Detection of Indian Sign Language using Deep Learning

Akshay Madhusudan Verulkar [1], Prof. Dr. Suhas Patil[2]

[1]M. Tech Student, Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be University) College of Engineering, Pune (India)

[2]Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be University) College of Engineering, Pune (India)

[1]akshay.verulkar@gmail.com , [2]shpatil@bvucoep.edu.in

Abstract:

Hand motion detection is a critical step in the automatic translation of sign language, which allows deaf people to communicate. This research pursued to investigate the difficulties in character categorization in Indian Sign Language (ISL). In the related discipline of Sign Language, a considerable of the study was done, and the same could be said for ISL. The absence of appropriate datasets, occluded characteristics, and language variability with neighborhood have all been major roadblocks, resulting in little ISL research. In this paper, we proposed ISL detection and recognition using a deep learning framework. The Convolutional Neural Network (CNN) has been used for detecting the activity using the RESNET-100 deep learning framework. The feature extraction and selection have been done in each convolutional layer and optimization has been done in the pooling layer. Multiple activation functions have been utilized in dense layers for classification on real-time hand gesture datasets as well as MNIST datasets. An extensive experimental analysis proposed system results has been evaluated with various machine learning algorithms, that produces almost 5-7% higher accuracy with state-of-art methods.

Keywords: Hand gesture, Sign language, Indian signs, Deep Learning, Convolutional Neural Network, Object detection, Object movement.

DOI Number: 10.14704/nq.2022.20.9.NQ44358 Neuro Quantology 2022; 20(9):3101-3108

I. INTRODUCTION

This template Sign language is a form of correspondence used by people who are deaf or hard of hearing. Deactivated People use multiple languages movements as a non-verbal communication technique to communicate their feelings and ideas to other people. However, since these ordinary people have a hard time understanding their expressions, experienced sign language experts are required for medical and legal appointments, as well as educational and training sessions. There has been an upsurge in demand for all these services in recent years. Other types of services, such as video distant location human interpreter using an increased Connection to the internet, have been presented, providing an easy-to-use sign language interpreter service that can be used and stands to benefit, but with significant limitations. To address this, we created a custom CNN model to recognise sign language gestures. Four Convolution, multiple Max-Pooling layers, two dense layer upon layer, one smoothing layer, and one absent / if these factors are used to create a deep neural network with several layers. To train the model to recognise the gesture, we utilise the MNIST American Sign Language Dataset. The collection includes the characteristics of several augmented gestures. Using Open-CV, created a custom CNN (Fully Convolutional) model to recognise a sign from a video frame. as shown in Figure 1, sign language innovations are linked in a way that has an impact on one another. The precise extraction of hand movements actions, as well as face emotions, is vital to the development of machine learning which are responsible for reliable sign identification. Furthermore, the performance of sign vernacular interpretation and portrayal methods is significantly influenced by accurate sign language recognition. Developments in sensory devices and artificial intelligence have cleared the path for the

creation of sign language apps that will greatly assist hearing-impaired persons in their daily lives.

Previous studies have mostly focused on particular sign computational linguistics, such as video as well as sensor-based multiple languages recognition [3–7] and sign translation software [8,9]. Recent reviews have presented sign language systems to assist wanting to hear individuals in education and learning, as well as in voice and text perception systems [10,11]. Notwithstanding, there is no comprehensive overview of all sign speech processing and their interrelationships. This study intends to close this gap by presenting AI advances in all sign machine translation, from intercepting and appreciation to translation as well as recognition, and concluding with a description of recent gesture applications that can significantly improve communication between wanting to hear as well as speaking people. The major goal of this review is to show how important it is to use AI technology in sign language to help deaf and wanting to hear persons communicate with other people. Furthermore, the goal of this study is to familiarise academics the with nation in all show machine translation and to suggest scope of research that will aid in the creation of even more precise techniques and conventional goods for the Deaf population. The following are examples of the study's objectives in more detail:

- A complete review of the application of AI technologies in different sign language tasks (i.e., capture, identification, interpretation, and reproduction) is offered, as well as their significance in their area.
- The benefits and drawbacks of modern symbol language technologies, as well as their interrelationships, are discussed and analysed.
- Future paths in the integration of Advanced technology for hand signals are offered to help aspiring sign language academics.

