International Journal of Future Generation Communication and Networking
Vol. 14, No. 1, (2021), pp. 832–843

Recognition of South Indian Sign Languages for Still Images Using Convolutional Neural Network

Ramesh M. Badiger[1], Dr. Dharmanna L[2]

[1]*Department of CSE, Tontadarya College of Engineering, Gadag, Karnataka, India*
[2] *Department of ISE, SDMIT Ujire, Karnataka, India*

rameshmbadiger@gmail.com , dharmannasdmit@gmail.com

Abstract:
Recognition of gestures is a growing area of study. Being an integral part of hand gestures for non-verbal communication plays a crucial role in our everyday lives. The Hand Gesture recognition system provides us with a artistic, natural, comprehensible way to connect with the machine that human beings are more familiar with. This paper aims to present a hand gesture recognition based on input image segmentation and hand gesture classification based on the Convolutionary Neural Network by keeping in mind the resemblance of the human hand form with four fingers and one thumb. Here we have suggested a method for understanding the gestures of isolated images of South Indian Sign languages (Kannada and Telugu) that is a stride towards supporting and refining people with hearing and speech disability. With a Convolutional Neural Network, we have carried out the classification of gestures. In the elements gestures recognition, and it's needed to differentiate the each spatial and worldly developments and this paper proposes a way for dynamic gesture recognition.

Keywords: Gestures recognition, Convolutional Neural Network (CNN), Deep Neural Network (DNN), Static hand gesture recognition, South Indian Sign language recognition.

I. INTRODUCTION:

Extraordinary strategies of contact with the machine have been established with the advancement of technology. More historically, it has been used with a mouse, electromagnetic gloves and then track pads, joysticks, keyboards etc. It's been seen before. Apart from these techniques, gesture recognition is often used and this work can be contemplated because it mimics daily verbal exchange with a human being in a more realistic way of interplay. Hand gesture recognition has many packages they are robot tool manipulation, widespread laptop interplay, signal languages recognition, game play, etc.

Hand Gesture may be noticed has a significant body movement. In order to convey the message or significant information to interact with the environment hands, body which can involve hands and the body is used these gestures may be static, dynamic or each [1].

In overall, motion recognition cans be isolated into remarkable regions and first one is vision- primarily based completely on methodologies and techniques which require uncommon equipment like gloves, arm band and outline pack [2, 3]. The second kind is relatively tough for the consumer in view that it is required to carry efficient hardware. This type of systems uses efficient specified appeal like Unmanned Arial Vehicle (UAV) [4]. On the other hand, vision-primarily based totally systems use techniques like, pattern recognition, Image

ISSN: 2233-7857 IJFGCN

segmentation and item detection, image processing, image extraction, [1]. The development of a vision-based approach gives the client with a non-obstructive interface and in this way makes a lift in this locale [5].

Here we have trained 1867 Images using both CNN and DNN and compared the results. In this paper we have developed a Vision Based system which is centered on hand gesture detection form still image. Here we created dataset consisting of 2042 still images having 34 different hand signals corresponding to South Indian Sign Languages. Then we applied preprocessing and segmentation techniques on images and then used Convolutional Neural Networks for classification.

II. LITERATURE SURVEY

S. Mitras and T. Acharyas [1] make a case for that the signal recognition could be a vital subject that translates human gesture through Computer vision calculations. There unit various real movements which might start signal however the common sort of motion beginning comes from the front hands. The full strategy of chase motion to their outline and changing them to a few deliberate commands is caught on as signal recognition. Various innovations has been utilized for the see and usage of such very gadgets, in any case contact based generally and vision based advances unit 2 fundamental shapes of innovations utilized for strong, adjust and dependable hand motion acknowledgment frameworks

T. Baudel and M. Beaudouin [2] appears elective ways these modalities may too be utilized and portrayal a few common interface pointers rundown of promising ranges for future investigation. Our inspiration for scripting this is often to goad designers construct compelling interfacing which will make discourse and signal as common on the desktop since the console and mouse.

K.-Y. Lian, C.-C. Chiu, Y.-J. Hong et.al,[3] analyses the progressive of period hand gesture recognition models victimization myogram knowledge and machine learning. He analyzed sixty five primary studies following the Kitchenham methodology. Supported a typical structure of machine learning-based systems, and additionally analyzed the structure of the projected models and standardized ideas in reference to the categories of models, knowledge acquisition, segmentation, preprocessing, feature extraction, classification, post process, real-time operation, varieties of gestures, and analysis metrics

Y. Ma, Y. Liu, et. al,[4] given the equipment fashion, mechanical wonder recognition of arm development and so the expound structure of a convolutional neural organize (CNN) framework utilized for period hand motion acknowledgment backed MMG signals. This strategy accomplished 94% precision for 5 motions with straight forward action for each client, subsequently giving a natural signal based for the most part UAV framework.

