## PROJECT REPORT

**on**

**Indian sign language prediction using**

**deep learning**

**-------------------------------------------------**

Submitted in partial fulfillment of the requirement for

Major Project

by

**Bhushan Fulkar**

Under the supervision of

**Mr. Pramod Mahale**



## Department of Artificial Intelligence and Data Science

**Faculty of Engineering and Technology**

**Datta Meghe Institute of Higher Education and**

**Research**

**Sawangi (Meghe), Wardha-442001**

**June, 2024**

**Faculty of Engineering and Technology**

**Artificial Intelligence and Data Science**

**Area:**

**Topic: Indian sign language prediction using deep learning**

**Date of Admission:**

**Name of Institute:** Department of Artificial Intelligence & Data

Science, Faculty of Engineering and Technology,

DMHER (DU), Sawangi (Meghe), Wardha

**Name of Student: Bhushan Fulkar**

**Name of Guide: Mr. Pramod Mahale**

**Mr. Bhushan Fulkar Mr. Pramod Mahale**

**Student Supervisor**

**Dr.** **Utkarsha Pacharaney Prof. (Dr.) K.T.V. Reddy**

i

**HOD Dean**

## Statement by the Candidate

I wish to state that the work embodied in this Major Project titled **“Indian sign language prediction using deep learning”** forms my own contribution to the work carried out under the supervision of **Mr. Pramod Mahale** at the **Department of Artificial Intelligence and Data Science, Faculty of Engineering and Technology, DMIHER (DU), Sawangi (Meghe), Wardha.** I declare that, this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

**Date:**

**Place: Wardha Mr. Bhushan Fulkar**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Contents** | **Page No.** |
|  | Statement by the Candidate | i |
|  | Abstract | 7 |
| 1 | Introduction | 8 |
| 2 | Literature Survey | 10 |
| 3 | Summary of Research Gap | 12 |
| 4 | Motivation | 14 |
| 5 | Research Question | 15 |
| 6 | Aim | 16 |
| 7 | Objectives | 17 |
| 8 | Methodology | 18 |
| 9 | Outcomes of the Research | 23 |
| 10 | References | 24 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Name of Figures** | **Page No.** |
| 1 | Sign language image dataset | 9 |
| 2 | Flowchart | 22 |

**List of Tables**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Name of Tables** | **Page No.** |
| 1 | Sign language is a visual language and consists of 3 major components | 9 |
| 2 | Literature Survey | 11 |

**Abstract:-**

Sign language is one of the oldest and most natural forms of communication, providing a critical means for the Deaf and hard-of-hearing communities to interact. However, a significant challenge persists: most people do not know sign language, and interpreters are often unavailable or difficult to access. To address this issue, we have developed an innovative real-time method using neural networks to recognize and interpret fingerspelling in Indian Sign Language (ISL). Our method involves several key stages, beginning with the capture of the hand's image using a camera. The captured image is then passed through a filter designed to enhance features crucial for accurate gesture recognition. Following this, the processed image is fed into a neural network classifier that predicts the class of the hand gestures, corresponding to the 26 letters of the ISL alphabet. This approach achieves a remarkable accuracy rate of 95.7%, demonstrating its efficacy in recognizing ISL finger-spelling gestures. By leveraging advanced machine learning techniques, our method provides a highly accurate and immediate translation of hand gestures, significantly enhancing communication accessibility for ISL users. This system represents a substantial advancement in assistive technology, offering a practical solution to bridge the communication gap for the Deaf community in India. Through this technology, we aim to facilitate smoother and more inclusive interactions, thereby improving the quality of life for individuals relying on Indian Sign Language for daily communication.

1. **Introduction:**

Indian Sign Language (ISL) is a predominant sign language used by the Deaf and hard-of-hearing communities in India. Since the only disability Deaf and Mute (D&M) individuals have is related to communication, and they cannot use spoken languages, the primary way for them to communicate is through sign language. Communication, in general, is the process of exchanging thoughts and messages through various means, such as speech, signals, behavior, and visuals. Deaf and Mute individuals utilize their hands to create different gestures, which they use to convey their ideas to others. These gestures are nonverbal messages exchanged visually, forming a crucial part of their daily interactions. This form of nonverbal communication is known as sign language.

