Performance Analysis of Different Classifiers for Indian Sign Language Recognition

[1]Rajesh B. Mapari, [2]Dr. G.U.Kharat
[1]Asst. Prof, [2]Professor
[1]Anuradha Engineering College, Chikhli, India
[2]Sharadchandra Pawar College of Engineering, Otur, India
[1]rajesh_mapari2001@yahoo.com., [2]gukharat@gmail.com

Abstract: Many approaches for Sign Language recognition have been tried by researchers including Web camera to latest device like Kinect. The Indian Sign language (ISL) consists of signs performed using one hand and two hands as well. In this paper, an Indian Sign Language recognition approach is presented using Leap Motion Sensor. The leap motion sensor captures the hand gesture and gives finger and palm position in 3D format (X, Y, Z axis values). While performing sign, the Leap Motion Camera is kept about 10 degree inclined so that depth information is properly extracted. Different distance and angle measures are used to form feature vector of 97 values. Dataset consists of signs performed by 10 signers (students of age 20-22 years) who have given training about how to perform signs. we have considered 33 signs. Every signer has performed each sign ten times which results in total dataset of 3300 signs. Out of this, 90% dataset is used for training and 10% dataset is used for testing/Cross validation i.e. first 9 student's dataset for training and last student's dataset for Cross validation. we have trained and tested different Neural Network classifiers like MLP, GFFNN, SVM. we have got maximum classification accuracy as 96.36% on CV/testing dataset using GFF Neural Network.

Keywords: ISL, MLP, GFFNN, SVM.

I. INTRODUCTION

Sign language is a means of communication for deaf and mute people community through gestures and visions. Both manual and non-manual expressions are conveyed through vision. If dean-mute has talk to any normal person using sign language then it's almost impossible for normal person to understand the meaning of what deaf-mute is talking about. However this can be done if any expert translator is present who can tell the meaning. But the translator may not be available at any time and at any place. The solution to this problem is Human-Compute Interaction system which can be installed at many places like post office, Railway station, Banks etc. So that meaning of sign performed by Deaf-Mute can be understood. There are different sign languages all over the

Many Sign Languages such as American Sign Language (ASL), British Sign Language (BSL), Australian Sign Language, Indian Sign Language (ISL), Chinese Sign Language (CSL) etc. Basically Sign Language recognition approaches can be classified as instrumented glove based and vision based. In first approach equipped sensors measures information related to the shape, orientation, movement, and location of the hand. As it is based on direct coordinate values the segmentation is easily achieved which is difficult compared to bare hand segmentation in vision based system. However due to wearing of many sensors on wrist and arms, it creates difficulty for signer to perform the sign in its natural way.

In contrast, vision based system supports to both manual and non manual signs. However the segmentation is color space based which creates difficulty if background is not uniform and has matching color objects with hand color. Other issue is occlusion handling which may create when hand and face overlap or one hand overlap on another.

Recently a new method which considers both the local feature and global feature of gesture is introduced using Kinect sensor. But the problem with this sensor is it's not support minute details like shape of hand.

In this paper we have proposed one solution to recognize static Signs of ISL using local features. We have used Euclidean distance and Cosine angle as features based on six 3D key points (five corresponding to five fingers and one for center of palm).

II. RELATED WORK

Most of the research work in sign language recognition system is concern to translation of sign language to text or spoken word. Some systems are as follows.

