PROJECT TITLE

PROJECT REPORT
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE
DEGREE OF BACHELOR OF TECHNOLOGY
IN
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

BY

NAME	Registration No:	Roll No:
Mohini Brahma	**211420100110169**	**14200121194**
Arunangshu Garai	**211420100110175**	**14200121187**
Diptimoy Hazra	**211420100110162**	**14200121207**
Sagnik Banerjee	**211420100110158**	**14200121185**
Md. Yaseer Hassan	**211420100110174**	**14200121198**
Debojyoti Maity	**211420100110170**	**14200121195**

Under The Supervision
of
Ms. Tanushree Chakraborty
Asst. Professor, Dept. of Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MEGHNAD SAHA INSTITUTE OF TECHNOLOGY

Techno Complex, Madurdaha, Beside NRI Complex, post-Uchhepota, Kolkata-700150

Affiliated by

MAULANA ABUL KALAM AZAD
UNIVERSITY OF TECHNOLOGY,
WEST BENGAL

Month Year

CANDIDATE'S DECLARATION

I hereby declare that the work which is being presented in the project entitled **……** in fulfillment of requirements for the award of degree of B.Tech. in CSE, submitted in the Department of Computer Science & Engineering at **MEGHNAD SAHA INSTITUTE OF TECHNOLOLY** under **MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, KOLKATA** is an authentic record of our own work carried out during Session 2024-2025 under the supervision of **……. ……**. The matter presented in this project has not been submitted by us in any other University / Institute for any award.

_____

Signature of the Students
With Date

Meghnad Saha Institute of Technology

Nazirabad, P.O. :Utchepota, Via Sonarpur, Kolkata 700 150

CERTIFICATE

This is to certify that the Project entitled ……. ……. is being submitted by**…….. ……** in partial fulfillment of the requirement for the award of the degree of B.Tech.in Computer Science and Engineering to the Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, is a record of bonafied work carried out by him under my guidance and supervision from _____ to _____

The results presented in this thesis have been verified and are found to be satisfactory. The results embodied in this thesis have not been submitted to any other University for the award of any other degree or diploma.

……
Name of Project Guide
<Designation>
<**Department**>
MeghnadSaha Institute of Technolgy
Kolkata-700150

Date:

Place:

Meghnad Saha Institute of Technology
Nazirabad, P.O. :Utchepota, Via Sonarpur, Kolkata 700 150

CERTIFICATE OF APPROVAL

The foregoing project entitled……..**……**
……..is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the degree for which it has been submitted. It is to be understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

Project Guide _____

Project Coordinator _____

Head of the Department _____

External Examiner _____

Date:

Place:

[Departmental & College Seal]

Acknowledgement

Life as a researcher is a zig-zag way and has many ups and downs, but I overcome from them because of the support and faith of many individuals. First and foremost I would like to express my sincere gratitude to **Dr./Mr./Ms. ……** for providing me an opportunity to work with him. Your creativity encouraged me, your energy and dedication towards work motivates me. Your perfection always inspired me to do things perfectly. I feel motivated and

encouraged every time I attend your meeting. Without your encouragement, inspiration and guidance this thesis / dissertation would not have materialized. Thank you so much for not only guiding me in this thesis / dissertation but for every knowledge you give me. This thesis / dissertation is dedicated to your creativity and enthusiasm of doing work.

I wish to thanks many people for their direct or indirect help in achieving this goal. Special thanks to my parents for their support and encouragement throughout my study, their faith in me ignite that spark to do anything for their happiness.

Last but not least I would like to thanks all my friends for being with me at each step when I need their support. This thesis never be successful without your support and love.

_____

(Signature of the Student-1)

<Name of Student-1>

_____

(Signature of the Student-2)

<Name of Student-2>

_____

(Signature of the Student-3)

<Name of Student-3>

_____

(Signature of the Student-4)

<Name of Student-4>

Abstract

Abstract = Times (Roman) 12 pt leading. All manuscripts should include an abstract of 200-300 words summarizing the significant findings. The abstract should be set up as a justified paragraph (i.e. not flush left).

Keywords:11pt, Italics.Six to twelve keywords or phrases should be supplied to aid inindexing the article.

TABLE OF CONTENTS

Page No.

List of Tables

List of Figures

1. INTRODUCTION

India’s immense linguistic diversity, with 22 official languages and countless dialects [1], presents significant communication challenges, especially for individuals who are deaf and speech-impaired. These individuals often face barriers in education, employment, and daily life

due to the lack of accessible tools for translating between sign language, text, and speech. This communication gap limits their social inclusion and growth opportunities.