II. LITERATURE SURVEY

Communicating is one of the most fundamental prerequisites for societal life. Individuals with hearing loss communicate with one another using gestures, which are difficult to understand for non-hearing persons. In today's world, almost every spoken languages has its own established gesture-based communications. It is critical to provide an understanding of sign speaking and listening for various populations that are unfamiliar with the signs [2]. It's a system of hand signals that includes visual gestures and signs. Conversational signals, regulating gestures, manipulating motions, and instructive motions are some of the other types of hand motions [1]. The use of a few parts of our body, such as our fingers, hand, and arm, to express data is known as sign language. A system that recognises hand movements from a digitally enhanced dataset is known as recognition and image. This technique is now used in a wide range of applications, including robotics and telerobotics, games and virtual reality, and human–computer interaction (HCI). The majority of the presented models rely on traditional example recognition, which necessitates human competence for extraction and recognition. Deep learning's recent success, particularly the Residual Neural Network (ResNet) for machine vision [4], [23], is being utilised to detect Indian Sign Language (ISL) as an object recognition challenge. This ResNet model is built on a deep Learning Neural Network (CNN) [8]. This is one of the top image categorization models.

Starner et al. [9] suggest employing Markov Chain Models to categorise the outcome, trajectory parameters, and alignment structure of the Multiple languages (HMM). This technique's intrinsic features make it suitable for use in gesture recognition. Grobel K et al. [10] provided a technique in which a total of 262 signs were gathered from two distinct endorsers, with a precision of 95% using an HMMs classifier. When the database is taught and evaluated using the signals of various people, the accuracy is considerably reduced. Artificial neural network-based approaches for real-time American Sign Language (ASL) identification were suggested by N. Sarawate et al. [11]. The Microsoft Kinect sensor is utilised in the study to detect signs for two applications: arithmetic calculation and the rockpaper-scissors game, with an accuracy of more than 90% for ASL. Setting subordinated HMMs and a technique for linking three-dimensional strategies on ASL were used by Vogler C et al. [12]. The framework grouped 53 ASL with an accuracy of 89.91

percent. Convolutional Neural Network (CNN) was utilised by N.Priyadharsini et al. [13] to recognise sign for Indian Sign Languages. To identify Arabic Sign Language, O Al-Jarrah et al [14] employed a neural network. Deora, D., and N.Bajaj [15] developed an Indian Sign Identification System for twenty-five English letters and nine number signs. This study used PCA for sign classification and Segmentation for fingertip method for feature extraction. The accuracy of this method was 94 percent. G. Balakrishnan, P. S. Rajam, and others suggested a technique for detecting Tamil Sign letters in their paper [3]. This approach employed images with a resolution of 640x480 pixels. The photographs are then transformed to greyscale images. This technique achieved a precision of 96.87 percent for the static method and a rate of 98.75 percent for the dynamic method.

The suggested technique is divided into many parts. We collected data from a standard dataset [22] in the first stage. We turned each picture to grayscale and resized it to 6464 pixels after gathering the data. After that, we used normalising methods to transform the grey level data from 0-255 to a range of 0 to 1 values. In the second stage, we transform each picture from two to three dimensions. We employed a residual block in the third phase, which comprises of convolution operation with much the same number of power channels, a batch normalisation layer, and a ReLU activation function [23]. We skip two convolution processes in the convolution layer and put the input well before final Relu activation illustrated in figure 1. In this stage, we applied the algorithm to recognise Indian Sign Language by extracting the attributes of picture data atomically. To increase efficiency, we used an optimization approach (Adam Optimizer [19]) in the fourth phase. Finally, we employed real-time augmentation to improve the variety of data accessible for the training set in the final stage.