Karishma Dixit and Anand Singh Jalal [1] have worked on recognition of signs in ISL. Once hand is extracted from image, it is filtered using median filter & later converted to binary image. From binary image Hu invariant moment set (consists of seven values) and structural shape descriptors (Aspect ratio, Compactness, Elongation, Spreadness, and Orientation) are used as a feature vector input to multi-class Support Vector Machine (MSVM). Out of 720 images database collected for 12 different signers for 60 different words, 600 images are used for testing and 120 images for training. Recognition rate is 96 %. Adithya v [2] proposed a method which uses digital image processing techniques and artificial neural network for recognizing different signs of ISL. The signs considered for recognition include 26 letters of the English alphabet and the numerals from 0-9. Database for training and testing made in controlled environment in lab with black background. After hand segmentation, image is in binary format where distance transform is used as a feature. The experiments are conducted for 36 signs with 15 images of each. 10 images of each sign are used for training the system and 5 images of each sign are used for testing the system. Features extracted from the training image set are used for training the feed forward neural network. Average recognition rate of 91.11%

In [3], J. Rekha et al. have worked on recognition of 23 static and 3 dynamic signs of ISL using different classifier. Using YCbCr skin color model hand region is segmented. To obtain the shape and texture information Principal Curvature Based Region detector (PCBR) and 2-D Wavelet Packet Decomposition (WPD) methods are used. Set of feature vector given by the PCBR, WPD-2 & finger count for each hand signs (alphabets of ISL) are stored in a database. Finally, the generated vector is fed into the multi-class SVM training classifier model that was built in the training stage to classify and recognize the hand gestures. For training of classifier 920 static and 66 dynamic samples are used. In Testing phase 22 videos are used. It is observed that SVM classifier has good recognition 91.7%. In 2015, Ankita Saxena et al.[4] presents principal component analysis which is a fast and efficient technique for recognition of Indian sign language gestures from video stream. In proposed technique 3 frames per second captured from video stream. By comparing three frames static posture is extracted. Static posture image of size 60X80 pixels matched with dataset using PCA and result is image of sign. The recognition rate of all gestures in between 70-80%.

In year 2006, M.K.Bhuyan & P.K. Bora [5] worked on Indian Sign language recognition and achieved recognition rate of 95.2%,95.5% and 92.5% for Static signs, Dynamic signs & Sentences respectively. In year 2011 offline approach to recognize sign [6], M.K. Bhuyan et al. have recognized few static hand postures (8 no's) by analyzing texture and key geometrical features of the hand. After segmentation of hand, features are extracted from joints of fingers and texture analysis of the circular strips in the segmented hand image. Average recognition rate using fingertip detection algorithm for 8 hand postures is 93.4%.

However due to recent development of inexpensive depth cameras, e.g., the Kinect sensor & Leap Motion, new opportunities opened doors for hand gesture recognition. In 2013 [8], Zhou Ren et al. have used advanced sensors like Kinect to recognize signs from 1 to 10. Using Template matching and Finger-Earth Mover's Distance (FEMD), experiments carried out which demonstrate that hand gesture recognition system is 93.2 % accurate.

A.S.Elons et al. [9] have captured hands and fingers movements in 3D digital format using Leap motion. The sensor throws 3D digital information in each frame of movement. These temporal and spatial features are fed into a Multi-layer Perceptron Neural Network (MLP). The system was tested on 50 different dynamic signs (distinguishable without non manual features) and the recognition accuracy reached 88% for two different persons. In 2014 [10],Giulio Marin et al. [10] proposed a novel 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. It is observed that due to combination of Leap and Kinect the recognition accuracy achieved is 91.28% for 10 static signs of ASL.

III. EXPERIMENT SETUP

A. Data Collection

The Leap Motion controller is a small USB peripheral device which is designed to be placed on a physical desktop, facing upward. Using two monochromatic IR cameras and three infrared LEDs, the device observes a roughly hemispherical area, to a distance of about 1 meter. Leap Motion sensor is a small size sensor which is easy to use and of low cost.

