International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03236
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Indian Sign Language Recognition: A Review of CNN-Based Approaches and Challenges

[1] Prajakta Dawange, [2] Omkar Pattnaik, [3]Neha Ranjan, [4]Shalini Mhaske, [5]Gitanjali Khatik

[2]Associate Professor, School of Computer Science and Engineering, Sandip University, Nashik, Maharashtra-422213, India.

Mail ID: omkar.pattnaik@sandipuniversity.edu.in

[1,3,4,5] Student, School of Computer Science and Engineering, Sandip University, Nashik, Maharashtra-422213, India.

prajdawange2001@gmail.com, ranjanneha536@gmail.com, shalinimhaske26199@gmail.com, gitanjalikhatik@gmail.com,

**Corresponding authors: prajdawange2001@gmail.com, omkar.pattnaik@sandipuniversity.edu.in*

Abstract- Communicating with the person having hearing failure is ceaselessly a major challenge. The work shown in paper is an exertion(extension) towards looking at the challenges in classification of characters in Indian Sign Language (ISL). Sign Dialect is not adequate for communication of people with hearing capacity or people with talk inadequacy. The signals made by the people with inadequacy gets mixed or disarranged for some person who has never learnt this lingo. Communication needs to be in both ways. In this paper, we display a Sign Dialect affirmation utilizing Indian Sign Lingo. The client must be able to capture pictures of hand movements utilizing a web camera in this examination, and the system must anticipate and show up the title of the captured picture. The captured picture encounters the course of action of planning steps which consolidate distinctive Computer vision strategies such as the change to gray-scale, extending and cover operation. Convolutional Neural arrange (CNN) is utilized to plan our show and recognize the pictures. Our appear has finished precision around 95%.

Keyword - Convolutional Neural Network (CNN), Hand Gesture, Hearing-impaired people, Sign Language, Sign Language Recognition (SLR).

1. INTRODUCTION

One of the foremost vital components for survival in society is communication. Hearing impeded individuals' conversation to each other utilizing sign dialect, but non-hearing-impaired individuals have difficulty understanding them. Although there is a part of investigate on the acknowledgment of American Sign Dialect, Indian Sign Dialect is exceptionally diverse from American Sign Dialect. ISL employments two hands (20 out of 26 hands) to communicate, whereas ASL employments one hand to communicate. Since both hands cover when utilizing both hands, the highlights are regularly not obvious. Too, need of documentation and contrasts in sign dialect on the location moreover restrain ISL gesture research. This paper points to be the primary step towards utilizing Indian Sign Dialect to bridge the communication between hearing and hearing-impaired individuals. This extend extends into a single word and a common expression, not as it were encouraging Hearing impeded people's communication with the exterior world but moreover offer assistance them create the skills to get it way better. The point of this paper is to memorize the Indian Sign Dialect letter set utilizing hand motions. Understanding motions and sign dialect could be a well-studied topic in American Sign Dialect but has gotten small consideration in Indian Sign Dialect. We need to illuminate this problem but rather than utilizing progressed innovation like gloves or Kinect, we need to memorize approximately signals from photographs (available through webcam) and after that utilize computer vision and machine learning to extricate interesting highlights and classify them.

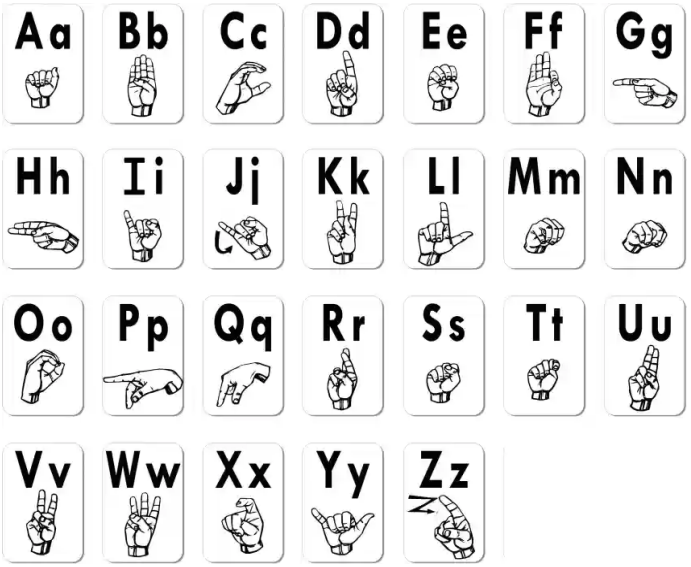

Fig. 1: Indian Sign Language Alphabets.

