Indian Sign Language Recognition using Fine-tuned Deep Transfer Learning Model

Chandra Mani Sharma[1], Kapil Tomar[1], Ram Krishn Mishra[2] and Vijayaraghavan M. Chariar[1]
[1]*Indian Institute of Technology, New Delhi, India*
[2]*BITS Pilani, Dubai Campus, Dubai, U.A.E.*

Keywords: Transfer Learning, Indian Sign Language, DCNN.

Abstract: The World Health Organization (WHO) estimates that over 5% of the world population suffers from hearing impairment. There are over 18 million hearing-impaired people living in India. In some cases, deafness comes from birth, hampering the speech learning capabilities of a child. Therefore, for this type of population, it is difficult to use spoken languages as a medium of communication. Sign languages come to the rescue of such people, providing a medium of expression and communication. However, it is difficult to decode sign language for other people who do not understand it. Computer vision and machine learning may play an important role in understanding what is said using sign language. The Indian sign language (ISL) is very popular in India and in many neighboring countries. It has millions of users. The paper presents a deep convolutional neural network (DCNN) model to recognize various symbols in ISL, belonging to 35 classes. These classes contain cropped images of hand gestures. Unlike other feature selection-based methods, DCNN has the advantage of automatic feature extraction during training. It is called end-to-end learning. A light weight transfer learning architecture makes the model train very fast, giving an accuracy of 100%. Further, a web-based system has been developed that can easily decode these symbols. Experimental results show that the model can classify Indian sign language symbols with accuracy and speed, ideal for real-time applications.

1 INTRODUCTION

Sign languages have existed since ancient times. Almost every country has its own sign language. There are 150 recognized sign languages (David et al, 2021). The total number of sign languages may be more than that. Sign language is a mechanism of expression by means of signs. Mostly, hand gestures and facial expressions are used for communication. Similar to any spoken language, sign language is also a form of natural language (Ghotkar, 2014). Sign language has its own set of alphabets and vocabulary. In deaf cultures, sign languages have played a significant role in connecting people based on ethnicity and cultural similarities.

Sign language relies heavily on hand gestures, body movements, and facial expressions.There could be thousands of words, expressions, and their associated meanings in ISL. However, numerals and alphabets are the basic building blocks for spelling out complex words using ISL. Fig. 1 shows some of the symbols in ISL (Prathum Arikeri, 2021).

2 LITERATURE SURVEY

A language is a medium used for communication between two people by sharing their information with each other (Gupta & Kumar, 2021). Similarly, sign language is used for communication by mentally impaired people (Rao & Kishore, 2018). Languages such as Hindi or English use a structured way for verbal and written communication, while sign languages use facial expressions and signs shaped by hand movements for communication (Rao & Kishore, 2018). Various researchers targeted the impaired people of the country and tried to develop an automation system using various methods and technologies that can use one hand and two hand symbols for communication (Sharma et al., 2020). The Sign language recognition (SLR) system measures and predicts human actions (Kishore et al., 2018).This automated system can be used by differently-abled people for communication using Indian Sign Language (ISL).

Various types of sign languages are used across the world, such as American Sign Language (ASL),

Sharma, C., Tomar, K., Mishra, R. and Chariar, V.
Indian Sign Language Recognition using Fine-tuned Deep Transfer Learning Model.
DOI: 10.5220/0010790300003167
In *Proceedings of the 1st International Conference on Innovation in Computer and Information Science (ICICIS 2021)*, pages 63-68
ISBN: 978-989-758-577-7

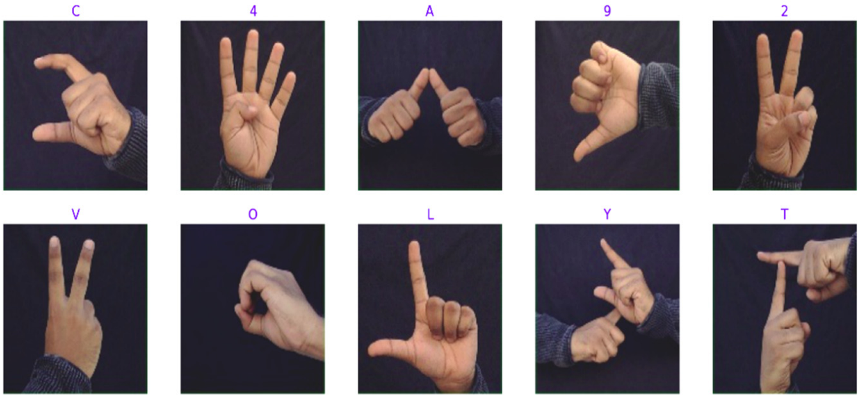

Figure 1: Some Representative Symbols in Indian Sign Language (ISL).