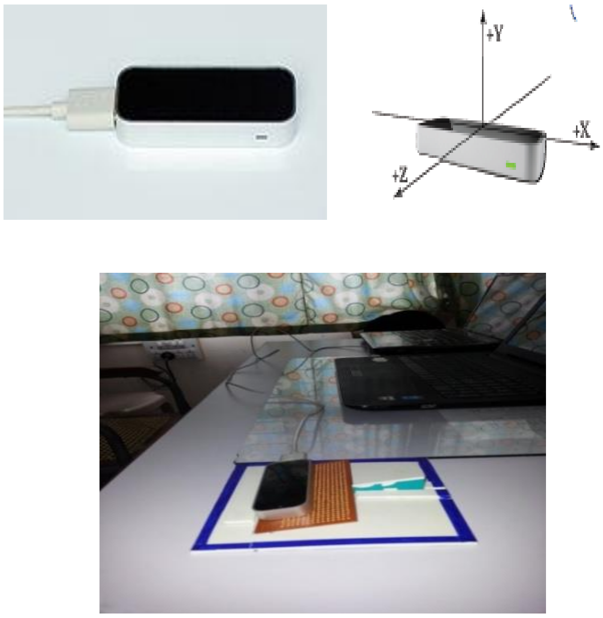

Fig. 1. Leap Motion Controller and its mounting stand

This sensor not only tracks the hand movements but also it has the ability to distinguish the fingers' joints and track their movements. This information is supported by Leap Motion Vendors "https://www.leapmotion.com" [11].

This sensor not only tracks the hand movements but also it has the ability to distinguish the fingers' joints and track their movements. While using Leap Motion we have kept it 10 degrees inclined as shown in Fig 1.

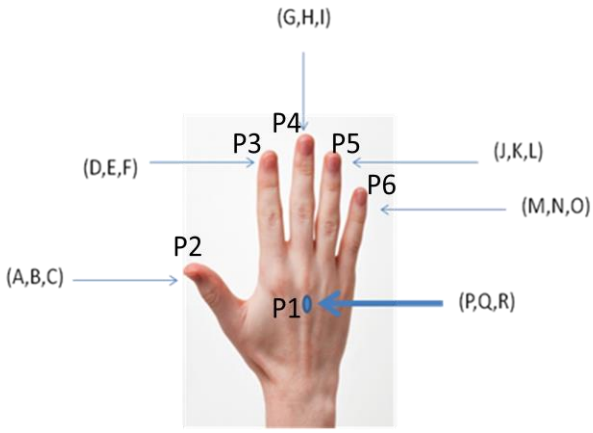

Fig. 2. Data Acquisition through Leap Motion Sensor

The 3D co-ordinates of finger tip and palm is accessed using Leap Motion API as shown in Fig. 2. Samples of signs on Visualizer tool of Leap Motion Sensor is shown in Fig. 3.

Fig. 3. Sample of ISL Signs on Visualizer Tool

B. *Feature Extraction*

The feature set consists of positional values of finger and palm, distance between positional values, angle between positional vales with respect to plane. Understanding the fact that every person has different hand shape and size, a database is created so as to have all possible samples of hand pose for concern posture.

$$D1 = \sqrt{(P-A)^2 + (Q-B)^2 + (R-C)^2} \quad (1)$$
$$D2 = \sqrt{(P-D)^2 + (Q-E)^2 + (R-F)^2} \quad (2)$$
$$D3 = \sqrt{(P-G)^2 + (Q-H)^2 + (R-I)^2} \quad (3)$$
$$D4 = \sqrt{(P-J)^2 + (Q-K)^2 + (R-L)^2} \quad (4)$$
$$D5 = \sqrt{(P-M)^2 + (Q-N)^2 + (R-O)^2} \quad (5)$$
$$D6 = \sqrt{(A-D)^2 + (B-E)^2 + (C-F)^2} \quad (6)$$
$$D7 = \sqrt{(A-G)^2 + (B-H)^2 + (C-I)^2} \quad (7)$$
$$D8 = \sqrt{(A-J)^2 + (B-K)^2 + (C-L)^2} \quad (8)$$
$$D9 = \sqrt{(A-M)^2 + (B-N)^2 + (C-O)^2} \quad (9)$$
$$D10 = \sqrt{(D-G)^2 + (E-H)^2 + (F-I)^2} \quad (10)$$
$$D11 = \sqrt{(D-J)^2 + (E-K)^2 + (F-L)^2} \quad (11)$$
$$D12 = \sqrt{(D-M)^2 + (E-N)^2 + (F-O)^2} \quad (12)$$
$$D13 = \sqrt{(G-J)^2 + (H-K)^2 + (I-L)^2} \quad (13)$$
$$D14 = \sqrt{(G-M)^2 + (H-N)^2 + (I-O)^2} \quad (14)$$
$$D15 = \sqrt{(J-M)^2 + (K-N)^2 + (L-O)^2} \quad (15)$$

