Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250095
Published: November 6, 2025

I-SRAVIA-Breaking the Wall of Silence: Prototyping Testing and Validating

Shraavya Mishra
High School Student (Class 10th)
Kendriya Vidyalaya No. 2, Bhubaneswar, Odisha, India

Sumona Karjee Mishra
R&D head Prantae Solutions Private Limited Bhubaneswar, Odisha, India
shraavyamishra@gmail.com

Abstract

Approximately 63 million individuals in India have significant hearing and speech impairments. This creates substantial communication barriers that restrict social, educational, and occupational inclusion. Current solutions rarely support bidirectional, real-time interaction tailored for Indian Sign Language (ISL). In this work, we introduce I-SRAVIA (Indian Sign-Language Responsive and Voice Intelligent Assistant), a computer vision-driven prototype enabling two-way communication between ISL users and hearing individuals.The system employs a dataset of 1,149 ISL gesture images across nine classes captured by a webcam. The image dataset was augmented to 14,937 samples by treating the real dataset with various parameters. The feed-forward Multi-Layer Perceptron (MLP) sequential model specialized for Convolutional Neural Network (CNN) to classify images into nine ISL words for the model generation that. was trained over 25 epochs (batch size 32). Software evaluations—comprising training/validation accuracy, loss metrics, and confusion matrix analysis—demonstrated near-perfect performance with zero misclassification among five tested gestures. Real-time trials involving 160 gesture inputs produced a Mean Magnitude of Relative Error (MMRE) of 15.6%, equivalent to 84.4% prediction accuracy. Orientation robustness tests confirmed reliable gesture recognition within ±25° deviations. The user interface was developed using Flask, HTML/CSS, and JavaScript using principles of human factors engineering. The interface supports both gesture-to-text/voice and voice-to-text modes.
These findings demonstrate the feasibility of a reliable, two-way ISL-based communication platform. Future work would expand the gesture lexicon library, leverage enhanced computational resources, and conduct usability testing in real-world environments to take this work toward real case deployment.

Keywords

Indian Sign Language, I-SRAVIA, Two-way communication, Computer Vision, Gesture Recognition, Convolutional Neural Network (CNN)

1. Introduction

Approximately 430 million individuals, or one-third of the disabled population across the globe, are deaf and mute (WHO, 2025). In India alone, an estimated 63 million people face hearing and speech impairments, forming a

significant portion of the disabled community (Directorate General of Health Services Government of India, 2025). These individuals encounter profound barriers that hinder their ability to connect, participate, and thrive.
The communication gap or the **"Wall of Silence"** creates significant challenges such as:

a) **Limited Job Opportunities:** A lack of effective communication tools restricts access to employment, leading to financial instability.
b) **Social Isolation:** Exclusion from conversations and social activities fosters loneliness and alienation.
c) **Mental Health Challenges:** Isolation and frustration often lead to depression and anxiety.
d) **Non-Inclusive Society:** Communication barriers perpetuate stereotypes and hinder societal inclusion.

Bridging this divide through innovative solutions is critical. Empowering the deaf and mute community with accessible communication tools is essential to fostering equality, ensuring full participation in society, and building a truly inclusive world.

1.1. Objectives

1.1.1. To create a model for two-way communication with primary focus on Indian sign Language.
1.1.2. To evaluate model performance

2. Literature Review

As per the World Health Organisation (WHO) approximately 20% of the global population, or 1.6 billion people are living with some degree of hearing disability. Out of those with hearing loss, 430 million have disabling hearing loss, meaning their hearing loss significantly impacts their daily lives (WHO, 2025). In India, there are approximately 63 million people, who are suffering from significant auditory impairment (Directorate General of Health Services Government of India, 2025). In a world where communication is central to the society structure, millions with hearing and speech impairments remain unheard and excluded.