Pavlovic, R. Sharma and T. Huang[5] said and upheld gestural frameworks furthermore as diverse potential applications of vision-based signal acknowledgment and conjointly highlighted bearings of future examination in motion acknowledgment, at the side its integration with distinctive characteristic modes of human-computer interaction.

ISSN: 2233-7857 IJFGCN

Bhowmick, Sourav; Sushant;et.al. [6] studied the different approaches to modeling, investigation and acknowledgment of hand motions for visual elucidation said 2 categories of models utilized inside the visual translation of hand motions. The essential depends on 3D models of the human hand, though the moment utilizes the looks of the human hand inside the picture. The 3D hand models give a costly portrayal and segregation capability that might empower a expansive category of motions to be recognized coming about in characteristic HCI.

Chung, Hung-Yuan; Chung, Yao-Liang [7] utilized a advanced camera to immediately track the locale of intrigued (ROI), to be specific, the hand locale, inside the picture shift and build up hand signals for family apparatus administration (in arrange to form sensible homes) or human-computer interaction areas. Firstly, author used color discovery and morphology to induce rid of unnecessary foundation information from the picture, and after that utilize foundation subtraction to find the ROI. Following, to dodge foundation impacts on objects or commotion moving the ROI, we tend to utilize the kernelized relationship channels (KCF) run the show to follow the identified ROI.

III. MATERIALS AND METHODS

In the present work we have considered 34 different Sign of South Indian Sign Languages. The dataset constructed by capturing the images which contains the sign using the mobile camera. Table 1 summarizes the technical specifications regarding the image acquisition.

Table 1: Technical Specifications for Image Acquisition

Sl. No	Specification
1	Type of the Mobile: Samsung A51
2	Resolution:1080*2400 pixels
3	Focal Length: f/2.0 aperture
4	Lightning: Normal Day Light
6	Distance of Object from the camera:1mts
7	Back Ground: Dark(Black, Green)

This chapter explains the various steps associated with recognition of South Indian Sign Languages using still Images Figure 1 represent the block diagram of proposed methodology.

As mentioned earlier there are 2042images constructed. This dataset consists of 34 different hand sign of the South Indian Sign Languages. Table 2 summarizes the details of the sign languages and number of images corresponding to two sample signs. A complete list of the same is provided in Appendix A.

Table 2: Details of the signs and number of images corresponding to each image of South Indian Sign Languages.

Sign Language	Sample Image	Number of Images	Letter Corresponding to Image
Kannada		60	ಕ
Telugu		60	క

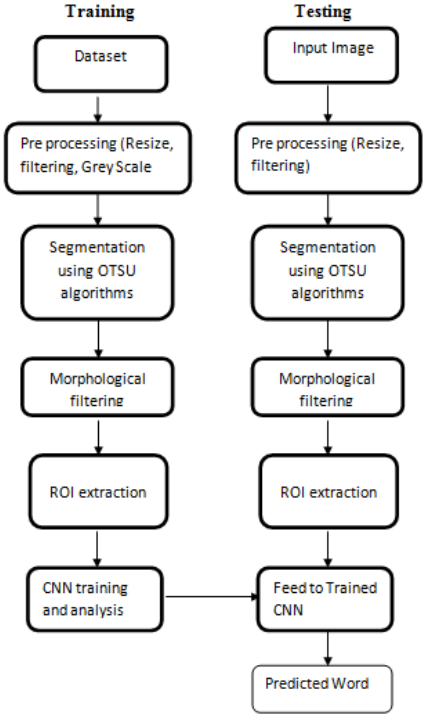

Figure 1: Block Diagram of Proposed method

Preprocessing: It will become an unwritten rule or a mandate for any image processing gadget to undergo a inflexible pre-processing. The manner begins off evolved with filtering noise, accompanied via way of means of an image adjust; which has proved to be an amazing congregation shape for pre-processing images.

Thresholding

The pre-processing additionally includes image thresholding the usage of Otsu's approach for disposing of history scene, when you consider that it'd emerge as an inconvincible hassle to image with many noise interferences from the history objects.