Similarly angles between every two positional values is calculated as

Cos theta1=dot(P1,P2)/(norm(P1)*norm(P2)) (16)
thetha_deg1=acos(Costheta1)*180/pi (17)

Likewise for all possible combination of point p1 to p6, total 15 angles (thetha_deg1, thetha_deg2,... thetha_deg15) are calculated. Thus for one hand we get 48 values (18 positional values, 15 distance values and 15 angle values).

Similarly for another hand we get 48 values. Distance between centers of palm of two hands is also calculated which results in total feature set of 97 values. Ten students have performed 33 signs. Each sign is repeated 10 times. So we have obtained the feature matrix of size 3300 × 97 for all 33 signs. Out of complete dataset, 90% dataset (belongs to 9 student) is used for training and 10% dataset (belongs to 10th student) is used for CV.

IV. CLASSIFICATION

A. *Generalized Feed Forward Neural Network*

Following trials have been performed on Multilayer Perceptron Neural Network (MLP) 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. 4 shows that the minimum MSE is given by processing element (PE) number 13.

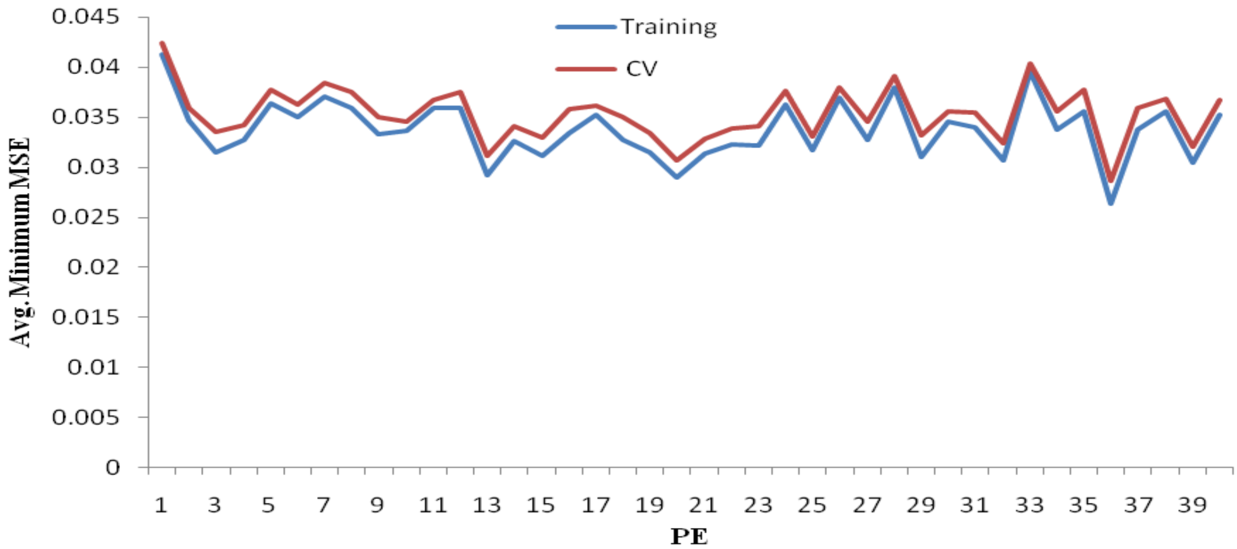

Fig. 4. Processing Element (PE) Vs Minimum MSE

Different transfer function like Tanh, LinearTanh, Sigmoid, LinearSigmoid, 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. GFNN with the following parameter setting gives maximum Percentage classification accuracy of 97.34 % on training and 96.36 % on CV dataset. Table I & II shows confusion matrix for signs of CV data and correct classification accuracy respectively. Tagging of Data: 90% for Training & 10% Cross validation