These individuals encounter profound barriers that hinder their ability to connect, participate, and thrive. The wall of silence keeps these individuals away from equal opportunities in education, profession as well as social acceptability. Of these, a large percentage is children between the ages of 0 to 14 years (Directorate General of Health Services Government of India, 2025) that constitute a large future population is at risk of being lost from the workforce to the disability. Sign language is the only weapon that has the power to break this wall of silence. Sign language is a form of language that expressed and visually perceived primarily through hand gestures instead of spoken words. Hand gestures would assist and facilitate communication among people by providing a meaningful interaction. As with any spoken language, sign language has grammar and structure rules, and more than 300 different sign languages in the world has evolved over time (National Geographic Education, 2025). Indian Sign Language (ISL) or a closer form Indo-Pakistan Sign Language (IPSL) is the most used sign language globally but unfortunately there is limited technological advances has been used on ISL or IPSL (Saini et al., 2023). The major volume of work has been done on American Sign Language where numerous systems have also been developed using sensors, image processing, and smartphones to facilitate communication between the two parties. However, most of these systems are limited to one direction of communication from the deaf-mute to non-deaf-mute people and not vice versa (Amal et al., 2023).
Artificial Intelligence (AI) drives computer vision—a subfield enabling machines to 'see' and interpret images using algorithms and deep learning—empowering robotics, healthcare diagnostics, automotive systems, manufacturing, surveillance, and more (Juneja et al., 2021). Recent breakthroughs in computer science have elevated human–computer interaction beyond mouse, keyboard, and touch—now embracing voice and gesture recognition, enabling more natural, intuitive, and hands-free control (Zhang et al., 2021).

In this paper, we are presenting our work I-SRAVIA (Indian Sign-Language Responsive And Voice Intelligent Assistant for the Deaf and Mute) that is an assistive system designed to enable the deaf and mute community by two-way communication. Unlike the American sign language (ASL) that has received consistent technological attention, I-SRAVIA is a multi-modal solution, specifically tailored to Indian Sign Language (ISL), combines cutting-edge

computer vision, machine learning, and natural language processing to facilitate real-time communication between deaf and mute individuals and the hearing population (Zengeler et al., 2018).

3. Methods

The development of I-SRAVIA followed a structured methodology comprising data collection, pre-processing, model training, testing, and user interface development. The workflow of the I-SRAVIA has been illustrated in Figure 1

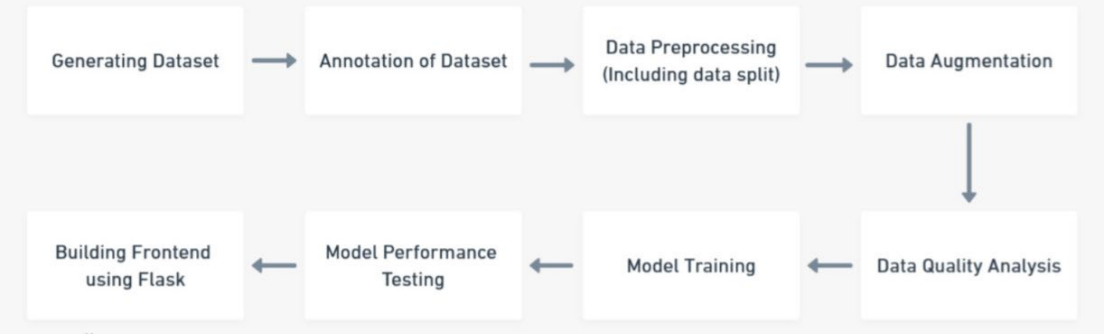

Figure 1: Flow diagram representing the process of I-SRAVIA development.

3.1. Pre-processing

The gesture data collected as described in Section 4, was processed using the library sklearn.preprocessing.LabelEncoder for encoding the labels into numbers, sklearn.preprocessing.StandardScalar for normalizing the hand key points to achieve Standard deviation = 1 and Mean = 0, sklearn.model_selsction.train_test_split that compartmentalized the dataset 8:2 ratio for training and validation, respectively. This pre-processed data and label codes were converted .npy files by numpy.

3.2. Data augmentation

In order to increase the data set and to provide on field scenario input we augmented the collected data set for both training and validation with the following parameters: a) horizontal flip, b) random rotation within the limit of $\pm$ 25°, c) random brightness adjustment $\pm$ 20% and d) blurring effect with random sigma. We used imgaug.augmenters library of python.

3.3. Data Quality Analysis

The check the quality of the total data set we performed Principal Component Analysis (PCA) on the dataset using sklearn.decomposition pca and matplotlib to visualize the output graphically.