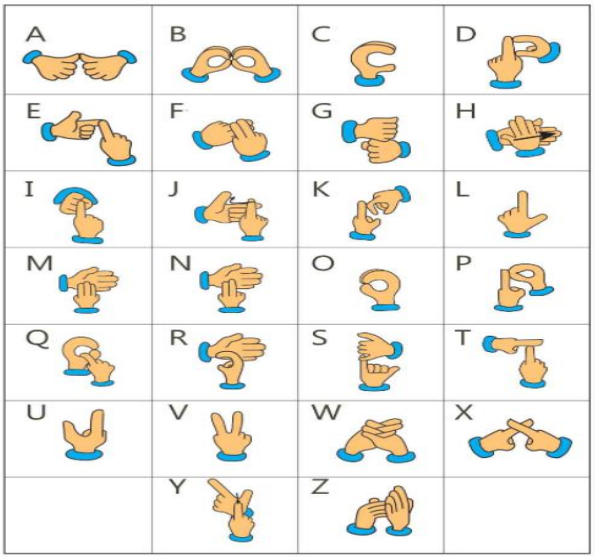

Figure 1 : Indian sign language sample dataset

The many pictures of Indian Sign Languages come from a common dataset [22]. We gathered 24 classes of distinct signals from this dataset, all 4972 imges. Then, in order to provide fixed length input to our suggested model, we turned each picture to grayscale and resized it to 6464 pixels. Figure 2 shows a sample from the database.

In the field of image processing, computer science and pattern matching are a rapidly developing discipline. ResNet is a crucial component of computer vision. Convolutional layers are the most significant layer for extracting features from images in ResNet [23]. To comprehend ResNet, think of it as a collection of residual blocks, each of which has a convolutional layer, batch normalisation, and the Relu activation function. ResNet also includes a skip layer that aids in the resolution of the vanishing gradient issue [24]. To create this model, we utilised Keras API with Tensorflow as a backend. We employed 50 layers in just this model, with the first layer being maxpool (Maxpool2D) as seen in picture 4. The output is then combined with the result of the first layer, a process known as multiple hops, and lastly the transfer functions Relu (figure 5) is applied to the output. Figure 6 depicts the model architecture. This was maintained until the last layer, when we utilised Flatten layer and linked thick layer with 24 classes, followed by Softmax function for posterior distribution.

A.Gap Analysis

Video and image processing are used in the current approach for sign language recognition. On the input side, a colour recording device is used. They can compute the relevant features at a frame rate of times per second using a CPU with an integrated attempt to classify. Following the capture, the image segmentation process starts. Photographs of trains are contrasted to photographs that have been segmented. Webcam image retrieval requires a lot of memory to preserve videos in real-time, thus our Gesture Recognition Sign may help you find what you're searching for. There will be fragmentation issues, sign recognition mistakes in real time, and a high degree of complexity. Quality and utilization are difficult to achieve in an inertial sensors system because tilts the figure in the xyz dimension but instead of bending it. This research advises utilising filters inside the sign automatic translation algorithm since the current technique is inaccurate owing to issues about skin tone detection. Although achieving there might be tough, sign language translation reliability can reach as high to 90%. The current system could not attain this degree of accuracy because detecting skin complexion in low-light environments took too long.

III. PROPSOED SYSTEM DESIGN

Initially, the information is divided from a video stream from the Webcam, as indicated in the above figure. To prevent background conflicts, the frames are discarded from the video with an area of Interest. The model is a custom CNN with multiple layers. The grayscale picture of the move separated from the video frame is then created. Because the model is trained using the characteristics of grayscale pictures, the input image from the camera is transformed to grayscale. The MNIST dataset is a pre-processed collection of RGB images that are transformed to grayscale. The transformed picture is then scaled to match the size of an image used to construct the system. After scaling and transformation, the picture is sent into the pre-trained proprietary CNN model. The CNN model's gesture prediction is derived, and it is then categorised using the category label. The gesture is labelled and presented as text.

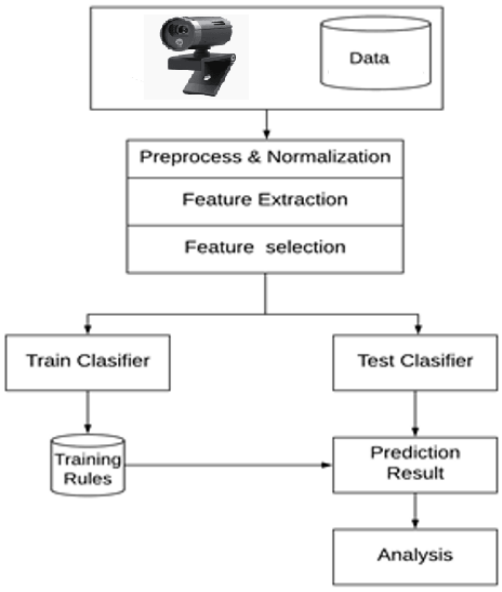

Figure 2 : Proposed system architecture

The above figure 2 demonstrates execution of proposed system with real time as well as synthetic dataset.Deep Learning includes the Convolutional Neural Network (CNN). Too far, CNN has shown to be a highly efficient and successful method of achieving handwriting recognition. Convolutional Neural Networks are neural networks that employ multiple layers of filters to extract information from images.