Input Layer:
Input PE - 97 Exemplars - 2970
Hidden Layer:
PE - 13 Transfer Function - Tanh
Learning Rule - Momentum Momentum - 0.7
Step Size - 0.1
Output Layer:
Output PE's - 33
Transfer Function - Tanh Learning Rule - Momentum
Momentum - 0.7 Step Size - 0.1

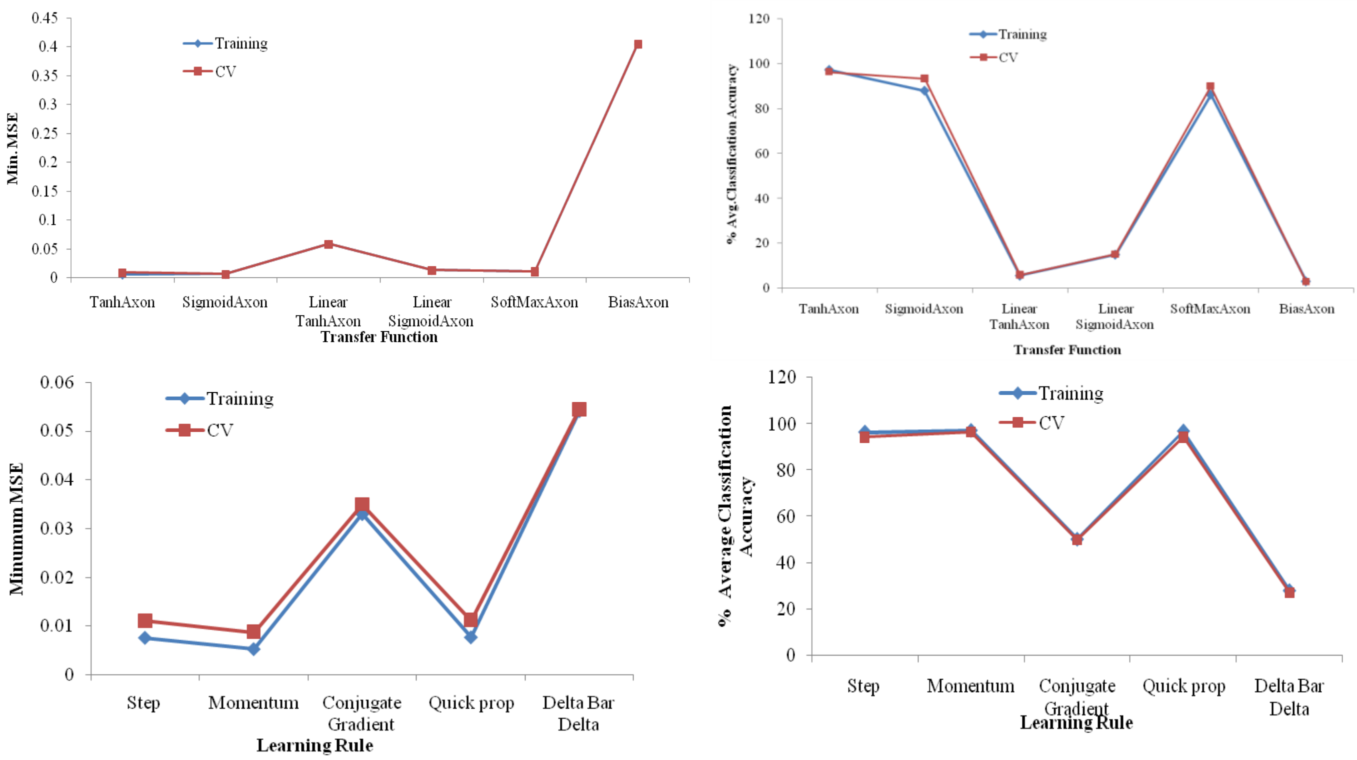

Fig. 5. Variation of Minimum MSE and Percentage average classification accuracy with different transfer functions and learning rules