Indian Sign Language (ISL)[2] users frequently find that current multilingual translation and sign language interpretation solutions fall short of their demands. Many tools struggle with accuracy in complicated situations or real-time processing, and many are unable to adjust to the specific language and gestures of ISL. These flaws make it difficult to communicate effectively and lead to frequent misunderstandings.

We suggest an AI-powered multilingual translation and sign language interpreter made especially for Indian languages and ISL in order to solve these problems. The system guarantees accurate and real-time hand gesture identification by utilizing the YOLOv11 deep learning model.

This system bridges the gap between text, speech, and sign language, empowering hearing- and speech-impaired individuals. By fostering accessibility and inclusivity, this research contributes to a more equitable and digitally empowered society. This solution promotes inclusivity and smooth communication between hearing and speech-impaired people and the general public by supporting many regional Indian languages

1.1Purpose

The primary purpose of this research is to develop an innovative, AI-driven solution to bridge the communication gap faced by individuals with hearing and speech impairments in India. This study focuses on creating a robust and accurate multilingual system that translates Indian Sign Language (ISL) gestures into text and speech. The research emphasizes the application of the YOLOv11 deep learning model for precise gesture recognition and using Generative AI to ensure contextually relevant translations. By addressing existing limitations, such as the lack of real-time gesture interpretation, limited datasets, and inadequate support for diverse Indian languages, the study aims to deliver a user-friendly system that integrates seamlessly into daily life.

1.1. Domain Definition

- Assistive Technology:

  Provides the framework for creating tools that enhance accessibility for the Deaf and Hard of Hearing (DHH) community.

  - Facilitates effective communication between Indian Sign Language (ISL) users and non-users.

- Artificial Intelligence (AI):

  - Utilizes technologies like machine learning, computer vision
  - Supports real-time recognition of ISL gestures and their conversion into text or speech.
  - Enables translation of text or speech back into ISL for two-way communication.

- Human-Computer Interaction (HCI):
  - Provides the framework for creating tools that enhance accessibility for the Deaf and Hard of Hearing (DHH) community.
  - Facilitates effective communication between Indian Sign Language (ISL) users and non-users.

1.2. Motivation

- Addressing Communication Barriers:
  - The Deaf and Hard of Hearing (DHH) community in India faces significant communication challenges due to the limited understanding of Indian Sign Language (ISL) among the general population.
  - This gap results in social isolation and restricted access to essential services, education, and employment opportunities.

- Promoting Inclusivity:

  The project aims to bridge this communication gap by developing a system that enables two-way communication between ISL users and non-users, fostering a more inclusive society.

- Improving Quality of Life:

  By breaking communication barriers, the project seeks to empower DHH individuals, enabling them to actively participate in:
  - Social interactions.
  - Educational environments.
  - Professional domains

2. SYSTEM LITERATURE REVIEW

REFERENCE	FOCUS	TECHNOLOGICAL APPROACH	UNIQUE FEATURES	RESEARCH GAP
Prof. Mrs. Maheshwari Chitampalli et al. (2023) [3]	Static Sign Language Recognition (SSLR)	1. Cleaning of dataset removing irrelevant information 2. Use computer vision techniques, such as background subtraction or skin color detection, to segment the gestures from the background. 3. Extract relevant features from segmented gestures 4. CNN is used to train the model on the extracted feature	Average accuracy of the model being used is about 95%	Accurately form complete sentences from continuous gesture (CSLR), lack of dataset of images having different skin tones and in different lighting condition
Deep Kothadiya et al. (2022) [4]	A deep learning-based model for detecting and	1. Combination models (e.g., LSTM-GRU) for enhanced recognition.	1. IISL2020 (customized dataset). The IISL2020	1. The model focuses on isolated signs, not continuou

	recognizing words in Indian Sign Language (ISL) from video frames	2. InceptionResNet-v2 (a convolutional neural network (CNN) that can classify images into 1,000 object categories) is used for feature extraction from video frames 3. Custom Dataset (IISL2020) is used to train the model.	dataset consists of 11 words; for each word, there are about 1100 video samples of 16 research participants, including males and females and was made without extra brightness, orientation, background adjustments, gloves, etc. 2. Combination of 2 layers Long short-term memory (LSTM) and gated recurrent unit (GRU) is used for	s sign recognition or sentence formation. 2. Dataset lacks variety of words and phrases to ensure broader generalization 3. Can produce instable result in varied light conditions.