3.4. Model Architecture

TensorFlow and KERAS (Chollet et al., 2015) and their modules has been used to bult, compile and train the feed-forward Multi-Layer Perceptron (MLP) sequential model specialized for Convolutional Neural Network (CNN) to classify images into nine ISL words. The training has been done on 25 epoch and batch size 32. Each image input data has been converted into compact 63-dimensional feature vector (typically 21 hand landmarks × 3 coordinates {x,y,z}) per sample. Dense layer of 256 neurons and 128 neurons have been used with Rectified Linear Unit (ReLU) along with the dropout layer to prevent overfitting. The final dense layer with 9 neurons and Softmax.

3.5. Model Validation

3.5.1. Software based testing: Model has been analysed with the testing subset of the data and model's performance has been evaluated on metrices such as testing and validation accuracy, testing and

validation loss. Confusion matrix has been generated to identify areas where the model struggled with classification.

3.5.2. Gesture based testing: Real four ISL hand gestures have been made and prediction accuracy and mean magnitude of relative error has been calculated (Jorgensen et al., 2022).

Where, MMRE = (1/n) * Σ [|predicted_i - actual_i| / |actual_i|] for i = 1 to n.

Prediction Accuracy = 1- MMRE

The ISL gestures selected for the testing A= Yes, B= Please, C= Hello and D= No

3.6. UI/UX for the frontend

Flask, HTML, CSS and JavaScript along with the input from the volunteers to imbibe the principles of Human Factor Engineering to make it aesthetically pleasing and easy to use interface.

4. Data Collection

Gesture data corresponding to Indian Sign Language (ISL) was captured using an inbuilt webcam (Asus Zenbook). The resultant pictures are annotated with their corresponding words and arranged in the folder with the label.

Image Capture and Annotation:

Minimum of 100 images were captured for each ISL gesture. A Python script (101 lines) was developed using the following libraries:

MediaPipe Hand Landmarker: to detect hand keypoints and render visual effects

OpenCV: to capture images

NumPy: for numerical computation

os: for file management

xml.etree.ElementTree: for XML-based annotation

5. Results and Discussion

I-SRAVIA platform construction initiated from the data-collection with the real-life image capture through the customized program coded. As a very limited amount of work has been done on ISL so using existing dataset was not the possibility unlike ASL based models [11]. As a prototype we started with 9 common ISL words, viz., "OK", "Hello", "Yes", "No", "Please". "I", "You" and "Thank You". We have captured a minimum of 100 images for each word totalling to 1149 images (Table 1: Real data column). These images are then annotated and stored in separate labelled folders for each image set corresponding to the ISL word. This annotated image pool of 1149 has been randomly divided into 8:2 ratio and classified for training and testing data set. with a condition provided more than 50% and less than 10% of image pool corresponding to a single word should not go to testing data set. This will ensure all words are well represented in the training as well testing dataset. This pool was then augmented irrespective of its classification based on parameters mentioned in section 3.2. Thereby we have generated 14937 datasets, where 11950 corresponds to training and 2987 testing data set to create I-SRAVIA (Table 1).

Table 1. Representing the real captured data and augmentation adopting different parameters where a) horizontal flip, b) random rotation within the limit of ± 25°, c) random brightness adjustment ± 20% and d) blurring effect with random sigma.

S.No.	Word	Real data	Parameters				Total
			a	b	c	d	
1	'OK'	123	123	492	615	369	1599
2	'Hello'	161	161	644	805	483	2093
3	'Yes'	102	102	408	510	306	1326
4	'No'	131	131	524	655	393	1703
5	'Please'	142	142	568	710	426	1846
6	'I'	105	105	420	525	315	1365
7	'You'	119	119	476	595	357	1547
8	'Thank You'	145	145	580	725	435	1885
9	'Good'	121	121	484	605	363	1573
Total data set							**14937**