In Computer imaginative and prescient and image processing, Otsu's method, named after Nobuyuki Otsu, is used to carry out computerized image thresholding. In the only form, the set of rules returns an unmarried depth threshold that separate pixels into classes, foreground and background. This threshold is decided with the aid of using minimizing intra-magnificence depth variance, or equivalently, with the aid of using maximizing inter-magnificence variance.

Morphological operation:
Once Otsu's thresholding is applied, then we have applied morphological operation for removing unwanted region and providing proper shape to segmented region followed by bitwise operation for extraction color hand region with respect to segmented image.

Region of Interest
Thus we achieve ROI (Region of Interest) extraction using OTSU thresholding method.

Convolutional neural networks (CNN)
CNNs had been one of the maximum influential improvements withinside the discipline of laptop imaginative and prescient as highlighted in Figure 2. They have achieved lots higher than conventional laptop imaginative and prescient and feature produced present day results.

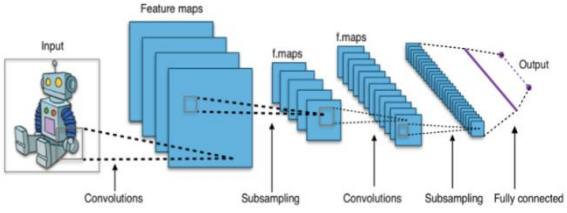

Figure 2: Convolutional Neural Network

The convolution layer computes the yield of neurons which could be related to adjacent zones or responsive areas withinside the center, each computing a speck item among their weights and a little open subject to which they're related to withinside the enter volume. Each computation comes about in extraction of a work outline from the enter image. Completely different words, consider you've got an image spoken to as a 5x5 network of values, and you're taking a 3x3 lattice and slide that 3x3 window or bit over the image. At each work of that network, you duplicate the values of your 3x3 window with the help of utilizing the values withinside the image which could be directly being covered with the help of utilizing the window.

Gray scale conversion formula
For each pixel (r g b) at (I, J) do:
grayPixel[I,J] = 0.21*r + 0.71*g + 0.07*b
Resized Image Resolution: 221*480

As a result, we get a single wide assortment that speaks to all of the values in that window of the pictures. We utilize this store to sifting: since the window developments over the image, you test for styles in that stage of the image. This works due to channels, that are expanded with the help of utilizing the values outputted with the help of utilizing the convolution.

The objective of subsampling is to induce an enter outline through way of implies of diminishing its measurements, which encourages in diminishing over fitting. One of

the methodologies of subsampling is max pooling. With this method, you choose out the exceptionally best pixel taken a toll from a place relying on its length. In several words, max pooling takes the greatest taken a toll from the window of the picture directly included by means of way of means of the part as appeared in Figure 3. For illustration, you'll have a max-pooling layer of length 2 x 2 will choose out the foremost pixel profundity fetched from 2 x 2 places. You're appropriate to accept that the pooling layer at that point works plenty just likes the convolution layer! You additionally take a bit or a window and stream it over the picture; the leading refinement is the highlight usually carried out to the bit and the picture window isn't direct.

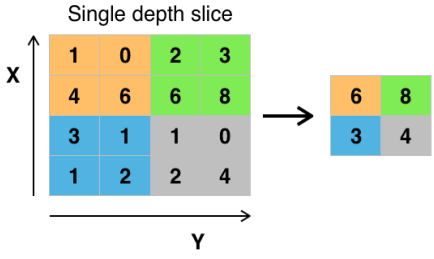

Figure 3: Max-Pooling from

3. The goal of the completely linked layer is to flatten the high-stage capabilities which can be found out via way of means of convolutional layer and mixing all of the capabilities. Its passes the flattened output to the output layers in which you operate a softmax classifies or an sigmoid to are expecting the enter magnificence labels.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 126, 126, 32)      896
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 63, 63, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 61, 61, 64)        18496
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 30, 30, 64)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 28, 28, 64)        36928
_________________________________________________________________
flatten_1 (Flatten)          (None, 50176)             0
_________________________________________________________________
dense_1 (Dense)              (None, 64)                3211328
_________________________________________________________________
dense_2 (Dense)              (None, 34)                2210
=================================================================
Total params: 3,269,858
Trainable params: 3,269,858
Non-trainable params: 0
_________________________________________________________________
Total number of Images 1867
```

Figure 4: Proposed DNN Layer

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_9 (Conv2D)            (None, 30, 30, 32)        896
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 15, 15, 32)        0
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 13, 13, 64)        18496
_________________________________________________________________
max_pooling2d_7 (MaxPooling2 (None, 6, 6, 64)          0
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 4, 4, 64)          36928
_________________________________________________________________
flatten_3 (Flatten)          (None, 1024)              0
_________________________________________________________________
dense_6 (Dense)              (None, 64)                65600
_________________________________________________________________
dense_7 (Dense)              (None, 36)                2340
=================================================================
Total params: 124,260
Trainable params: 124,260
Non-trainable params: 0
```