Understand the meaning of Hearing impaired and dumb signs and translate them into understandable language (text)

2. LITERATURE SURVEY

Misfortune of hearing can cause individuals to ended up separated and forlorn, having a huge impact on both their social and working life. Hearing individuals utilized sign dialect for communication

among themselves. An translator is required when a ordinary individual needs to communicate with a hearing impeded. Numerous individuals never attempt to memorize the sign dialect for collaboration with hard of hearing individuals. It gets to be troublesome for finding a well experienced and taught interpreter for sign dialect each time and all over. S. Reshna has presented that the Robotization of Sign Language Recognition framework may be a step forward for making a difference the hearing disabled community and has been investigated for numerous years.[1][5] Rajat Soni has proposed communication through motions affirmation system contains two guideline stages, to be particular data securing, planning, and testing of the final CNN classifier. The going with fig. delineates the common system stream chart that addresses the working show of the system. The taking after stage is planning and testing the CNN classifier: The proposed demonstrate is ready with the NVIDIA GeForce GTX 1650 Graphical Preparing Unit (GPU), 4 GB of Slam, 16 GB of sporadic get to memory (Slam) and 1000 GB of solid-state drive (SSD). Around 30 diverse models were created with diverse combination of convolutional layers, pooling layers, smooth, dropout (for dodging overfitting of demonstrate) and thick layers which were at that point tuned for getting best hyperparameter values utilizing keras tuner.[2] Ankita Wadhawan and her group has work on Information Securing and Data preprocessing. In Information Securing three-channel picture outlines (RGB) are recovered from the camera, and after that these pictures are passed to the picture preprocessing module. The dataset comprises of the collection of the RGB pictures for diverse inactive signs. The dataset comprises 35,000 pictures which incorporate 350 pictures for each of the inactive signs. There are 100 unmistakable sign classes that incorporate 23 letter sets of English, 10 digits and 67 commonly utilized words (e.g., bowl, water, stand, hand, fever, etc.). The dataset comprises inactive sign pictures with different sizes, colors, and taken beneath distinctive natural conditions to help within the way better generalization of the classifier. In Information preprocessing the application of diverse morphological operations that are utilized to expel commotion from the information. In this stage, the sign pictures are preprocessed utilizing two strategies that are picture resizing and normalization. In picture resizing, the picture is resized to 128 9 128. These pictures are at that point normalized to alter the extend of pixel intensity values which comes about in cruel 0 and variance.[3] The essential Goals of investigate was to produce a plentiful sum of dataset for Indian Sign Dialect and to Plan the Finest show for CNN to the preprocessed picture and accomplish the most extreme conceivable accuracy.[4] The point of Rachana Patil was to utilized the corresponding gesture to perceive letter sets in Indian Sign Dialect. The distinguishing proof of motions and sign dialects was a well examined subject in American Sign Dialect, but it has gotten small consideration in Indian Sign Dialect. The most moto of inquire about was to fathom this issue, but rather than utilizing high-end advances like gloves or the Kinect, they need to perceive motions from photos (which can be accessed from a webcam), and after that utilize computer vision and machine learning methods to extricate particular highlights and classify them.[6] Chandra Mani Sharma and group said that the Cutting edge mechanical headways can help the hearing and discourse impeded populace to viably communicate, and interface with other individuals. Mechanized sign dialect acknowledgment is one such range that has attracted researchers from numerous areas of ponder. In this work, a computer vision based profound learning approach has been utilized to recognize ISL primitive images from 35 diverse classes. The demonstrate can accomplish 100% exactness on inconspicuous test information and has additionally great misfortune & exactness amid preparing. It can be utilized as a valuable instrument to empower hearing or discourse disabled individuals to communicate with the rest of the world.[8] The American Sign Language (ASL), British Sign Dialect (BSL), Indian Sign Dialect (ISL), and others are all distinctive sign dialects. Comparable to how talked dialects have a lexicon of words, sign dialects as well have a lexicon of signs. The linguistic use of sign dialects changes from nation to nation and isn't standardized or all inclusive. A manual sign dialect mediator isn't continuously a great thought and as often as possible barges in on the subject's right to security. This issue can be settled by utilizing a robotized sign dialect interpreter that can decipher sign language into talked or composed language.[9] Pias Paul presented different CNN-based models to classify 24 characters from the ASL FingerSpelling Dataset. They displayed models which were custom made for this issue, as well as models which use exchange learning. One of our custom models accomplished a test precision of 86.52%, which is way better than the current best distributed result.[10]