The model architecture has been built using the training dataset generated. We used a lightweight, feed-forward Multi-Layer Perceptron (MLP) to classify nine ISL gestures/words. This keeps the model small (~52k parameters), fast, and appropriate for real-time inference. In order to match the data input as per software requirement we have converted each image into compact 63-dimensional feature vector (typically 21 hand landmarks × 3 coordinates {x,y,z}). Thus, instead of raw images the model was fed with the feature vectors. On such structured numeric inputs, an MLP is simpler and more data-efficient than conventional CNN models (which is designed for 2-D pixel grids) resulting in significant reduction in computational cost and memory requirements for training the MLP (Liu et al., 2024). Furthermore, feature vectors often capture invariant properties of the image, such as edges, textures, or shapes, rather than raw pixel values which are sensitive to variations in lighting, rotation, or scaling. This pre-processing helps the MLP to generalize better to unseen data and become more robust to variations in the input images. In addition, it also reduces the risk of overfitting. The CNN model architecture was built with total 393 Neurons and output was collected with Softmax for output probability distributions in the output layer for multi-class classification. The model has been built with 25 epoch and 32 batch size with repeated batch normalization and dropout layers resulted in a robust Final Layer. The output model has been challenged with the testing dataset (Figure 2).

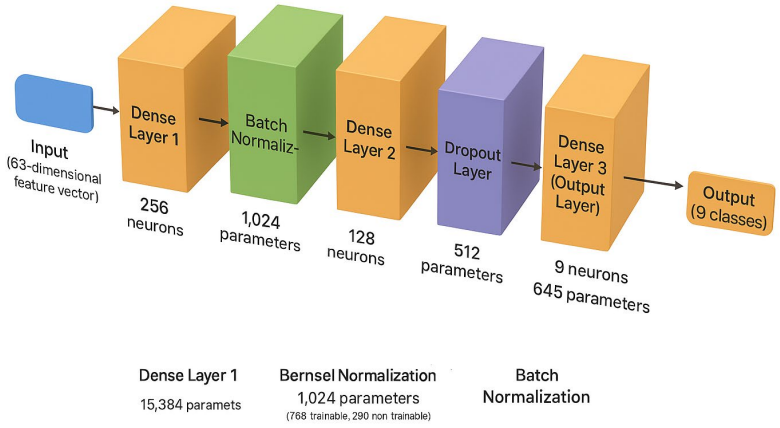

Figure 2. I-SRAVIA: Model Architecture

The model has been trained with Convolutional neural network (CNN) model During software-based testing, the model achieved high training and validation accuracy, corroborated by a low validation loss (Figure 3).

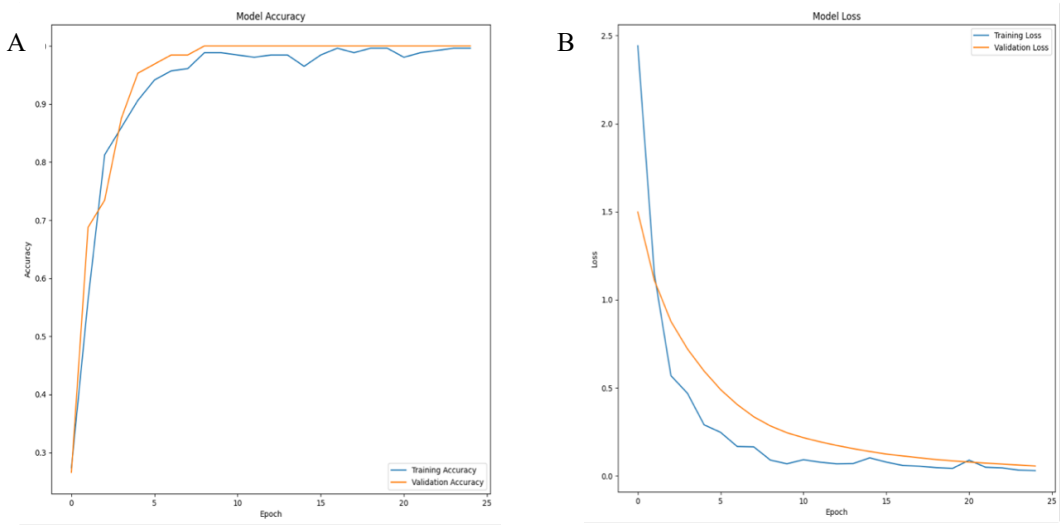

Figure 3: Graphical representation of Training and Validation (A) Accuracy and (B) Loss