			isolated signs, achieving a high accuracy rate on the custom dataset 3. Cross-Dataset Validation is done to demonstrate the robustness of the model	
Hope Orovwode et al.(2023) [5]	Real-time sign language recognition system using machine learning	Convolutional Neural Network (CNN) model with three convolutional layers and a SoftMax output layer to classify hand gestures13. The model was trained on a dataset of 44,654 images	1. This system is the utilization of the HandDetector module to detect and crop out the signer's hand from the camera's field of view4. This minimizes the effect of the user's background on the model	Only can be used for SSLR

			prediction 2. This system achieved a remarkable test accuracy of 94.68%	
Iftikhar Alam et al. (2024) [6]	1. Systematically reviews the use of smartphones for sign language detection and interpretation utilizing machine learning and deep learning approaches. 2. Explores the advancements in vision-based and sensor-based techniques	1. Machine Learning Techniques: SVMs, KNN and Random Forest 2. CNNs, LSTM and RNNs 3. Smartphone-based Applications: Utilizing smartphone cameras and processors for real-time detection and translation of gestures into text or speech.	1. Analyzes studies from 2012 to 2023, focusing on both machine learning and deep learning methods for sign language. recognition. 2. Role of smartphones as portable and affordable tools for accessibility solutions. Explores variations in datasets and approaches for region-specific sign languages.	1. Lack of standardized, universal sign language datasets or models to address regional variations. 2. Challenges in achieving low latency and high accuracy in real-time applications using smartphone hardware. Issues with lighting, skin tone variations, and complex backgrounds

				affecting vision-based recognition systems.
Sachin Tripathi et al.(2021)[7]	A sign language recognition mobile application named "Sanket" to bridge the communication gap between deaf-mute individuals and the general population	1. CNN used for recognizing hand gestures from images 2. Image Dataset: Contains 87,000 images across 29 classes (A-Z, space, delete, nothing) 3. Flutter: For building a user-friendly interface. 4. RESTful API to process images and return recognized text.	1. Mobile Application is used. 2. Offline Capability: Designed for offline use, ensuring accessibility without internet dependency. 3. Lightweight Architecture: Utilizes a CNN-based model with 11 layers, including dropout and batch normalization, optimized for efficiency. 4. User-Centric Design: Simplified for easy	1. Does not have real-time gesture recognition capabilities. 2. The focus is solely on ASL 3. Performance varies significantly with lighting conditions, requiring further optimization. 4. Restricted to isolated gestures, with no support for continuous gesture recognition or sentence formation.

			use by non-technical users.	
Rajesh B. Mapari and Dr. G.U. Kharat [8]	Focuses on creating a reliable and effective method for categorizing Indian Sign Language (ISL) alpha-numeric characters, with a particular emphasis on static signs created with one or two hands.	An RGB camera is used to take pictures of signers executing ISL signs against a black background. Discrete Cosine Transform (DCT) and regional attributes are utilized to generate a distinct feature vector for every sign. This feature vector is then used to train and evaluate the system's recognition of ISL characters using different neural network classifiers.	Unlike ASL and other sign languages, it focuses on the requirement for an ISL recognition system. To produce a complete feature vector, it uses DCT to merge frequency domain data with spatial features. When comparing the performance of three different neural network classifiers (MLP, GFFNN, and SVM), MLP produces the best accuracy.	Real-world application is limited by the system's use of a fixed black background and its inability to recognize dynamic indications or facial expressions. Even though MLP achieved 86.27% accuracy, certain signs (K, S, T, and Y) still require development in order to increase reliability.
Rosemary Antony, Stephy Paul, Silji Simon C, and Assistant Prof. Scaria Alex (2020) [9]	Translating sign language into voice for people who are unable to communicate. It attempts to close the gap in communication	Uses a webcam to collect hand motions from live footage. It then extracts hand features and guesses the appropriate symbol. Special languages are used to handle connecting words such as "is" and "are". The	Real-time sign language gestures are recorded and processed by the system. Anyone can use the system because it translates recognized signs into spoken speech. To take	An important issue with current sign language translation systems, according to the report, is their inability to correctly interpret connective terms ("is," "are," "and,"