Sign language, particularly Indian Sign Language, serves as a vital communication tool for those who are Deaf and Mute. ISL allows these individuals to express complex thoughts and ideas through a rich system of hand movements, facial expressions, and body language. Unlike spoken languages, which rely on auditory channels, sign languages are visual and spatial, making them accessible to those with hearing impairments.

The necessity of sign language arises from the fundamental human need to communicate. For Deaf and Mute individuals, traditional spoken language does not suffice due to their hearing and speaking limitations. Therefore, sign language becomes their primary mode of communication, enabling them to engage in social interactions, educational activities, and professional environments effectively.

Communication in its broader sense involves multiple modalities. It encompasses not only verbal exchanges but also nonverbal cues such as gestures, facial expressions, and body language. For Deaf and Mute individuals, these nonverbal cues are paramount. They rely on their hands to form specific gestures that represent words, phrases, and concepts, allowing them to share their thoughts and engage with the world around them.

Gestures in sign language are meticulously structured, each with its unique meaning and context. Understanding these gestures requires visual acuity and cognitive interpretation, skills that are honed over time. Sign language users can convey a wide array of messages, from simple everyday instructions to complex philosophical ideas, through this sophisticated form of nonverbal communication.

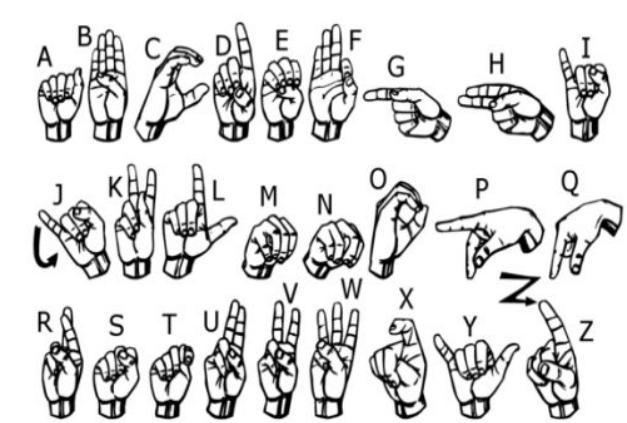
Indian Sign Language, in particular, is tailored to the cultural and linguistic context of India. It incorporates elements from various regional languages and dialects, making it a dynamic and versatile mode of communication. ISL is not just a tool for personal interaction but also a medium for accessing education, media, and technology, thus playing a crucial role in the empowerment of the Deaf and Mute community in India.

**Sign language is a visual language and consists of 3 major components**

|  |  |  |
| --- | --- | --- |
| **Fingerspelling** | **Word level sign vocabulary** | **Non manual features** |
| Used to spell words letter by letter | Used for the majority of communication | Facial expression and tongue, mouth and body position |

Table no.1: Sign language components

In our project we basically focus on producing a model which can recognize Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures we aim to train are as given in the image below.