Figure 5: Proposed CNN Layers

IV. DEEP NEURAL NETWORK (DNN)

As DN community (DNN) is on synthetic neural community (ANN) with more than one layer's among the center and output layer's. There are specific sorts of neural network however they usually encompass identical additives: functions, synapses, biases, weights and neurons. These additives functioning much like the humans brain it may be educated like some other Machine Learning algorithms. For an ex, a DNN this is educated to apprehend hand gesture will moved over given image and it calculates the possibility that the hand vicinity withinside the image it is a positive gesture.

Consumer can be evaluation the effects and pick out the possibilities the community need to displays (above a positive thresholds, etc.) and go back the propose labels. Each one of the mathematical manipulations and is taken into consideration a layers, and the complicated and in DNN has many layers, therefore the name "deep" networks as in the Figure 4. DNN has version complicated in discriminate affinity. DNN architectures generate compositional fashions in which the item is expressed as a layered composition of primitives. The more layers permit composition of capabilities from decrease layers, probably modeling complicated records with fewer devices than a further acting shallow community.

ISSN: 2233-7857 IJFGCN

V. RESULTS AND DISCUSSIONS:

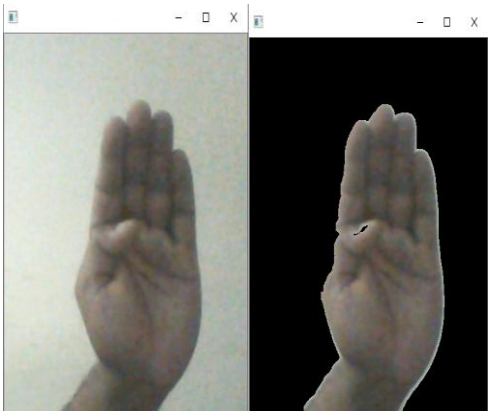

Figure 6: Input Image and Segmented Image

Figure 6 appears the Input Image and Fragmented image. We will isolated or fragment the picture into distinctive parts called Portions. It is anything but a great thought to bargain with the total picture at the same time as there will be areas within the picture which do not contain any information. By segregating the picture into fragments, we will be ready to utilize the critical fragments for handling the image.

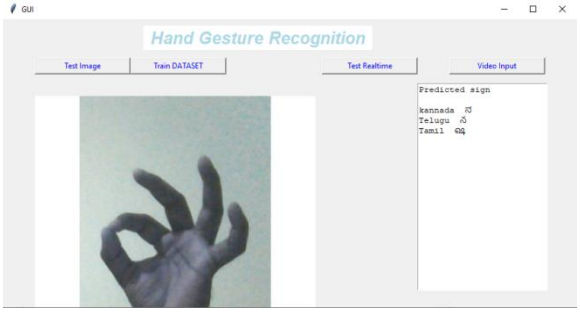

Figure 7: GUI Output of Hand gesture Prediction of Still Image in South Indian Sign Languages (Kannada, Telugu and Tamil).

Figure shows the GUI Output of Hand gesture Prediction of Still Image in South Indian Sign Languages (Kannada, Telugu and Tamil). Figure 8 and Figure 9 shows the DNN Training Accuracy and Training Validation.