3. PROPOSED METHODOLOGY

Experiments are performed on the continuous ISL dataset. It is created with different background and illumination conditions with variations in surrounding, clothing, and gesture movement. Features that can identify the sign are extracted from the hand image and classified to recognize the sign using CNN [1][2][14]

3.1 Image Acquisition:

The color images are retrieved from the camera using data acquisition software developed by us and these images are then passed on to the image pre-processing module. The dataset comprises the assortment of RGB pictures for different static characters. The dataset contains 74,188 pictures of the static characters. There are 33 distinctive person classes that incorporate 23 English letter sets and 0 - 9 digits. The dataset comprises static sign pictures of various sizes, colors and recorded under various ecological conditions to help the better speculation of the classifier.[2][3]

3.2 Preprocessing

The data preprocessing is the application of different morphological operations that are used to remove noise from the data.[3] In this phase, the sign images are preprocessed using three methods that are dilation, and Gaussian smoothing. When a color image is converted to grayscale, the size of the image is reduced. One way to reduce data processing is to convert the image to grayscale. The preprocessing step is as follow:[4][10]

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03236
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

3.2.1 Morphological Change (Morphological Change):
similar-sized yield image.

It compares the comparing pixel within the input image with its neighbors to decide the esteem of each pixel within the yield image. There are two different sorts of morphological changes Disintegration and Expansion.[6]

Figure 2: Original Hand Gesture Recognition

1. Dilation: The greatest esteem of all pixels in the neighborhood is the esteem of the yield pixel. A pixel in a double picture is set to 1 on the off chance that all of its neighbors have the esteem 1 Morphological enlargement increments the perceivability of relics and fills in little gaps.[6]

2. Erosion: The o/p pixel's esteem is the minimum of all pixels within the neighborhood. A pixel in a binary picture is set to in case all of its neighbors have the value 0, small relics are dissolved absent by morphological disintegration, taking off behind significant objects[6]

3.2.2 Blurring:

Including a low-pass channel to a picture is a case of blurring. The word "low-pass channel" alludes to killing noise from a picture whereas clearing out the rest of the picture intaglio in computer vision. An obscure operation may be a straightforward operation that must be completed sometime recently for other tasks such as edge detection.[6]

3.2.3 Thresholding:

Thresholding could be a form of picture division in which the pixels of an image are changed to make it less demanding to decipher the picture. Thresholding is the method of converting a color or grayscale picture into a parallel picture, which is basically dark and white. We commonly utilize thresholding to choose regions of interest in a picture whereas ignoring the segments we are not concerned with.[6]. Thresholding is a type of picture segmentation where the pixels are altered to make the image simpler to understand. Thresholding is a technique that converts a color or grayscale image into a binary image containing only black and white pixels. It is most frequently used to choose specific regions of an image while disregarding other regions. Suppose a pixel's intensity in the input image exceeds the threshold. In that case, the corresponding output pixel is labeled as white (foreground), and if it is equal to or less than the threshold, it is labeled as black (background)[4].