British Sign Language (BSL), German Sign Language (GSL), and Indian Sign Language (ISL), etc. All these languages are different from each other in terms of symbols, syntax and morphology (Hore et al., 2017). Even in India, due to being a diverse country, many languages exist. So, recognition of sign language has been a major challenge and an area of research in the last few decades (Badhe & Kulkarni, 2020). Researchers have published various studies on sign languages other than India. (Beena et al 2019) used the Convolutional Neural Network (CNN) and Support Vector Machines (SVM) on American Sign Language (ASL). They extracted important features using a variety of techniques, including HOG, LBP, and 3D-voxel. Similarly, (Quesada et al. 2017) worked on American Sign Language (ASL) to recognize the finger-spelling alphabets automatically. They also used Support Vector Machines (SVM) to classify the gestures. And (Raheja et al., 2016) used Support Vector Machine (SVM) as a classifier to classify the input gesture from the sign database. (Ghotkar, 2014) used computer vision and natural language processing (NLP) to get the meaning of hand gestures.

Sign languages rely on gestures and shapes made with one or both hands.of automated sign language systems can be divided into three parts, such as training, testing and the recognition phase (Dixit & Jalal, 2013).

The efficiency of a sign language recognition (SLR) system depends on how fast and accurately it tracks the features and orientations of the hand (Rao & Kishore, 2018). There are many algorithms for machine learning and for feature extraction. The Convolutional Neural Network (CCN) is one of the best algorithms used in feature extraction in the area of deep learning (Islam et al., 2018). Sign languages are the most common application of the vision-based approach. Earlier, many researchers used the Hidden Markov Model (HMM) in sign languages to extract handcrafted features. But due to the recent success of deep learning techniques in the area of object identification, image classifications, natural language processing and human activity recognition, recently, many researchers have used deep convolutional neural networks (DCNN) in sign languages for feature extraction (Al-Hammadi et al., 2020) (Oyedotun & Khashman, 2017).

The Sign Language Recognition (SLR) system proposed by (Rao & Kishore, 2018) gave an average 85.58% performance for Word Matching Score (WMS) using a minimum distance classifier (MDC) and performance increased up to 90% for Artificial Neural Network (ANN) after making some little variations. It is possible to obtain more accurate classifiers using neural networks (Rao & Kishore, 2018). (Gupta & Kumar, 2021) found in their study that the highest accuracy for ISL recognition can be achieved using Label Powerset (LP) based SLR with the lowest error of 2.73%. (Sharma et al., 2020) developed the dataset manually and used pre-trained VGG16 for training. More than 150,000 images of all 26 alphabets were included in the dataset. To keep the data consistent, the same background was used for each image.94.52% accuracy was obtained using the Deep Convolutional Neural Network (DCNN) for ISL recognition (Sharma et al., 2020). Tracking of hands, shapes made by fingers, head movement, and recognition of patterns from the database are the limitations of the Automated Sign Language Recognition (ASLR) system (Kishore et al., 2018). The accuracy of the ASLR systems mostly depends on feature extraction (Tyagi et al., 2021). Enhancing the feature extraction technique (Rajam & Balakrishnan, 2011) achieved 98.125% in their proposed model. Machine learning is widely used in various applications to extract features from a dataset (Oyedotun & Khashman, 2017).

Table 1: Approaches for Indian Sign Language (ISL).