	between people who use sign language and those who don't.	technology uses machine learning and image processing techniques to turn recognized signs into audible speech	connecting words into consideration, the algorithm uses unique linguistic properties.	etc.). The translated output becomes less grammatically correct and fluent as a result.
Nishchal Gowda H S, Preksha V L, Prathibha M, Prathiksha N V, and Dr. Surekha T P [10]	focuses on investigating different technical methods for translating sign language into audio and text forms. The importance of these methods in facilitating communication between the hearing community and deaf and mute people is emphasized in the study.	Examines machine learning, image processing, and sensor-based approaches for sign language identification. Web cams are used by systems to record motions, extract features, compare them to a database, and then turn them into text and voice.	In order to improve gesture recognition, the system uses depth cameras to translate sign language in real time. It is available via mobile apps, provides two-way communication, and concentrates on certain sign languages like ISL.	The limited range of depth cameras hinders their practical application. The accuracy of sign language systems varies, and performance is affected by things like noise and lighting. Certain systems' adaptability is limited by their need for regulated conditions.
V. Valarmathi, S. Sowmiya, and M. Viswanathan (2022)[11]	By addressing the under-utilization and lack of knowledge of sign language,	Python is the main technology used in the system. Using a machine-learning model based on Natural Language Processing (NLP), the system records	Bidirectional communication is made possible via EquiSign, which translates voice to sign language and vice versa. It	At the moment, EquiSign concentrates on ISL; additional data is needed to expand to other sign languages. It has trouble

	the article focuses on creating a dynamic sign language translator to promote communication between individuals with hearing impairments and the broader public, so promoting inclusion and inclusivity.	vocal input in real time and transforms it into text. After that, the text is converted into the designated sign language.	provides real-time translation and supports a number of sign languages. Speech input, speech-to-text conversion, sign production, gesture capture, display output, and data management are all included in the system to guarantee accuracy.	becoming accurate because of things like lighting and signing styles. Depth cameras have a limited range, which impacts their practical application, and many systems require controlled conditions, which restricts their adaptability.
Aaditya. C. Punekar, Saranya. A. Nambiar, Raunak Singh Laliya, and Prof. Saumya. R. Salian (2020) [12]	In order to improve communication between the deaf/mute community and hearing people in India, the paper highlights the creation of a system that translates Indian Sign	It combine image processing and machine learning in a vision-based methodology. The device uses a CNN to identify motions, records hand gestures in real time, and extracts data using methods like Gaussian Blur and Contour Detection. Then, using the pyttsx3 package for audio output, the identified gesture is transformed into speech.	It is more accessible and reasonably priced because it only requires a laptop and camera to function, eliminating the need for additional hardware. Additionally, it offers speech and text outputs to satisfy different user preferences.	Vocabulary has to be expanded for practical use because it now only includes a few words and fundamental alphabets. Although it attains 88% accuracy for words and 96% accuracy for alphabets, further work is required to make it more resilient, particularly under different

	Language (ISL) into text and speech.			**circumstances. More research is also required to address the difficulties of managing regional ISL variants and real-time translation in order to guarantee smooth communication across various populations.**
Harsh Kumar Vashisth et al. (2023) [13]	**Hand Gesture Recognition in Indian Sign Language Using Deep Learning**	**Convolutional Neural Networks (CNN) on a self-constructed dataset of 7800 images, employing HSV color space for enhanced hand segmentation and training on multi-layer CNN.**	**1. Achieved 99% accuracy on static hand gestures in ISL.** **2. The use of HSV color space optimized gesture detection.** **3. The model is designed for real-world application settings.**	**1. Limited to static gestures; does not account for dynamic gestures essential for certain letters in ISL.** **2. limited vocabulary beyond alphabets; (A – Z)**
L. Latha, M. Kaviya (2019) [14]	**Real-Time System for Two-Way**	**1. Raspberry Pi, camera, and microphone for**	**1. modular design with Raspberry Pi,**	**+**

	Communication for Hearing and Speech Impaired People	**gesture and speech input;** **2. pre-processing using background subtraction;** **3. voice generation for gestures and gestures for speech** **4. Processes input images and speech for gesture-to-speech and speech-to-gesture translation using Local Binary Patterns and SVM.**	**camera, and microphone for gesture and speech input;** **2. Portable hardware-based system facilitating two-way communication, including gesture-to-speech and speech-to-gesture conversion.** **3. Two-way communication system.**	
- V. Valarmathi (2022) [15]	**Enhancing Inclusion Using a Dynamic Sign Language Translation System**	**1. Speech-to-Sign Conversion:** • **Natural Language Processing (NLP) techniques (e.g., tokenization, stopword removal).** • **Conversion of text to ISL gestures displayed as GIFs or animations.**	**Multi-Modular Architecture: Includes speech processing, text processing, gesture recognition, and display modules.** **Dynamic Output Display: Outputs ISL gestures either as 3D animations or pre-recorded videos/GIFs.**	