3.2.4 Recognition:

Sign language recognition (SLR) is the task of recognizing sign language glosses from video streams. It is a very important research area since it can bridge the communication gap between hearing and Deaf people, facilitating the social inclusion of hearing-impaired people. Moreover, sign language recognition can be classified into isolated and continuous based on whether the video streams contain an isolated gloss or a gloss sequence that corresponds to a sentence.[11] We'll utilize classifiers in this case. Classifiers are the strategies or calculations that are utilized to decipher the signals. Well known classifiers that distinguish or get it sign dialect incorporate the Covered up Markov Demonstrate (Gee), Closest Neighbor classifiers, Bolster Vector Machine (SVM), Manufactured Neural Arrange (ANN), and Guideline Component Examination (PCA), among others. In any case, in this venture, the classifier will be CNN. Since of its tall accuracy, CNNs are utilized for picture classification and acknowledgment. The CNN employments a progressive demonstrate that builds a arrange, comparative to a pipe, and after that yields a fully-connected layer in which all neurons are associated to each other and the yield is handled[6]

3.2.5 Text Output:

Understanding human conduct and distinguishing various stances and body developments, as well as deciphering them into content.

3.3 Segmentation:

The process of separating an object or symbol from the content of the captured image is called segmentation. Scene extraction, skin tone detection and edge recognition are used in the segmentation process. In order to recognize gestures, hand movements and positions must be detected and segmented.[6]. An image segmentation technique divides a digital image into several groupings, or "image segments," each containing a variety of the image's elements. Focusing on a specific item inside an image is done by image segmentation, which is necessary and helpful for our processing. Image segmentation simplifies image processing and analysis while also bringing down the complexity of the original image. In segmenting a picture, each pixel is given a name, and a section of the image

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03236
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

is allocated to a category with a specific label. Thus, a specific area of the picture with the same label is discovered instead of the entire image[4].

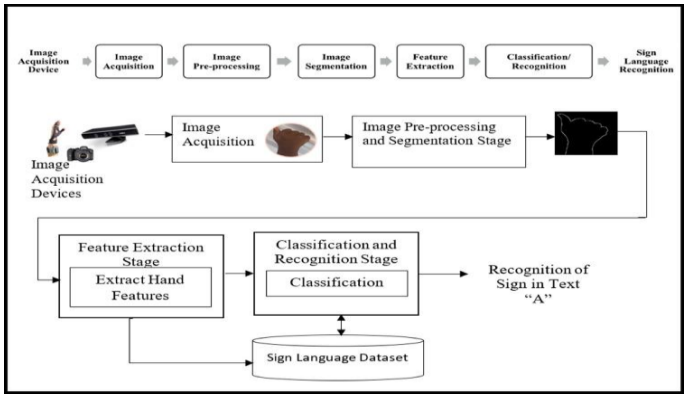

Figure 3: Proposed Methodology

3.4 Features Extraction using CNN:

Predefined features such as shapes, contours, geometric features (position, angle, distance etc.), color features, histograms etc. are first extracted from the image and then used for the classification of symbols or information. Feature extraction is a step in the dimensionality reduction process used to classify and organize large amounts of raw data. Reduce less, manage more groups and this will make the job easier. These large files have many variables and these are the most important ones. Many calculators are required to manage these changes. Therefore, function extraction helps to extract the best features from large datasets by selecting variables and assigning them to functions, reducing file size. This feature is easy to use and also describes the actual data collection process accurately and specifically.[6][10]. The Convolutional Neural Network (CNN) is a type of neural network that is commonly used for image classification and recognition tasks. In the proposed method for sign language detection, the CNN is responsible for feature extraction from the input image sequence. The CNN is composed of multiple convolutional layers followed by pooling layers. The convolutional layers apply a set of filters to the input image, producing a feature map that highlights the presence of certain features. The filters in the convolutional layers are learned during training, and each filter is responsible for detecting a specific feature in the input image.[12]