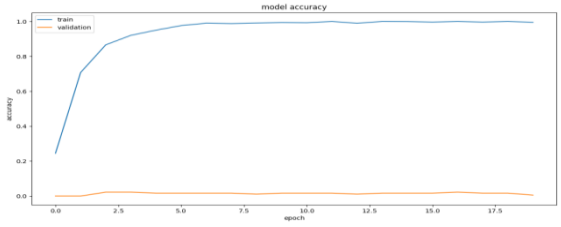

Figure 8: DNN Training Accuracy vs Training Iteration

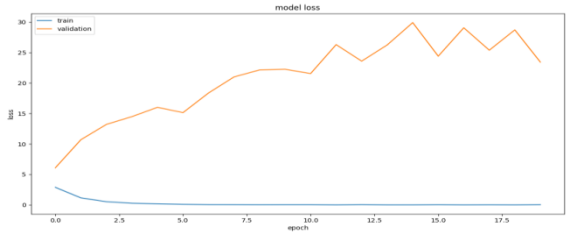

Figure 9: DNN Training Validation vs Training Iteration

Figure 10 and Figure 11 shows the CNN Training Accuracy and Training Validation

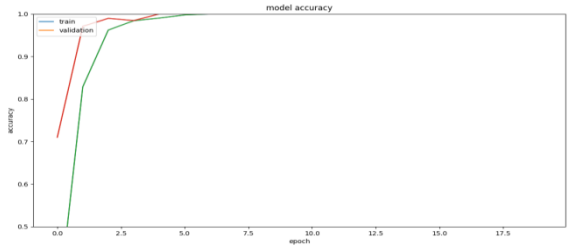

Figure 10: CNN Training Accuracy vs Training Iteration

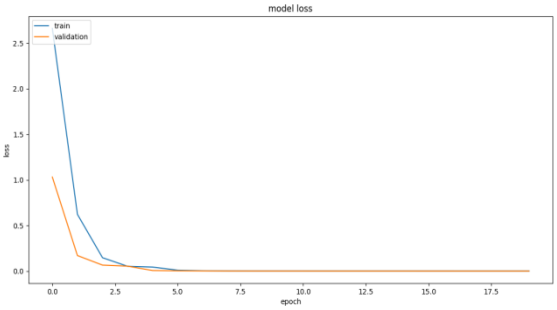

Figure 11: CNN Training Validation vs Training Iteration

Figure 12 shows the comparison of training accuracy for CNN and DNN.

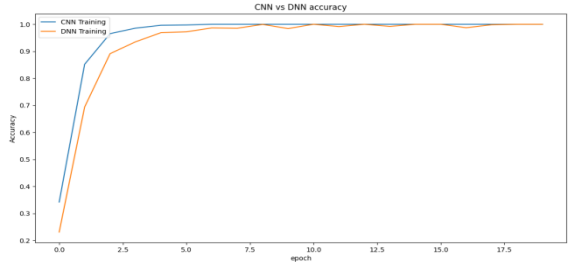

Figure 12: CNN VS DNN Training Accuracy

Figure 13 shows the comparison of training loss for CNN and DNN.

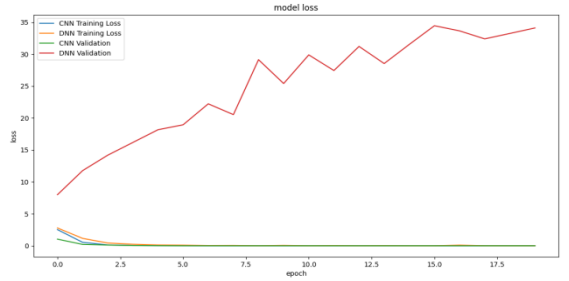

Figure 13: CNN and DNN Training loss and Validation

The results for the parameters like Accuracy, Validation Loss and Validation Accuracy are summarized in Table 3.

TABLE 3: Machine Learning Analysis

Training Results	Accuracy	Validation Loss	Validation accuracy:
DNN	**0.986**	**25.0464**	**0.9969**
CNN	**0.997**	**1.8926e-04**	**1.0000**

The results of recognition of South Indian Sign Languages from still Images prefer the CNN. In the implementation of recognition with the parameters like Accuracy, Validation Loss and Validation Accuracy CNN shows the high accuracy over DNN. Validation loss in the CNN is less compared to DNN. Validation Accuracy of CNN is grazing above DNN. Thus the CNN module to recognize the South Indian Sign Language is preferred.

VI. CONCLUSION

In this paper, we have compared two approaches for machine vision based analysis for hand gesture and correspondingly recognized South Indian Sign languages (Kannada, Telugu) and alphabets. The work preceded with images of South Indian Sign languages as input. The input images are segmented and classified using both CNN and DNN. The comparison of these two approaches is analyzed based on three metrics viz. (1) Accuracy (2) Validation Loss and (3) Validation accuracy. The results are also represented using graphical forms. The comparative analysis led to following conclusions: (1) Accuracy of CNN is better as compared with DNN (2) Loss in CNN is also less as compared to DNN and (3) Thus Convolutional Neural Networks is preferred in recognition of South Indian Sign Languages for still images.