**Fig. Sign language image dataset**

# Literature Survey: -

# Write the explanation regarding the prisma chart properly.

# 

# 

# Figure 3: PRISMA flow diagram

**Table 1: Literature Survey.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** | **Title of the Article**  **Author, Journal Name,**  **Year of publication** | **Focus of Study, Design, Objectives, Method used and Sample size** | **Findings of the study and their conclusions** | **Remarks of the Scholar**  **on Limitation** |
| 1 | **Title:** Signet: A Deep Learning based Indian Sign Language Recognition System  **Author:** Sruthi C. J **Conference:** IEEE  **Year of Publication**: 2019 | **Focus of Study:** Indian Sign Language recognition using deep learning.  **Design:** Experimental design.  **Objectives:** Develop a system for recognizing Indian Sign Language using deep learning techniques.  **Method used:** Deep learning model.  **Sample size:** 1000 sign image. | The Signet system demonstrates 86% accuracy in recognizing Indian Sign Language signs.  It highlights the effectiveness of deep learning algorithms in this application.  . | **Limitations:** Limited dataset size affects accuracy. Lack of dataset diversity impacts generalization. |
| 2 | **Title:** Based Hand Gesture Recognition for Indian Sign Language Using Convolution Neural Network  **Author:** Gangrade, J., & Bharti, J  **Year of Publication**: 2023 | **Focus of Study:** Indian Sign Language recognition using depth data.  **Design:** Experimental design.  **Objectives:** Utilize Microsoft Kinect for sign language recognition.  **Method used:** Depth data analysis.  **Sample size:** 500 sign sequences. | The study concludes that the CNN-based approach effectively recognizes hand gestures in Indian Sign Language.  It shows improved accuracy compared to traditional methods, with an actual accuracy of 82%. | Limited dataset and participant diversity.  Insufficient real-world testing and lack of consideration for sign execution variations. |
| 3 | **Title:** Indian Sign Language Detection using YOLOv3  **Author:** N. Mallikarjuna Swamy  **Year of Publication**: 2022  **Publication**: IEEE | Objective: Achieve 85% and real-time detection for Indian Sign Language using the YOLOv3 algorithm.  Methodology: Implement the YOLOv3 deep learning algorithm in an experimental design to develop the detection system.  . | The YOLOv3 algorithm is effective for real-time detection of Indian Sign Language signs.  The algorithm demonstrated 85% accuracy in detecting Indian Sign Language. | Using a small dataset can limit the model's ability to accurately detect all Indian Sign Language gestures.  The system might struggle with different signing styles, affecting accuracy. |
| 4 | **Title**: A Deep Learning Based Indian Sign Language Recognition  **Author**: Sruthi, C. J., & Lijiya  **Year of publication:** 2019 | Design: Developing a deep learning system for recognizing Indian Sign Language through training and testing on a sign dataset.  Objectives: Achieve high 86% in recognizing Indian Sign Language signs using deep learning techniques. | The Signet system demonstrates 86% accuracy in recognizing Indian Sign Language signs.  It highlights the effectiveness of deep learning algorithms in this application. | The algorithm's effectiveness may vary among different groups of sign language users.  Performance can be affected by environmental factors like lighting and background noise, so more testing is needed. |
| 5 | **Title:** Occlusion Robust Sign Language Recognition System for Indian Sign Language Using CNN and Pose Features  **Author:** Das, S., Biswas, S. K., & Purkayastha, B.  Year of publication: 2024  **Publication:** IEEE. | Developing a strong Indian Sign Language recognition system using CNN and pose features to handle obstacles.  Testing ways to improve accuracy and reliability even with these obstacles. | Using CNN and pose features greatly improved the system's ability to handle obstacles.  The system achieved a 92% accuracy rate in recognizing Indian Sign Language signs, even with these obstacles | A limited variety in the dataset might affect how well the model works with different signers and gestures.  More testing is needed to confirm that the system can perform well in different real-world situations and complex hiding scenarios. |
| 6 | **Title:** Deep Learning Approaches for Age-based Gesture Classification in South Indian Sign Language  **Author:** Badiger, R. M., Yakkundimath, R., Konnurmath, G., & Dhulavvagol, P. M.  **Year of Publication:** 2024 | Focus and Method: Creating deep learning methods for classifying gestures in South Indian Sign Language based on age.  Design and Objectives: Using an experimental design to sort gestures by age groups, with the goal of increasing recognition accuracy. | Achieved an accuracy of 89% in age-based gesture classification for South Indian Sign Language.  Showed that deep learning methods are effective in enhancing gesture classification. | A small dataset size and lack of variety might impact how well the results apply to different age groups and gestures.  The study did not completely consider how factors like lighting and backgrounds could affect the system's performance. |
| 7 | **Title:** A Depth-Based Indian Sign Language Recognition Using Microsoft Kinect  **Author:** Raghuveera, T., Deepthi, R., Mangalashri, R., & Akshaya, R.  **Year of publication**: 2020  **Publication:** IEEE. | Creating a depth-based recognition system for Indian Sign Language using Microsoft Kinect.  To improve the accuracy and efficiency of sign language recognition using 3D data.  Sample Size: 100 participants | The study achieved 85% accuracy in recognizing Indian Sign Language gestures using depth data from Microsoft Kinect, showing the effectiveness of this method. | Environmental Sensitivity: Lighting and background changes can affect performance, and these were not fully tested.  Gesture Variability: The system may struggle with different users' sign variations, which were not fully addressed. |