4. PROPOSED ALGORITHM

4.1 Creating the sign language recognition dataset:

Any outline that identifies a hand inside the ROI (locale of intrigued) produced can be exchanged to a registry that contains a combine of catalogs, prepare and take a look at, each containing 10 organizers containing pictures captured buse the deliver motion knowledge.py perform. Presently, to make the dataset, we have a propensity to use OpenCV to encourage the live cam nourish and make a ROI, that is exclusively the parcel of the outline wherever we have a propensity to need to discover the hand for the motions. For differentiating between the foundation we have a tendency to calculate the gathered weighted avg for the foundation at that point cipher this from the outlines that contain a few object ahead of the foundation which can be distinguished as foreground. This can be fulfilled by computing the collected weight for particular outlines and the context's gathered average. After we've the accumulated normal for the background, we tend to subtract it from each outline that we studied after sixty outlines to seek out any question that covers the foundation [6]

4.2 Calculate threshold value:

We presently degree the limit esteem for each outline and assess the forms utilizing cv. Discover Forms. The max contours i.e. object's most peripheral forms are returned utilizing the work portion. We may choose whether there's a few foreground objects distinguished within the ROI using the forms, in other words, whether there's a hand in the ROI. When show detects contours, it begins sparing the picture of the ROI within the prepare and test sets for the letter or number we're seeking out for (or a hand is show in the ROI). The dataset for is produced within the going before illustration, and the ROI's thresholded picture is shown in the taking after the window. Within the prepare dataset, we spare 600 images for each number to be recognized, and within the test dataset, we produce 80 pictures for each number.

4.3 CNN Layer:

The convolutional layer, the central component of a CNN, is where most computations take put. The primary convolutional layer may be taken after by a consequent convolutional layer. A bit or channel interior this layer moves over the image as responsive areas amid the convolution handle to decide whether a highlight is present. The bit navigates the complete picture over a number of cycles. A dab item between the input pixels and the channel is calculated at the conclusion of each cycle. A highlight outline or convolved include is the result of the specks being associated in a certain design. In this layer, the picture is eventually changed into numerical values that the CNN can get it and extricate germane designs from.[9][6]

Pooling Layer: The pooling layer additionally to the convolutional layer clears a bit or channel over the input picture. Opposite to the convolutional layer, the pooling layer has less input parameters but moreover causes a few data to be misplaced. Emphatically, this layer rearranges the CNN and increments its effectiveness.[9]

Fully Associated Layer: Based on the features extricated within the going before layers, picture categorization within the CNN takes put within the FC layer. Completely associated in this setting implies that each actuation unit or hub of the consequent layer is associated to each input or hub from the going before layer. The CNN does not have all of its layers completely associated since that would make an too much thick organize. It would fetched a part to compute, increment misfortunes, and have an affect on yield quality.[9]

4.4 Training CNN:

In the training of the model will be done after the preprocessing of the dataset. we will be training the model so that the image we will be selecting is trained and the parameters of image will be extracted from that and those calculations of the dataset will be shown at the time of training and that code will be executed with CNN. Then after the image will be trained under different parameters of the image that parameters. As a part of training the model, first we need to build the Convolution Neural Network (CNN). And the above code tells us how it is done. First, we need to initialize the CNN and build layers of Convolution layers and Pooling layer, here Max Pooling is used. After building Convolution layers, fully Connected layer will be added to CNN and compile CNN. After Building we need to use this CNN and prepare the train/test data, the following code will show how training of the model will happen.[5]

Layer(type)	Output Shape	Param#
conv2d(Conv2D)	(None, 62, 62, 32)	896
max_pooling2d(MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 31, 31, 64)	18496
max_pooling2d_1(MaxPooling2D)	(None, 15, 15, 64)	0
conv2d_2(Conv2D)	(None, 13, 13, 128)	73856
max_pooling2d_2(MaxPooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 64)	294976
dense_1 (Dense)	(None, 128)	8320
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 128)	16512
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 12	1548

Table 1: CNN Layer

4.5 predicting the gesture:

We produce a bounding box for recognizing the ROI and measuring the total normal, fair as we did when making the dataset. This can be worn out arrange to recognize a closer view substance. Presently we hunt for the greatest con- visit, and in case one is found, it implies a hand has been recognized, so the ROI's threshold is utilized as a test picture. Utilizing Keras models stack demonstrate, we stack the already spared show and after that bolster the edge picture of the ROI containing the hand to the show for forecast as an input.