Paper Ref	Algorithm or Methodology	Dataset	Accuracy	Limitations
(Rao & Kishore, 2018)	Artificial Neural Network (ANN)	Dataset of 1313 frames.	90.58 %	Number of hidden layers and computational time
(Kishore et al., 2018)	Kernel Matching Algorithm	Mocap dataset	98.9 %	Missing Nodes and inaccurate features
(Gupta & Kumar, 2021)	LP-based SLR	20,000 samples collected with multiple sensors	97.27 %	Categorization of the data.
(Sharma et al., 2020)	Hierarchical Network	Manually prepared dataset (150,000 images of all 26 categories)	98.52% - one-hand gestures 97% - for both hand gestures	Classification of sign
(Badhe & Kulkarni, 2020)	Artificial Neural Network	Created own dataset	Training accuracy - 98%. Validation accuracy - 63%	Small size of used dataset
(Dixit & Jalal, 2013)	Multi-class Support Vector Machine (MSVM)	720 images, data set created	96 %	
(Rajam & Balakrishnan, 2011)	Image Processing Technique	Created own dataset of 320 images for 32 signs, 10 images for each	98.125%	No standard Dataset for south Indian Language
(Yuan et al., 2019)	Deep Learning	Created Chinese Sign Language Dataset (CSLD) by discussing with expert	Not Mentioned	Accuracy of Data
(Aly & Aly, 2020)	Multiple Deep Learning Architectures (deep Bi-directional Long Short-Term Memory (BiLSTM) recurrent neural network)	Arabic Sign Language database for 23 words	89.59%	Segmentation of data

3 MATERIALS AND METHODS USED FOR ISL RECOGNITION

3.1 Software Tools and Setup

For deep learning model creation, compilation and training, the Python deep learning library Keras (with Tensorflow as a backend) was used. The training was performed in the cloud on an Nvidia K80 graphical processing unit. For creating the web application, the Python web framework Flask was used. The system was set up on the Heroku cloud platform.

3.2 Dataset

The Kaggle Dataset (Prathum Arikeri, 2021) was used to train the deep learning model for ISL recognition.This dataset is available openly under creative commons licensing and can be accessed easily. The dataset contains 42,745 RGB images belonging to 35 sign classes. There are 9 numeral signs (1-9) and 26 alphabet signs (A-Z). The number of samples in the *'C'*, *'O'*, and *'I'* classes is 1,447, 1,429, and 1,379, respectively. Each of the remaining classes contains 1200 images.The samples have been resized to a common dimension of 128x128.

3.3 Transfer Learning Model

In order to save time & computational sources, the transfer learning approach is very popular. It solves pattern recognition problems in various domains. There are a variety of pre-trained deep learning models that can be used after solving a given problem. However, it is unlikely that a model can solve all problems all the time. Fine-tuning is the process of finding out the optimal values for a given set of variables (hyperparameters). We tested various custom models on the ISL dataset, and found that one fine-tuned variant of the MobileNetV2 model is ideal for our purpose. It is a deep learning model, pre-trained with a huge ImageNet dataset.

The first half of the layers of the model were frozen (*N1*), and the rest of the layers (*N2*) were unfrozen to update their weights during training. One custom output layer was added with 35 neurons, representing each of the categories of ISL symbols in the dataset. Fig. 2 shows the schematic diagram of the model. In order to avoid overfitting, data augmentation was used. The images were randomly rotated, shifted, sheared, and cropped.

Initially, the number of epochs was set at 10. A call back function was used to update the hyperparameters listed in Table 1. The initial learning rate was set at 0.001. Later on, the factor to reduce it was set to 0.25. A rule of adjusting monitor accuracy when train accuracy is less than the threshold, otherwise monitoring the validation loss, was applied.

Table 2: Hyperparameters for Monitoring Training Process

#epochs	#patience	#stop_patience	threshold	factor	dwell	freeze
10	1	3	0.85	0.25	True	False