TABLE I. CONFUSION MATRIX FOR CROSS VALIDATION (CV)

Output / Desired	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	10	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	0	0
4	0	9	0	0	0	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	0	0
5	0	0	10	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
6	0	0	0	10	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
7	0	0	0	0	10	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	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	0	0	0	0	0	0
9	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	10	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
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	0	0
C	0	0	0	0	0	0	0	0	0	10	0	0	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	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	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	10	0	0	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	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	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	0	10	0	0	0	0	0	0	0	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	0	9	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	1	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	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	0	10	0	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	0	0	0	0	0	0	0	0	10	0	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	0	0	0	0	0	0	0	0	9	0	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	0	10	0	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	0	9	0	0	0	0	0	0
U	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	9	0	0	0	0	0
V	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	10	0	0	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	0	9	0	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	1	0	0	0	8	0	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	0	10	0
Z	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	0	0	10

TABLE II. CORRECT RATE OF SIGN POSTURE RECOGNITON USING GFF NEURAL NETWORK

Sign	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
% correct Classification	100	90	100	100	100	90	100	100	80	100	100	100	90	100	100	100	100	100	90	100	100	100	100	100	90	100	90	90	100	90	80	100	100

B. Multilayer Perceptron Neural Network

Like GFF Neural Network we have performed similar trials using MLP. With the following parameter setting we have got maximum Percentage classification accuracy of 95.01% on training and 94.84% on CV dataset. Tagging of Data: 90% for Training & 10% CV

Input Layer:
I/p Processing Element - 97 Exemplars - 2970
Hidden Layer:
Processing Elements - 24
Transfer Function - Tanh Learning Rule - Momentum
Momentum - 0.7 Step Size - 0.1
Output Layer: Output PE's - 33
Transfer Function - Tanh Learning Rule - Momentum
Momentum - 0.7 Step Size - 0.1

C. Support Vector Machine

We have varied epoch & number of runs by fixing the step size at 0.1. It is observed that from epoch 2 onwards, there is very small change is MSE as shown in Fig 6.

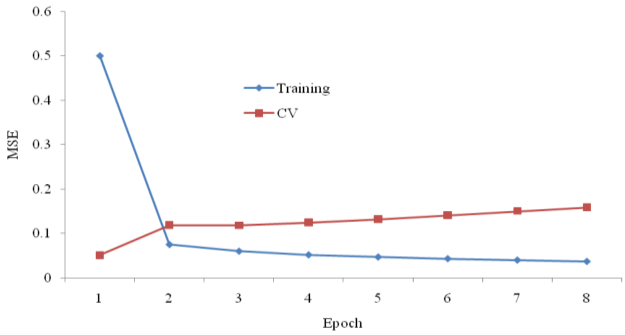

Fig. 6. MSE Vs Epochs

After experimentation we have observed that the best result is i.e. 96.86 % on training and 90.90% on CV data set with optimal parameter setting as below

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

No. of Epoch - 02 No. of Runs - 1
Input processing Elements - 97
Output Processing Elements - 33 Exemplars - 2970
Step Size - 0.8 Kernel Algorithm - Adatron

V. RESULT

We have obtained maximum Average classification accuracy as 96.36 % on CV dataset using GFF Neural Network as shown in Table II. As compared to other researchers as mentioned in related work, we can say that our work is much satisfactory. The approach presented in this paper is able to recognize all alphabets A to Z and numbers 0 to 9 of ISL. However number posture 0, 1, 2 are not included in recognition system due to similarity with alphabet posture O, I, V respectively. The approach has one constraint that Leap Motion Sensor has to be kept little inclined and only static signs are recognized.