5. CONCLUSION AND FUTURE SCOPE

The Sign language Acknowledgement (SLR) framework could be a strategy for recognizing a collection of shaped signs and interpreting them into content or discourse with the fitting setting. The significance of motion acknowledgment can be seen within the development of compelling human-machine intelligence. We endeavored to construct a show employing a Convolutional Neural Arrange to this extent. This comes about in a approval exactness of approximately 95%. The project emphasizes on making a framework that identifies Indian Sign Language in genuine time and shows the recognized words/letter to the conclusion client. It too changes over recognized words to content. This framework accommodates the difficulty of hearing individuals that need to communicate with orders by means of Sign Language. one of our primary centers is to produce way better exactness for acknowledgment of words by counting considerable sum of preparing.

REFERENCES

[1] S.Reshna, "Recognition of static hand gestures of Indian sign Language using CNN," Published online on 16 April 2020. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
[2] Rajat Soni, "Real-time Recognition Framework for Indian Sign Language Using Fine-Tuned Convolutional Neural Network." SCRS proceeding of International Conference of Undergraduate Student, 95-106. Published on 2021
[3] Ankita Wadhawan, "Deep learning-based sign language recognition system for static signs." London: Springer-Verlag, 2020
[4] Harleen Kaur, "Image-based Indian Sign Language Recognition: A practical Review Using Deep Learning Network" unpublished.
[5] Sukanya L, "Indian Sign Language Recognition using Convolution neural Network," *E3S Web of Conference.391.01058(2023)*
[6] Rachana Patil, "Indian sign Language Recognition using Convolutional Neural Network,"ITM Web of Conferences 40, 03004(2021)
[7] Imran Khan, "Design and Implementation of CNN for Sign Language Recognition." Jilin Daxue Xuebao (Gong Xue Ban)/Journal of Jilin University (Engineering and Technology Edition), E-publication Online Open Access, Vol:41 Issue:11-22
[8] Chandra Mani Sharma, "Indian Sign Language Recognition Using Fine-Tuned Deep Transfer Learning Model", unpublished.
[9] Meena Ugale, "A Review on Language Recognition Using CNN", Published on 2023
[10] Pias Paul, "Modern Approach for Sign Language Interpretation Using Convolutional Neural Network" published in 2019.
[11] Ilias Papastratis, "Artificial Intelligence Technologies for Sign Language", Published on 30 August 2021.
[12] Pratyush Sinha, "Real Time Sign Language Prediction using CNN and LSTM." volume:05/Issue:04 published in 2023.
[13] Aniket Kumar, "Indian Sign Language Gesture Recognition in Real-Time Using Convolution Neural Network." published in 2021 8th International Conference.

[14] Shashidhar R, "Indian Sign Language to Speech Conversion Using Convolution Neural Network." Published in 2022 IEEE.
[15] Muhummad saad Amin, "A Comparative Review on Applications of Different Sensors for Sign Language Recognition." Published on 2 april 2022.
[16] Dongxu Li, "Word Level deep Sign Language Recognitions from Video: A new Large-Scale Dataset and Method Comparison." Published in 2020.
[17] Sarfaraz Masood, "American Sign Language Character Recognition using Convolution Neural Network." Unpublished.
[18] Sakshi Sharma, "Recognition of Indian Sign Language(ISL) Using Deep Learning Model." Published Online:28 september 2021.
[19] Pranati Rakshit, "Sign Language Detection using Convolutional Neural Network." Published Online:26 march 2024.
[20] Muhammad AL-Qurishi, "Deep Learning for Sign Language Recognition:Current Techniques, Benchmarks, and Open Issue." Published on september 7 2021.