A. Convolutional Layer:

It's the foundation for constructing a CNN model. This layer conducted mathematical calculations on the picture that was used as input, as well as resizing the image into the M* M format. This layer's output describes the image's features, such as edge and corner mapping, also known as a feature map. The information was then added to the following layer.

B. Pooling Layer:

This is the layer that connects the convolutional and fully connected layers. This layer is used to decrease the network's parameters and computation. The maxpooling and average pooling methods are provided by this layer. The most frequent method is max pooling. The output of the preceding layer, the pooling layer, is sent to the fully connected layer. This layer is where the categorization process takes place.

C. Dense Layer:

This layer work for classification, it uses the training data as background knowledge and identify the current activity accordingly. These modules basically give the refection accuracy of proposed system.

In practice, input is given through a graphical user interface (GUI). Now, for the GUI, we've built a new file in which we've constructed an interactive window in which we can draw object on the canvas and identify them using buttons. The Tkinter package for Python was used to create the GUI. Tkinter is a standard Python GUI library. It allows you to build a GUI application quickly and easily. After giving the input, the model is loaded and stored, and predictions are made. h5 format. The supplied information is progressed further in order to resize in a certain format in order to get the real forecast. The resize picture is then passed on to the prediction model, where the provided input's features are extracted. The modelling then yields a prediction, which reflects the likelihood of the target variable based on the assessed importance of the input variables.Authors and Affiliations.

IV. ALGORITHM DESIGN

A. CNN Training:

Input: Training dataset collect from MNIST or real time sources TrainData[], CNN activation functions set[], Inputthreshold Th

Output: Extraction of various features from convolutional and pooling layer.

Step 1: Set input block of data TrainData [], selective activation function, iteration size as epoch,

Step 2: Features[] ← Extract_Features(TrainData [])

Step 3 : Optimized_Feature_set[] ← CNN,optimized(Features[])

Step 4 : Return Optimized_Feature_set[]

B. CNN Testing:

Input: Test Dataset which contains various test instances TestDBLits [], Train dataset which is build by training phase TrainDBLits[] , Threshold Th.

Output: HashMap <class_label, SimilarityWeight> all instances which weight violates the threshold score.

Step 1: For each read each test instances using below equation

$$\text{testFeature(m)}=\sum_{m=1}^{n}(.\ \text{featureSet[A[i]}\ldots\ldots\text{A[n]} \leftarrow \text{TestDBLits})$$

Step 2 :extract each feature as a hot vector or input neuron from $testFeature(m)$ using below equation.

$$\text{Extracted\_FeatureSetx[t}\ldots\ldots\text{n]} = \sum_{x=1}^{n}(t) \leftarrow \text{testFeature (m)}$$

Extracted_FeatureSetx[t] contains the feature vector of respective domain.

Step 3: For each read each train instances using below equation

$$\text{trainFeature(m)}=$$

$$\sum_{m=1}^{n}(.\ \text{featureSet[A[i]}\ldots\ldots\text{A[n]} \leftarrow \text{TrainDBList})$$

Step 4 :extract each feature as a hot vector or input neuron from $testFeature(m)$ using below equation.

$$\text{Extracted\_FeatureSety[t}\ldots\ldots\text{n]} = \sum_{x=1}^{n}(t) \leftarrow \text{testFeature (m)}$$

Extracted_FeatureSetx[t] contains the feature vector of respective domain.

Step 5 : Now map each test feature set to all respective training feature set

$$\text{weight=calcSim (FeatureSetx} \mid\mid \sum_{i=1}^{n} \text{FeatureSety[y])}$$

Step 6 : Return Weight

The different deep learning algorithms and solutions has utilized for to solve the problem of memory and execution cost. Initially, we define the time bound and space bound as outcome of process and different optimizers has used.