VI. CONCLUSION

Finally we came to conclude that although GFFNN classifier gives satisfactory percentage classification accuracy but from Table I it can be observe that few signs (4,8,B,F,L,R,T,U,W,X) are not perfectly recognized. However the classification accuracy of these signs can be improve if more or some distinct features are extracted and tested on similar or different classifiers.

REFERENCES

[1] Dixit, Karishma, and Anand Singh Jalal, "Automatic Indian sign language recognition system." IEEE 3rd International Conference on Advance Computing (IACC), pp. 883-887, 2013.

[2] Adithya V, Vinod P. R., Usha Gopalakrishnan, " Artificial Neural Network Based Method for Indian Sign Language Recognition ",Proceedings of IEEE Conference on Information and Communication Technologies, pp. 1080-1085, 2013.

[3] Rekha, J., J. Bhattacharya, and S. Majumder, "Shape, texture and local movement hand gesture features for indian sign language recognition." IEEE 3rd International Conference on Trendz in Information Sciences and Computing (TISC), pp. 30-35, 2011

[4] Saxena, Ankur, Deepak Kumar Jain, and Achintya Singhal. "Sign Language Recognition Using Principal Component Analysis." Fourth IEEE International Conference on Communication Systems and Network Technologies (CSNT), pp. 810-813, 2014.

[5] Bhuyan, M. K., Debashis Ghosh, and P. K. Bora. "A framework for hand gesture recognition with applications to sign language." Annual IEEE In India Conference, pp. 1-6, 2006.

[6] Bhuyan, Manas Kamal, Mithun Kumar Kar, and Debanga Raj Neog. "Hand pose identification from monocular image for sign language recognition." IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 378-383, 2011.

[7] P.V.V.Kishore, M.V.D.Prasad, Ch.Raghava Prasad, R.Rahul,"4-Camera Model for Sign Language Recognition Using Elliptical Fourier Descriptors and ANN", International Conference on Signal Processing And Communication Engineering Systems (SPACES), pp.34-38, 2015.

[8] Ren, Zhou, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. "Robust part-based hand gesture recognition using kinect sensor." IEEE Transactions on Multimedia, pp. 1110-1120, 2013.

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

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

[11] Wikipedia. "Leap Motion" Wikipedia.org. [Online]. Available https://en.wikipedia.org/wiki/Leap_Motion [Last Modified: 26 July 2015, 07:16].

Prof. Rajesh B. Mapari has received his B.E Degree from Sant Gadge Baba Amravati University Amravati, India, M.Tech.Degree from Rashtrasant Tukadoji Maharaj Nagpur University and currently doing Ph.D. degree in Engineering and Technology faculty at Amravati University. He is currently Assistant Professor in Anuradha Engineering College, Chikhli, India. He has published many papers in National & International Journals and conferences. His interests include Image processing, Sign language analysis, Human computer interaction etc.. He is currently Life member of Technical bodies like ISTE, IEI.

Dr. G. U. Kharat received his B.E. Degree in 1989 from Amravati University & M.E. Degree in 1999 from Shivaji University, India Ph.D. degree in 2009 in Engineering & Technology faculty at Amravati University. He has moer than 25 years teaching & 10 years of administrative experience. He has guided over 150 projects at U.G. level. He has over 25 papers in his credit at National & International Conference and Journals. He is Fellow member of Technical bodies like ISTE, IETE, IJERIA, IJMSEA. He is currently guiding 5 Ph.D. scholars. He is Professor/Principal at SharadChandra Pawar college of Engineering Otur, India. His current research interests include Signal processing, ANN, HCI, etc.