REFERENCE

[1] S. Mitra and T. Acharya , “Gesture recognition: A survey," IEEE Transactions on Systems, Man, and Cybernetics, vol. 37, no. 3, pp. 311 - 324, 2007. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[2] T. Baudel and M. Beaudouin-Lafon, "Charade: remote control of objects using free-hand gestures," Communications of the ACM, vol. 36, no. 7, pp. 28-35, 1993.
[3] K.-Y. Lian, C.-C. Chiu, Y.-J. Hong and W.-T. Sung, "Wearable armband for real time hand gesture recognition," in IEEE International Conference on Systems, Man, and Cybernetics, Canada, 2017.
[4] Y. Ma, Y. Liu, R. Jin, X. Yuan, R. Sekha, S. Wilson and R. Vaidyanathan, "Hand gesture recognition with convolutional neural networks for the multimodal UAV control," in Workshop on Research, Education and Development of Unmanned Aerial Systems, Sweden, 2017.
[5] V. Pavlovic, R. Sharma and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 677 - 695, Jul 1997.

ISSN: 2233-7857 IJFGCN

[6] Bhowmick, Sourav; Kumar, Sushant; Kumar, Anurag (2015). [IEEE 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS) - Kolkata, India (2015.7.9-2015.7.11)] 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS) - Hand gesture recognition of English alphabets using artificial neural network., 405–410. doi:10.1109/ReTIS.2015.7232913

[7] Chung, Hung-Yuan; Chung, Yao-Liang; Tsai, Wei-Feng (2019). *[IEEE 2019 IEEE International Conference on Industrial Technology (ICIT) - Melbourne, Australia (2019.2.13-2019.2.15)] 2019 IEEE International Conference on Industrial Technology (ICIT) - An Efficient Hand Gesture Recognition System Based on Deep CNN. , (), 853–858.* doi:10.1109/ICIT.2019.8755038

Bibliography

Prof. Ramesh M. Badiger completed B.E in 2009,M.Tech in 2011 and perusing PhD and has 10 year of teaching experience. Received Second Prize, International National Conference for the paper entitled "Text Extraction From Road Side Display Boards Using Wavelet and SVM" held at TRINITY (isle) Bangalore and also Second Prize, National Conference for the paper entitled "Cross Layer Adaptation in Media Delivery in Wireless Sensor Network "held at J.N.N I.T CHENNAI. Ha has published many national and International papers. Trained in CISCO CCNA by Amrita University, network Academy Coimbatore. Awarded Adarsh Vidya Saraswati Rastriya Puraskara by Global Management Council, Ahmedabad.

Dr. Dharmanna L, is currently working as HOD of information science and engineering at SDM Institute of technology, Ujire, Karnataka. He obtained M.Tech degree from VTU, Belagavi. He was awarded Ph.D. Degree in computer science and engineering from VTU, Belagavi. Earlier worked as Sr. Lecturer in the department of Computer Science Engineering from Nov 2007 to August 01 2009 at East West Institute of Technology (EWIT) Bangalore. Also he worked as Lecturer in the department of Computer Science Engineering from Nov 05 2005 to Nov 01 2007 at B.T.L. Institute of Technology Bangalore His research interests includes Medical Image Processing, Embedded System design, Pattern Recognition and Computer Networking. He has published around 20 papers in reputed Journals and conferences

Appendix A

Details of the sign languages and number of images corresponding to two sample signs

Letter in Kannada Language	Letter in Telugu Language	Number of Images Used	Sign	Letter in Kannada Language	Letter in Telugu Language	Number of Images Used	Sign	Letter in Kannada Language	Letter in Telugu Language	Number of Images Used	Sign
ಂ	ం	60		ದ	ద	60		ಮ	మ	60	
ಖ	ఖ	60		ಧ	ధ	60		ಯ	య	60	
ಗ	గ	60		ನ	న	60		ರ	ర	60	
ಘ	ఘ	60		ಪ	ప	60		ಲ	ల	60	
ಙ	ఙ	60		ಫ	ఫ	60		ವ	వ	60	
ಚ	చ	60		ಬ	బ	60		ಶ	శ	60	
ಛ	ఛ	60		ಭ	భ	60		ಷ	ష	60	
ಜ	జ	60		ಮ	మ	60		ಸ	స	60	
ಝ	ఝ	60		ಯ	య	60		ಹ	హ	60	
ಞ	ఞ	60		ಱ	ఱ	60		ಳ	ళ	62	
ಟ	ట	60		ಬ	బ	60					
ಠ	ఠ	60		ಭ	భ	60