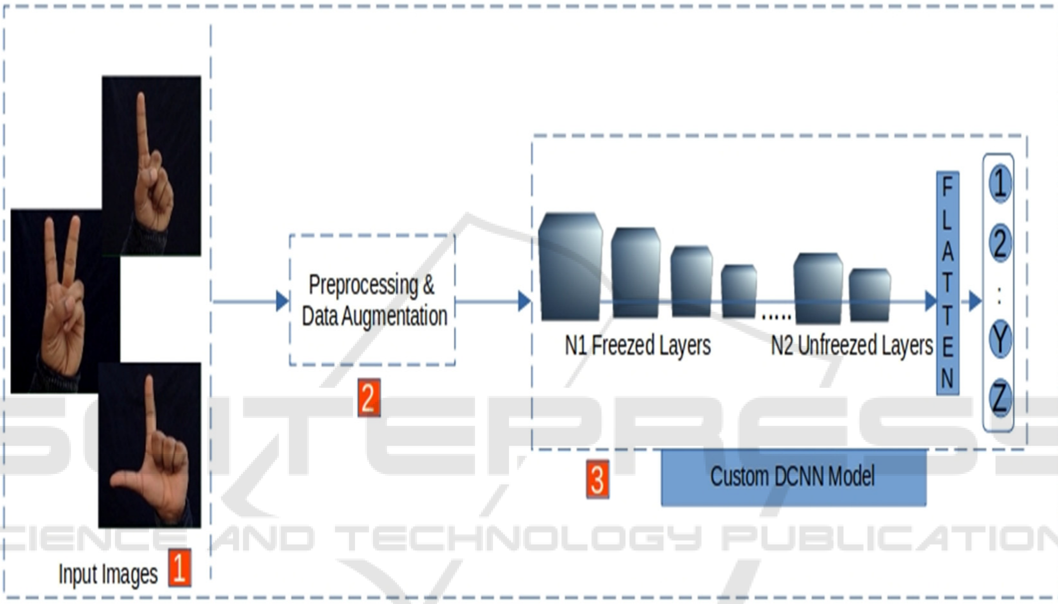

Figure 2: ISL Classifier Model Training.

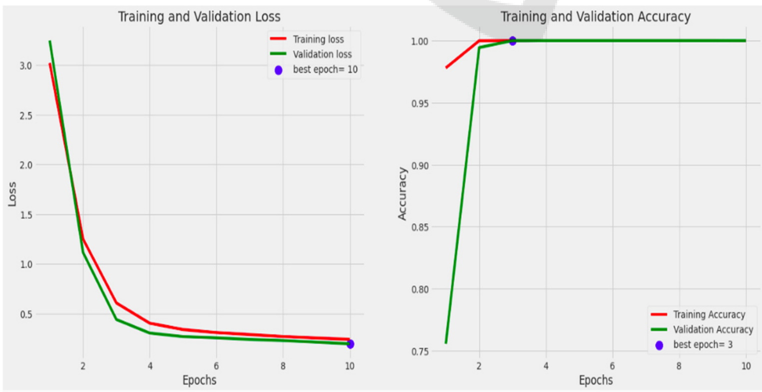

Figure 3: Loss and Accuracy Curve of the DCNN Training & Validation Phases.

4 RESULTS & DISCUSSION

Fig. 3 shows the loss and accuracy curves of the trained model on the ISL dataset. The left part of the figure shows the training and validation losses, drawn against each epoch. The best epoch, from the loss minimization point of view, happens to be the last one. Further, it is evident from the accuracy graphs that the model starts with an approximate accuracy score of 0.75, and improves to reach the perfect score of 1.00. The model converges from epoch number 3, as evident from the given figure. Experimental results show that the proposed technique can recognize the ISL symbols with promising accuracy. In the future, the work could be extended to recognize a wide variety of words with a few more applications.

Further, the comparison with other classification algorithms is shown in Fig. 4. The other classification algorithms, used for comparison, are the Neural Network (NN), Genetic Algorithm (GA), Evolutionary Algorithm (EA), and Particle Swarm Algorithm (PSA) to recognize Indian Sign Language (ISL) gestures. A k-fold cross validation was performed to calculate the accuracy of a total of 35 gestures and 30% of data of each gesture was used to analyze the performance. The data set is divided into two parts, such as training and testing. 70 % of the data set is used for training and the remaining data was used for testing the Neural Network. A comparison of accuracy with respect to multiple parameters is shown in Fig. 4 with the help of a bar graph.

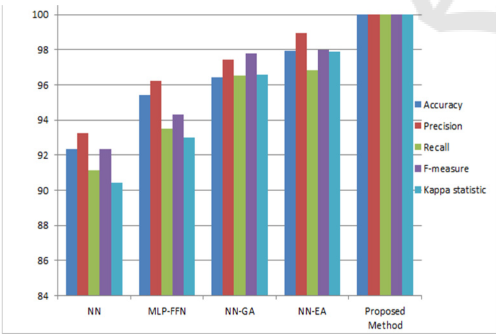

Figure 4: Accuracy of the suggested methodology.

5 CONCLUSION

Modern technological advancements can assist the hearing and speech impaired population to effectively communicate, and connect with other people. Automated sign language recognition is one such area that has attracted researchers from multiple fields of study. In this work, a computer vision based deep learning approach has been used to recognize ISL primitive symbols from 35 different classes. The model can achieve 100% accuracy on unseen test data and has similarly good loss & accuracy during training. It can be used as a useful tool to enable hearing or speech impaired people to communicate with the rest of the world.