The training and validation accuracy steadily increased across epochs, achieving near-perfect performance. During the first 5–7 epochs, both training and validation accuracy rose from around 25–30% to approximately 98–100%. By epochs 8–10, validation accuracy converged to nearly 100%, closely tracking the training accuracy curve. Simultaneously, loss values decreased markedly as the training loss plummeted from about 2.4 at epoch 0 to nearly zero by epoch 10. Training accuracy similarly plateaued near the maximum, with both curves stabilizing through to the 25th epoch (Figure 3 a). Further, the validation loss mirrored this trend, beginning around 1.5 and steadily declining toward zero by the final epochs (Figure 3b). Also, in the confusion matrix as well we get the evidence of all the five ISL words were precisely recognized with zero mis-classification with input n=43 for each word (Figure 4). This level of accuracy we could obtain because of the usage of the Adam Optimizer in our CNN model training. Adam optimizer known for its efficiency and effectiveness, we have used it in our model training deep neural networks with the following parameters learning rate 0.001, β_1=0.9, β_2=0.999, ε=1e-7, categorical cross-entropy and accuracy.

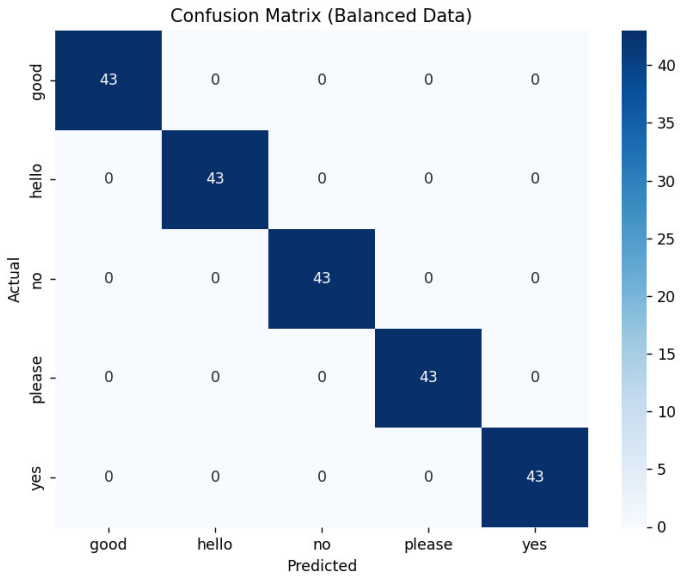

Figure 4: Graphical representation of confusion matrix where ISL words tested x = 5 and the number of times each word input is given n=43.

Thus, the model developed demonstrated rapid learning and strong generalization, indicated by alignment between training and validation curves. The sharp drop in loss values and convergence of accuracy to nearly 100% suggest effective learning and minimal overfitting within the training regime.

To further test the real-life applicability, we tested the model with the real time gesture analysis. For pilot evaluation, we tested the platform with a model sample of 4 ISL gestures, viz. A= Yes, B= Please, C= Hello and D= No, have been tested. Each of the 4 gestures have been randomly repeated 40 times generating n=160 inputs where the application has been able to recognize the gesture correctly 135 times while for the 25 times mis-interpreted (Table 2). Thus, the mean magnitude of relative error (MMRE) is around 15.6% and a prediction accuracy is 84.4%. As MMRE <25% is considered good accuracy the model performance can be categorized as satisfactory (Wu et al., 2025).

Table 2. Representing the real time gesture analysis with four ISL word hand gestures and the relative error and mean magnitude relative error.

ISL Gesture	No. of times gestures made	No. of times gestures correctly identified	Relative Error
A	40	32	0.2
B	40	35	0.125
C	40	32	0.2
D	40	36	0.1
Mean Magnitude Relative Error			**0.15625**
Mean Magnitude Relative Error%			**15.6%**

To test robustness against gesture orientation, single gesture ‘C= Hello’ were performed at different inclination angles (0°, 15°, 30°, 120°, 180°, 195°, 270° and 345°) with n=10 times same gesture. The system classified the gestures more correctly when the angles were 0°, 15°, 165°, 180°, 195° 270° and 345° than when the angles of gesture with respect to 0° were 30°, 120° and 270° (Table 3). The minimum relative error was obtained is 10% for 0° and 100% for 270°

positioning. This can be justified with the fact the model training has been done with the permissible angle of deviation from the ideal gesture position by ± 25° as well as through horizontal flips.