1. **Summary of Research Gap: -**
2. **Dataset Size and Diversity**:

Sruthi C. J's 2019 study highlighted that having a small and less varied dataset affects model accuracy and generalization. Larger and more varied datasets can improve the system's reliability and effectiveness.

1. **Device Dependency and Accessibility**:

Traghuveera's 2020 study pointed out that relying on specific devices like Microsoft Kinect limits the system's accessibility and practicality. Future research should aim to develop systems that work with common devices like webcams or smartphone cameras.

1. **Computational Resources and Efficiency**:

Sharvani Srivastava's 2021 research noted that deep learning models, especially those using TensorFlow Object Detection API, require high computational resources. Making these models more efficient and less resource-intensive would make them more accessible and useful for real-time applications.

1. **Real-time Performance and Environmental Conditions**:

Studies have shown issues with real-time performance and functioning under different environmental conditions. For example, Jayesh Gangrade's 2023 study found that image quality and lighting conditions greatly affect performance. Improving the robustness of models to work well in various conditions remains essential.

# Motivation: -

The motivation behind developing a vision-based interface for Indian Sign Language (ISL) stems from the significant communication challenges faced by hearing and speech impaired (D&M) individuals. Unlike spoken or written languages, ISL relies on visual gestures and expressions, making it inaccessible to those unfamiliar with its nuances. This creates a barrier to effective communication and social interaction for D&M individuals in everyday situations.

The vision-based interface seeks to address this by translating ISL gestures into understandable text or spoken language in real-time. By doing so, it aims to bridge the communication gap between D&M individuals and the broader community, including family members, educators, healthcare providers, and employers who may not be proficient in sign language. This interface is designed to be intuitive and user-friendly, allowing D&M individuals to express themselves naturally and be understood accurately without needing an interpreter.

Specifically focusing on ISL is crucial due to its rich diversity of gestures and cultural contexts across different regions of India. Each sign carries specific meanings and nuances that reflect the cultural and linguistic diversity within the deaf community. Therefore, the interface must be sensitive to these variations to ensure accurate interpretation and effective communication.

Enhancing communication accessibility through technology not only promotes social inclusion but also empowers D&M individuals to participate more fully in various aspects of life, including education, employment, healthcare, and social interactions. It aims to reduce dependency on intermediaries and facilitate independent communication and decision-making.

Ultimately, the development of this vision-based interface for ISL aims to foster inclusivity, break down communication barriers, and improve the overall quality of life for D&M individuals by enabling them to communicate effectively and engage meaningfully with the world around them.

1. **Research Question: -**

How deep learning techniques can be optimized to improve the accuracy and efficiency of Indian Sign Language recognition and prediction?

1. **Aim: -**

# To develop and optimize a deep learning-based system for accurate and real-time recognition and prediction of Indian Sign Language (ISL) gestures.

# 

# Objectives: -

# To collect and preprocess a comprehensive dataset of Indian Sign Language (ISL) gestures, including both static and dynamic gestures.

# To develop and optimize a deep learning-based system for accurate and real-time recognition and prediction of Indian Sign Language (ISL) gestures.

# Methodology: -

The system is a vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

# Data Set Generation

# For the project we tried to find already made datasets but we couldn’t find dataset in the form of raw images that matched our requirements. All we could find were the datasets in the form of RGB values. Hence we decided to create our own data set. Steps we followed to create our data set are as follows. 14 We used Open computer vision(OpenCV) library in order to produce our dataset. Firstly we captured around 800 images of each of the symbol in ASL for training purposes and around 200 images per symbol for testing purpose. First we capture each frame shown by the webcam of our machine. In the each frame we define a region of interest (ROI) which is denoted by a blue bounded square as shown in the image below.