REFERENCES

Eberhard, David M.; Simons, Gary F.; Fennig, Charles D., eds. (2021). Sign language, Ethnologue: Languages of the World (24th ed.), SIL International, retrieved 2021-05-30.

Hore, S., Chatterjee, S., Santhi, V., Dey, N., Ashour, A. S., Balas, V. E., & Shi, F. (2017). Indian sign language recognition using optimized neural networks. In Information technology and intelligent transportation systems (pp. 553-563). Springer, Cham.

Prathum Arikeri. (2021), “Indian Sign Language Character Level Dataset. Kaggle Dataset. Retrieved 2021-06-05.

Kishore, P. V. V., Kumar, D. A., Sastry, A. C. S., & Kumar, E. K. (2018). Motionlets matching with adaptive kernels for 3-d indian sign language recognition. IEEE Sensors Journal, 18(8), 3327-3337.

Rao, G. A., & Kishore, P. V. V. (2018). Selfie video based continuous Indian sign language recognition system. Ain Shams Engineering Journal, 9(4), 1929-1939.

Gupta, R., & Kumar, A. (2021). Indian sign language recognition using wearable sensors and multi-label classification. Computers & Electrical Engineering, 90, 106898.

Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya, D. (2020). Benchmarking deep neural network approaches for Indian Sign Language recognition. Neural Computing and Applications, 1-12.

Sridhar, A., Ganesan, R. G., Kumar, P., & Khapra, M. (2020, October). INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 1366-1375).

Tyagi, A., & Bansal, S. (2021). Feature extraction technique for vision-based indian sign language recognition system: A review. Computational Methods and Data Engineering, 39-53.

Badhe, P. C., & Kulkarni, V. (2020, July). Artificial Neural Network based Indian Sign Language Recognition using hand crafted features. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

Raheja, J. L., Mishra, A., & Chaudhary, A. (2016). Indian sign language recognition using SVM. Pattern Recognition and Image Analysis, 26(2), 434-441.

Dixit, K., & Jalal, A. S. (2013, February). Automatic Indian sign language recognition system. In 2013 3rd IEEE

International Advance Computing Conference (IACC) (pp. 883-887). IEEE.
Rajam, P. S., & Balakrishnan, G. (2011, September). Real time Indian sign language recognition system to aid deaf-dumb people. In 2011 IEEE 13th international conference on communication technology (pp. 737-742). IEEE.
Ghotkar, A. S., & Kharate, G. K. (2017). Study of vision based hand gesture recognition using Indian sign language. International journal on smart sensing and intelligent systems, 7(1).
Ariesta, M. C., Wiryana, F., & Kusuma, G. P. (2018). A Survey of Hand Gesture Recognition Methods in Sign Language Recognition. Pertanika Journal of Science & Technology, 26(4).
Oyedotun, O. K., & Khashman, A. (2017). Deep learning in vision-based static hand gesture recognition. Neural Computing and Applications, 28(12), 3941-3951.
Islam, M. R., Mitu, U. K., Bhuiyan, R. A., & Shin, J. (2018, September). Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 115-119). IEEE.
Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., ... & Mekhtiche, M. A. (2020). Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation. IEEE Access, 8, 192527-192542.
Aly, S., & Aly, W. (2020). DeepArSLR: A novel signer-independent deep learning framework for
isolated arabic sign language gestures recognition. IEEE Access, 8, 83199-83212.
Yuan, T., Sah, S., Ananthanarayana, T., Zhang, C., Bhat, A., Gandhi, S., & Ptucha, R. (2019, May). Large scale sign language interpretation. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5). IEEE.
Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 1-40.
Beena V, Namboodiri A, Thottungal R (2019) Hybrid approaches of convolutional network and support vector machine for American sign language prediction. Multimed Tools Appl. https://doi.org/10.1007/s11042-019-7723-0
Quesada L, Lo´pez G, Guerrero L (2017) Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments. J Ambient Intell Humaniz Comput 8(4):625–635. https://doi.org/10.1007/s12652-017-0475-7
DixitM, Upadhyay N, Silakari S (2015) An exhaustive survey on nature inspired optimization algorithms. Int J Softw Eng Appl 9(4):91–104