Gradient Descent has used with $\theta=\theta-\alpha\cdot\nabla J(\theta)$, Stochastic Gradient Descent utilized using $\theta=\theta-\alpha\cdot\nabla J(\theta;x(i);y(i))$, where {x(i) ,y(i)}. Mini-Batch Gradient Descent with $\theta=\theta-\alpha\cdot\nabla J(\theta; B(i))$, where {B(i)} and momentum $V(t)=\gamma V(t-1)+\alpha.\nabla J(\theta)$.The Nesterovic Accelerated Gradient $V(t)=\gamma V(t-1)+\alpha.\ \nabla J(\ \theta-\gamma V(t-1)\ )$ and then update the parameters using $\theta=\theta-V(t)$. Adagrad optimizer using $g_{t,i}=\nabla_\theta J(\theta_{t,i})$, adadelat using $E[g^2](t)=\gamma.E[g^2](t-1)+(1-\gamma).g^2(t)$, adam with

$$\theta_{t+1}=\theta_t-\frac{\eta}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t.$$

and evaluation has done in entire result section

V. RESULTS AND DISCUSSIONS

Intel i7 CPU 2.7 GHz has used with 16 GB Random Access Memory for execution. The RESENT (32,50, 101 and 152) version has used for experimental investigation of proposed systems including 5G network. The major factors has considered execution time (including data processing, data uploading and downloading etc.), memory consumption, network overhead and energy for valuation the efficiency of proposed systems.

Table 1. Comparison of various algorithms and their accuracy

Authors information	Methodology	Algorithms	Accuracy
Shape texture-based hand gesture [16]	Shape and localized features extraction technique has used for detection of sign language	SVM, KNN DTW	86.20%
KNN classification techniques [17]	KNN based image feature selection for classifcation	SIFT and HOG	95.70%
Real and static hand gesture [20]	OpenCV methodology has used for detection of sign language	Euclidian distance approach for template matching	62.10%
Wavelet descriptor and FMCC technique [21]	collaboration of wavelet descriptor as well as Mel Sec Frequency Cepstral Coefficients (MFCC) approach has used for extraction of features	KNN and SVM	93.20%
Proposed	CNN features has extracted	Deep CNN	95.90%

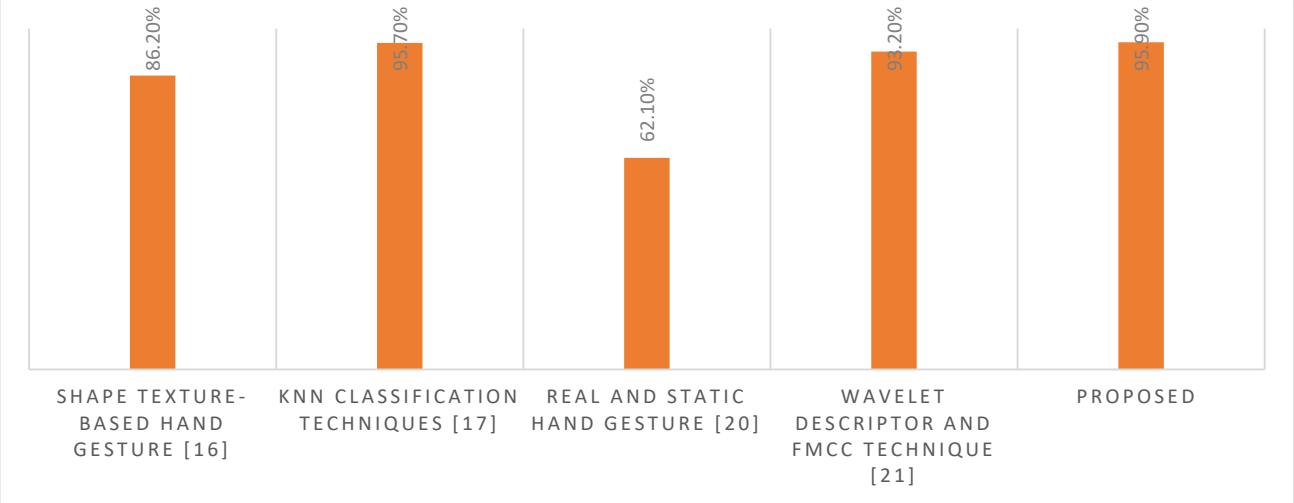

Figure 3: Analysis of proposed system using RESNET-100 with CNN various machine learning algorithms

The above results describe proposed CNN provides better detection accuracy over the traditional machine learning algorithms. The proposed CNN improve the accuracy almost 5-7% than other approaches.