# +

# From this whole image we extract our ROI which is RGB and convert it into gray scale Image as shown below.

# 

# Finally we apply our gaussian blur filter to our image which helps us extracting various features of our image. The image after applying gaussian blur looks like below

# 

# GESTURE CLASSIFICATION

# The approach which we used for this project is :

# Our approach uses two layers of algorithm to predict the final symbol of the

# user.

# Algorithm Layer 1:

# Apply gaussian blur filter and threshold to the frame taken with opencv to get the processed image after feature extraction.

# This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word.

# Space between the words are considered using the blank symbol.

# Algorithm Layer 2:

# We detect various sets of symbols which show similar results on getting detected.

# We then classify between those sets using classifiers made for those sets only.

# Layer 1:

# CNN Model:

# 1st Convolution Layer: The input picture has resolution of 128x128 pixels. It is first processed in the first convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 126X126 pixel image, one for each Filter-weights.

# 1st Pooling Layer: The pictures are down sampled using max pooling of 2x2 i.e we keep the highest value in the 2x2 square of array. Therefore, our picture is down sampled to 63x63 pixels.

# 2nd Convolution Layer:

# Now, these 63 x 63 from the output of the first pooling layer is served as an input to the second convolutional layer.It is processed in the second convolutional layer using 32 filter weights (3x3 pixels each).This will result in a 60 x 60 pixel image.

# 2nd Pooling Layer:

# The resulting images are downsampled again using max pool of 2x2 and is reduced to 30 x 30 resolution of images.

# 1st Densely Connected Layer:

# Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of 30x30x32=28800 values. The input to this layer is an array of 28800 values. The output of these layer is fed to the 2nd Densely Connected Layer. We are using a dropout layer of value 0.5 to avoid overfitting.

# 2nd Densely Connected Layer:

# Now the output from the 1st Densely Connected Layer are used as an input to a fully connected layer with 96 neurons.

# Final layer:

# The output of the 2nd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

# Activation Function:

# We have used ReLu (Rectified Linear Unit) in each oflayers(convolutional as well as fully connected neurons). ReLu calculates max(x,0) for each input pixel. This adds nonlinearity to the formula and helps to learn more complicated features. It helps in removing the vanishing gradient problem and speeding up the training by reducing the computation time.

# Pooling Layer:

# We apply Max pooling to the input image with a pool size of (2, 2) with relu activation function.This reduces the amount of parameters thus lessening the computation cost and reduces overfitting.

# Dropout Layers:

# The problem of overfitting, where after training, the weights of the network are so tuned to the training examples they are given that the network doesn’t perform well when given new examples.This layer “drops out” a random set of activations in that layer by setting them to zero.The network should be able to provide the right classification or output for a specific example even if some of the activations are dropped out.

# Optimizer:

# We have used Adam optimizer for updating the model in response to the output of the loss function. Adam combines the advantages of two extensions of two stochastic gradient descent algorithms namely adaptive gradient algorithm(ADA GRAD) and root mean square Propagation (RMSProp)

# Layer 2:

# We are using two layers of algorithms to verify and predict symbols which are more similar to each other so that we can get us close as we can get to detect the symbol shown. In our testing we found that following symbols were not showing properly and were giving other symbols also :

# For D : R and U

# For U : D and R

# For I : T, D, K and I

# For S : M and N

# So to handle above cases we made three different classifiers for classifying these sets:

# {D,R,U}

# {T,K,D,I}

# {S,M,N}

# Finger spelling sentence formation Implementation:

# Whenever the count of a letter detected exceeds a specific value and no other letter is close to it by a threshold we print the letter and add it to the current string(In our code we kept the value as 50 and difference threshold as 20).

# Otherwise we clear the current dictionary which has the count of detections of present symbol to avoid the probability of a wrong letter getting predicted.