Table 3. Representing angle deviation analysis for ISL gesture where 0° is the reference position representing theoretical gesture and all angles are relative to 0°.

ISL gesture inclination angle	No. of times gestures made	No. of times gestures correctly identified	Relative Error	% Relative Error
0°	10	9	0.1	**10**
15°	10	7	0.3	30
30°	10	2	0.8	80
120°	10	2	0.8	80
180°	10	8	0.2	20
195°	10	6	0.4	40
270°	10	0	1	**100**
345°	10	7	0.3	30

The frontend has been created with soft colour pallet to reduce distraction and make it easy on the eyes. Further we have made three distinct gateways according to the requirement of functionality Gateway1: ISL to text and/or voice (S ® T/V) Voice to text, Gateway 2: Voice to text (V ® T) and Gateway 3: Dual mode (S ® T/V : V ® T) (Figure 5). The I-SRAVIA working and usage has been explained in the video https://www.youtube.com/watch?v=IJzqUcsietY

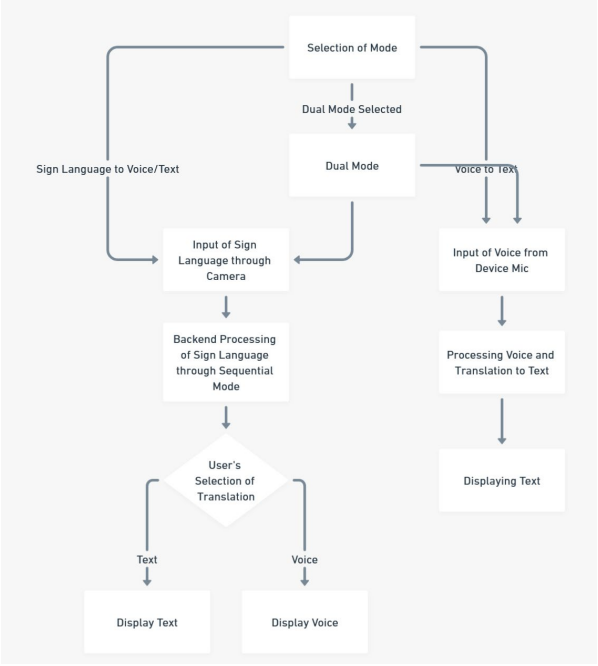

Figure 5: Flow diagram of the UI/UX logic of the I-SRAVIA platform with three gateways for one-way and two-way communication.

6. Proposed Improvements

The I-SRAVIA model has been successfully prototyped using only 9 ISL words and limited dataset. The following proposed improvements will help us further develop this platform for use in real world. (1) Expanding the gesture dataset to encompass a more comprehensive ISL lexicon (estimated at approximately 10,000 signs), (2) Enhancing

computational capacity—such as deploying more capable GPUs—to support larger datasets and deeper models (3) Integrating advanced algorithms or architectures (e.g., transformer models, larger CNNs) to improve accuracy and scalability and (4) Exploring real-world deployment contexts, such as educational environments, workplaces, and public services, to evaluate system usability and impact.

7. Conclusion

The I-SRAVIA prototype establishes a viable, bidirectional communication platform tailored to Indian Sign Language (ISL), addressing a critical accessibility gap for India's hearing- and speech-impaired population. By integrating computer vision, machine learning, and natural language processing, the system achieved near-perfect accuracy and minimal loss during software evaluation, while real-time testing demonstrated satisfactory performance with an MMRE of 15.6% (~84.4% prediction accuracy). Moreover, the model-maintained reliability within ±25° gesture orientation, confirming the efficacy of data augmentation methods. The human-centered interface—comprising gesture-to-text/voice and voice-to-text gateways built using Flask—illustrates the system's usability grounded in human factors engineering principles.