VI. CONCLUSION

In continuation, in order to enhance mobile artificial intelligence, we expect that this framework can provide knowledge on the use and implementation of deep learning. These techniques will stimulate additional research and implementation of scenarios that will allow the network and services to be increasingly automated in the future. A deep learning framework that allows flexible allocation of resources in mobile deep learning systems is defined in this paper. Time consumption as well as memory usage, the system trades off system performance. To this end, through code block execution, we proposed a comprehensive mechanism that does not lead to any loss of accuracy and is friendly to the resource restriction environment. The experiments investigate the performance, good configurability and original deep learning process of the system by installing the application with Caffe and TensorFlow frameworks. For future work to detect moving objects in runtime direction using hybrid deep learning will the interesting task for enhancement of current research.

REFERENCES

[1] Wu Y, Huang TS (1999) Human hand modeling, analysis and animation in the context of HCI. In: Image processing, ICIP 99. Proceedings. 1999 international conference, IEEE, pp 6–10

[2] "Demand for Sign Language Interpreters Expected to Rise Nearly 50%", A Sharp, http://wqad.com/2015/10/21/demand-for-signlanguage-interpreters-expected-to-rise-nearly-50/, accessed 5/17/18

[3] Balakrishna, G. and P. S. Rajam. 2012. Recognition of Tamil Sign Language Alphabet using image processing to aid Deaf-Dumb People. International Conference on Communication Technology and System Design. 30: 861-868

[4] Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 1097-1105.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436-444.

[6] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent advances in convolutional neural networks, Pattern Recognition 77 (2018) 354-377.

[7] F. Sultana, A. Sufian, P. Dutta, Advancements in image classification using convolutional neural network, in: 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 122-129.

[8] Alom, M.Z., Sidike, P., Taha, T.M., Asari, V.K. ",Handwrittenbangla digit recognition using deep learning", May 2017.

[9] U Starner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans Pattern Anal Mach Intell 20:1371–1375

[10] R.Grobel K, Assan M (1997) Isolated sign language recognition using hidden Markov models. In: Systems, Man, and Cybernetics, 1997. Computational cybernetics and simulation. 1997 IEEE international conference, IEEE, pp 162–167

[11] N Sarawate , M Chan LEU , Cemil OZ¨ , "A real-time American Sign Language word recognition system based on neural networks and a probabilistic model ", Turkish Journal of Electrical Engineering & Computer Sciences,2013, pp 2107-2123, 2013.

[12] Vogler C, Metaxas D (1998) ASL recognition based on a coupling between HMMs and 3D motion analysis. In: computer vision, 1998. Sixth international conference, IEEE, pp 363–369

[13] N.Priyadharsini, N.Rajeswari, "Sign Language Recognition Using Convolutional Neural Networks ", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 6, pp :625-628, June 2017.

[14] Al-Jarrah, Ali MohammdShatnawi, Alaa Halawani, "Recognition of gestures in Arabic Sign Language using neural networks", Artificial Intelligence and Soft Computing, Spain, August 28-30, 2006.

[15] Deora, D. and N. Bajaj. 2012. Indian sign language recognition. 1st International Conference on Emerging Technology Trends in Electronics Communication and Networking (ET2ECN). : 1-5.

[16] Rekha J, Bhattacharya J, Majumder S (2011) Shape, texture and local movement hand gesture features for indian sign language recognition. In: 3rd international conference on trendz in information sciences and computing (TISC2011), IEEE, pp 30–35

[17] Gupta B, Shukla P, Mittal A (2016) K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. In: 2016 international conference on computer communication and informatics (ICCCI), IEEE, pp 1–5.

[18] Weiyang Liu, YandongWen, Zhiding Yu, Meng Yang," Large-Margin Softmax Loss for Convolutional Neural Networks ", Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

[19] Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[20] Pansare JR, Dhumal H, Babar S, Sonawale K, Sarode A (2013)Real time static hand gesture recognition system in complex background that uses number system of Indian sign language. Int J Adv Res ComputEngTechnol IJARCET 2:1086–1090

[21] Baranwal N, Nandi G (2017) An efficient gesture based humanoid learning using wavelet descriptor and MFCC techniques. Int J Mach Learn Cybern 8(4):1369–1388 https://github.com/imRishabhGupta/Indian-SignLanguage-Recognition

[22] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[23] Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 6(02), pp.107-116