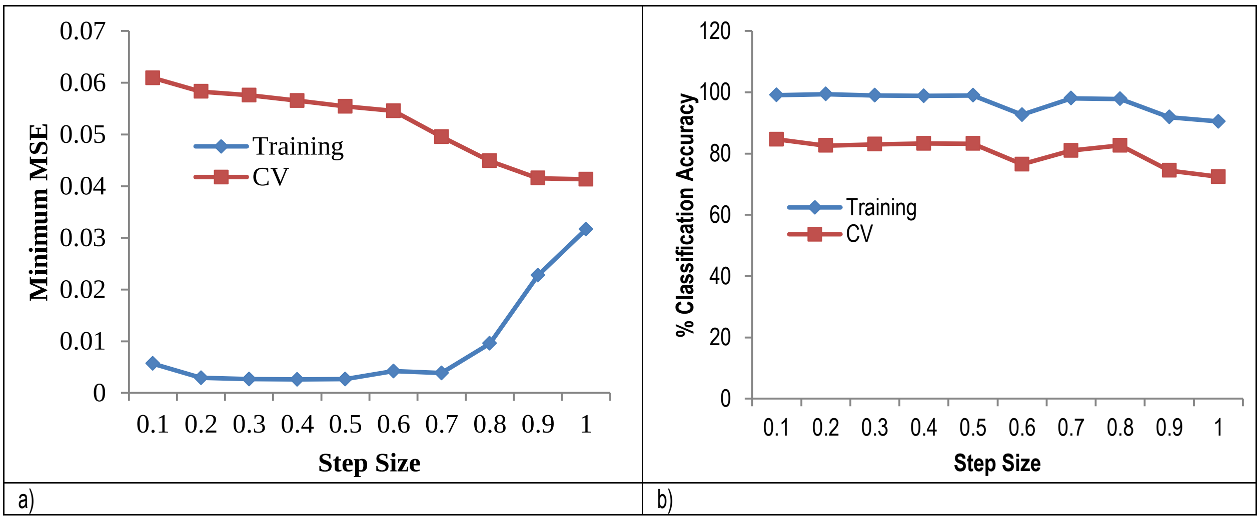

Fig-7 a) Minimum MSE Vs step size b) Percentage of classification accuracy Vs step size

Result

As the maximum Classification accuracy is obtained using GFF Neural Network we have shown details of confusion matrix and performance matrix of CV data only. It can be observed from confusion matrix shown in [Table-1] that percentage of correct classification of signs like 3, D,V,W are not much satisfactory because of samples are misclassified. So Average classification accuracy of these signs is poor as shown in [Table-2]. We have obtained maximum Average classification accuracy as 86.16 % on Cross Validation data with the optimal parameter setting as explained earlier using GFF Neural network as shown in [Table-3]. We have not considered dynamic signs like J, Z. However static signs 2 & 6 are also not considered because of exactly similar posture like V and W respectively.

Table-3 *Performance measure of different Neural Network classifiers*

Sr. No.	Neural Network Classifier	Percentage of Average Classification Accuracy		Elapsed Time (Sec.)
		Training	CV	
1	MLP	97.25	85.58	270
2	GFF	97.08	86.16	324
3	SVM	99.11	84.62	73

Conclusion

In this paper, we have presented two techniques namely DCT and Regional Properties of Sign images for the accurate classification of signs. From the [Table-3] it can be concluded that although GFF neural network is more precise in classification as compared to the other classifiers but, the computational time required for the classification is almost 4.5 times greater than SVM Neural Classifier.

Future Work

In future it is proposed to work for the recognition of signs by collecting more database of different languages which will make the system language independent.

Conflicts of Interest: None declared.

References

[1] Fu-Hua Chou, Yung-Chun Su (2012) *IEEE/ASME International Conference on Advanced Intelligent Mechatronics,* 885-889, July 11-14.

[2] Sinith M.S., Soorej G Kamal, Nisha B., Nayana S., Kiran Surendran, Jith P. S. (2012) *International Conference on Advances in Computing and Communications,* 122-125.

[3] Hee-Deok Yang, Seong-Whan Lee (2011) *Combination of Manual and Non-Manual Features for Sign Language Recognition Based on Conditional Random Field and Active Appearance Model, Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, 1726-1731.

[4] Fahad Ullah (2011) *American Sign Language Recognition System for Hearing Impaired People Using Cartesian Genetic Programming, 5th International Conference on Robotics and Application*, 96-99.

[5] Asha Thalange, Shantanu Dixit (2015) *Effect of Thinning Extent on ASL Number Recognition Using Open-finger Distance Feature Measurement Technique, International Conference on Signal Processing And Communication Engineering Systems (SPACES),* 39-43.

[6] Priyanka Mekala, Ying Gao (2011) *Real-time Sign Language Recognition based on Neural Network Architecture, in Proceedings IEEE*, 195-199.

[7] Taehwan Kim, Karen Livescu, Gregory Shakhnarovich (2012) *American Sign Language Fingerspelling Recognition with Phonological Feature based Tandem Models, Proceedings IEEE*, 119-124.

[8] Dominique Uebersax, Juergen Gall et al. (2011) *Real-time Sign Language Letter and Word Recognition from Depth Data, IEEE International Conference on Computer Vision Workshops*, 383-390.

[9] Jerome M. Allen, Pierre K. Asselin, Richard Foulds (2003) *American Sign Language Finger Spelling Recognition System" Proceedings of IEEE 29th Annual Northeast Bioengineering Conference*, 285-286.

[10] Vasiliki E. Kosmidou, Leontios J. Hadjileontiadis , Stavros M. Panas (2006) *Evaluation of surface EMG features for the recognition of American Sign Language gestures, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA,* 6197-6200.

[11] Weerasekera C. S., Jaward M. H., Kamrani N. (2013) *Robust ASL Finger spelling Recognition Using Local Binary Patterns And Geometric Features, Proceedings of IEEE,* pp. 1-8.

[12] Nanni L., Lumini A., Dominio F., Donadeo M. and Zanuttigh P. (2014) *International Journal of Automated Identification Technology.*

[13] Lucas Rioux-Maldague, Philippe Gigu`ere (2014) *Sign Language Fingerspelling Classification from Depth and Color Images using a Deep Belief Network, Canadian Conference on Computer and Robot Vision*, pp. 92-97.

[14] Elons A.S., Menna Ahmed, Hwaidaa Shedid and Tolba M.F. (2014) *Arabic Sign Language Recognition Using Leap Motion Sensor, 9th International Conference on Computer Engineering & Systems (ICCES),* 368-673.

[15] Cao Dong, Ming C. Leu, Zhaozheng Yin (2015) *American Sign Language Alphabet Recognition Using Microsoft Kinect, The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Boston, MA, , 2015).*, 44-52.

[16] Giulio Marin, Fabio Dominio, Pietro Zanuttigh (2014) Hand gesture recognition with leap motion and kinect devices, IEEE International Conference on Image Processing (ICIP), 1565 – 1569.