These findings underscore I-SRAVIA's potential to break the "Wall of Silence," enabling inclusive communication using ISL rather than over-reliance on West-centric sign systems. Nonetheless, the prototype's current scope—limited to nine gestures and constrained by hardware capabilities—points to necessary next steps. Future work will expand the gesture vocabulary to encompass a broader ISL lexicon, employ more powerful computational resources to accommodate deeper and more generalized models, and validate usability in real-world contexts such as educational, workplace, and public-service environments. These enhancements will be critical for evolving I-SRAVIA from prototype to widely deployable assistive technology.

Acknowledgement

Authors like to acknowledge the contribution of Shri Harsha Memorial School for Deaf and Mute for providing the support and inspiration for I-SRAVIA. We are thankful to Dr. Aseem Mishra for his technical guidance throughout the project. We are thankful to the Kendriya Vidyalaya No. 2 BBSR Principal Shri D. P. Sharma, teachers S.R. Maharana and R. Patel for their constant support and encouragement.

References

Amal, B., Hind, B., Ohoud, A., Dimah, A., Amal, B., Amal, A. and Hanadi, A., "Intelligent gloves: An IT intervention for deaf-mute people" *Journal of Intelligent Systems* 32, no. 1, 2023

Chollet, F., and others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

https://dghs.mohfw.gov.in/national-programme-for-prevention-and-control-of-deafness.php Accessed on 20th August 2025

https://education.nationalgeographic.org/resource/sign-language/ Accessed on 20th August 2025

https://www.who.int/health-topics/hearing-loss#tab=tab_2. Accessed on 20th August 2025

Jorgensen, M., Halkjelsvik, T. and Liestol, K., When should we (not) use the mean magnitude of relative error (MMRE) as an error measure in software development effort estimation?, *Information and Software Technology,* Vol 143, 2022.

Juneja, S., Juneja, A., Dhiman, G., Jain, S., Dhankhar, A. and Kautish S. computer Vision-Enabled character recognition of hand Gestures for patients with hearing and speaking disability. *Mobile Information Systems*. 2021;2021.

Liu, S.., Wang, L. and Yue, W. An efficient medical image classification network based on multi-branch CNN, token grouping Transformer and mixer MLP, *Applied Soft Computing,Volume* 153, 111323, 2024,

Obi, Y., Claudio, K.S., Budiman, V.M., Achmad, S., and Kurniawan, A., Sign language recognition system for communicating to people with disabilities, 216, 13-20, 2023.

Saini, B., Venkatesh, D., Chaudhari, N., Shelake, T., Gite, S. and Pradhan, B., A comparative analysis of Indian sign language recognition using deep learning models, *Forum for Linguistic Studies*, vol 5. 197-222, 2023.

Wu, Y., Liu, X., Morris, K. D., Lu, S., & Wu, H., An Exploratory Study of Body Measurement Prediction Using Machine Learning and 3D Body Scans. *Clothing and Textiles Research Journal,* 43(3), 187-203, 2025.

Zengeler N, Kopinski T, Handmann U. Hand gesture recognition in automotive human–machine interaction using depth cameras. *Sensors*. 19(1):59 2018.

Zhang, Q., Xiao, S., Yu, Z., Zheng, H. and Wang, P., Hand gesture recognition algorithm combining hand-type adaptive algorithm and effective-area ratio for efficient edge computing. *Journal of Electronic Imaging.* ;30(6):063026, 2021

Biographies

Shraavya Mishra, a Class 10 student at PM-SHRI Kendriya Vidyalaya No. 2, Bhubaneswar, is a young innovator passionate about science and technology. At 14, she explores AI in healthcare and agriculture, winning the INSPIRE-MANAK National award (2024) and joining the Lodha Genius Programme 2025, reflecting her drive for impactful research (https://shraavyamishra.netlify.app/).

Dr. Sumona Karjee is a R&D head of Prantae Solutions Private Limited. She has more than 16 years of experience spanning industry and academics. She has 12 peer reviewed scientific journal as author and has been inventor in 5 granted Indian patents and 2 International and 1 Indian patent applied. She did her PhD from International Centre for Genetic Engineering and Biotechnology, New Delhi in 2010.