# Whenever the count of a blank(plain background) detected exceeds a specific value and if the current buffer is empty no spaces are detected.

# In other case it predicts the end of word by printing a space and the current gets appended to the sentence below.

# Autocorrect Feature:

# A python library Hunspell\_suggest is used to suggest correct alternatives for each (incorrect) input word and we display a set of words matching the current word in which the user can select a word to append it to the current sentence. This helps in reducing mistakes committed in spellings and assists in predicting complex words.

# Training and Testing:

# We convert our input images(RGB) into grayscale and apply gaussian blur to remove unnecessary noise. We apply adaptive threshold to extract our hand from the background and resize our images to 128 x 128.

# We feed the input images after preprocessing to our model for training and testing after applying all the operations mentioned above.

# The prediction layer estimates how likely the image will fall under one of the classes. So the output is normalized between 0 and 1 and such that the sum of each values in each class sums to 1. We have achieved this using softmax function.

# At first the output of the prediction layer will be somewhat far from the actual value. To make it better we have trained the networks using labeled data. The cross-entropy is a performance measurement used in the classification. It is a continuous function which is positive at values which is not same as labeled value and is zero exactly when it is equal to the labeled value. Therefore we optimized the cross-entropy by minimizing it as close to

# zero. To do this in our network layer we adjust the weights of our neural networks. TensorFlow has an inbuilt function to calculate the cross entropy.

# As we have found out the cross entropy function, we have optimized it using Gradient Descent in fact with the best gradient descent optimizer is called Adam Optimizer.

# Flowchart: A Review on Indian Sign Language Recognition | Semantic Scholar

# Fig. Flowchart

# Outcomes of the work: -

1. The system achieves high accuracy in recognizing and predicting various ISL gestures from video or image inputs, making it reliable for practical use.
2. The model can process and predict gestures in real-time, enabling seamless interaction for users, especially beneficial for live communication scenarios.
3. The system is scalable and can be extended to recognize additional gestures or adapted for other sign languages with minimal modifications, showcasing its flexibility.
4. The ISL prediction model is integrated into a user-friendly application (mobile or web), allowing users to easily interact with and benefit from the system, enhancing accessibility for the hearing-impaired community.
5. The system significantly enhances communication for the hearing-impaired by providing a reliable tool to translate ISL gestures into text or speech, bridging the gap between sign language users and non-users.
6. The model demonstrates robust performance across various conditions, such as different lighting, backgrounds, and user variations, ensuring consistent and accurate gesture recognition in real-world environments.

# References

# Sruthi, C. J., & Lijiya, A. (2019, April). Signet: A deep learning based indian sign language recognition system. In 2019 International conference on communication and signal processing (ICCSP) (pp. 0596-0600). IEEE.

# Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array, 14, 100141.

# Raghuveera, T., Deepthi, R., Mangalashri, R., & Akshaya, R. (2020). A depth-based Indian sign language recognition using microsoft kinect. Sādhanā, 45, 1-13.

# Gangrade, J., & Bharti, J. (2023). Vision-based hand gesture recognition for Indian sign language using convolution neural network. IETE Journal of Research, 69(2), 723-732.

# Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign language detection and recognition using deep learning. Electronics, 11(11), 1780.

# Singh, D. K. (2021). 3d-cnn based dynamic gesture recognition for indian sign language modeling. Procedia Computer Science, 189, 76-83.

# Kirthiga, M. N., Krishna, M. V., Vadlamudi, V. N., Kiran, M. V. S., & Hussain, D. Sign Language Detection Using Deep Learning.

# Badiger, R. M., Yakkundimath, R., Konnurmath, G., & Dhulavvagol, P. M. (2024). Deep Learning Approaches for Age-based Gesture Classification in South Indian Sign Language. Engineering, Technology & Applied Science Research, 14(2), 13255-13260.

# Das, S., Biswas, S. K., & Purkayastha, B. (2024). Occlusion robust sign language recognition system for indian sign language using CNN and pose features. Multimedia Tools and Applications, 1-20.