		2. Sign-to-Text Conversion: • **Image processing using techniques like contour detection and Haar cascade.** 1. **Classification using pre-trained Convolutional Neural Networks (CNNs).**		
Rosemarry Antony et al. (2020) [16]	**Sign Language to Voice Translation for Dumb People**	**Webcam-based system capturing real-time gestures, using Principal Component Analysis (PCA) for feature extraction, and converting signs to audible voice output with text-to-speech technology.**	2. **Real-Time Translation** 3. **The final output is converted into speech, making it audible to everyone and more accessible.** 4. **Special provisions for handling connecting words like "is"**	**The use of PCA for feature extraction may not achieve the precision provided by modern deep learning models like Convolutional Neural Networks (CNNs).**

			and "are" are included, which are often overlooked in existing systems. 5. designed for use in public places like ticket counters and hospitals and has potential applications in sign language teaching.	
Elakkiya R et al. (2012) [17]	Intelligent System for Human Computer Interface Using Hand Gesture Recognition	1. Image Processing Techniques: • Uses coloured markers on fingers for gesture identification. • Implements Hue Histogram and Projection Techniques for color-based	1. For the low-level hand-posture detection, it uses a statistical approach based on Haar-like features to train. 2. Pseudo-Two-Dimensional Hidden	• The system's dependency on coloured markers limits its applicability in dynamic lighting conditions or when markers are unavailable.

		segmentation. 2. Gesture Recognition: • Tracks hand motion using frame differencing, grayscale conversion, and erosion/dilation techniques. • Analyzes trajectories using Circular Fuzzy Neural Networks (FIM) for decision-making.	Markov Models (P2-DHMMs) are used for the hand gesture recognition. The fuzzy neural network architecture incorporates the idea of fuzzy Adaptive Resonance Theory Mapping (ARTMAP) in hand gesture recognition neural networks.	• Noise handling and tracking accuracy could be improved, especially in scenarios with occlusions or overlapping objects. • The system lacks modern machine learning techniques (e.g., deep learning models) that could improve generalization and adaptability to diverse hand shapes and movements.
Sharvani Srivastava et al. (2021) [18]	Development of a real-time Indian	• TensorFlow Object Detection	• Real-time detection capability	• Small dataset with

	Sign Language (ISL) recognition system using machine learning.	API with transfer learning • SSD MobileNet v2 as the pre-trained model.	. • Low-cost data acquisition using webcams and Python OpenCV. • Binary files usage for storage of the data.	limited variability. • Vision-based inaccuracies due to overlapping of hands and fingers. • Lack of scalability for continuous sign language recognition.
S.G. Mundada, K. Khurana, A. Bagora (2019) [19]	Automating gesture recognition and converting gestures to speech using vision-based methods.	K-means clustering and thresholding for background elimination, convex hull and proposed slope-based peak detection algorithm for fingertip identification, centroid trajectory for motion tracking, and text-to-speech (TTS) APIs.	Introduces a two-phased approach: static and dynamic gesture conversion; real-time centroid trajectory tracking to detect movement directions and gestures; uses slope-based detection for precise fingertip and motion recognition.	Works only for single-hand gestures. Dataset size is limited, leading to constrained gesture variety. Lack of incorporation of machine learning or AI models limits adaptability for larger and more diverse real-time datasets.
Beena M.V., Dr. M.N. Agnisarman Namboodiri (2017) [20]	Development of a static gesture recognition	Implemented a lightweight Convolutional Neural Network (CNN) using PDNN	High accuracy of 94.6774% achieved through depth images and	Limited to static gestures (excludes dynamic gestures like "J"

	system for American Sign Language (ASL) alphabets and numbers.	in Theano with GPU acceleration; trained on a dataset of 33,000 depth images (28x28 grayscale); created a Java GUI application for real-time testing and speech translation.	GPU-accelerated CNNs; optimized for embedded devices with limited computational resources; included a user-friendly GUI application to convert gestures into speech output.	and "Z"); dataset collected only from five subjects, restricting diversity; does not integrate multimodal data (e.g., facial expressions, hand motion trajectories) for comprehensive recognition.
V. Adithya and R. Rajesh (2020) [21]	Creation of a video dataset for recognizing Indian Sign Language (ISL) gestures used in emergencies.	Video dataset of 824 samples of eight ISL gestures captured using a Sony CyberShot camera; Pre-processing included cropping and resizing; Classification using Multiclass SVM and Deep Learning (GoogleNet+LSTM).	Focused on emergency-related ISL gestures; included dynamic and static gestures; dataset captured from diverse subjects with varying skin tones; provides a benchmark for ISL-based emergency systems; achieved up to 96.25% accuracy using CNN-LSTM.	Dataset is limited to only eight ISL emergency gestures, restricting its scope; lacks consideration of diverse backgrounds, complex lighting conditions, or real-world noisy environments; additional multimodal features (e.g., facial expressions) not included.
Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, Indra R. (2020) [22]	Conversion of real-time sign language to text and speech through an Android application.	Utilized Histogram of Oriented Gradients (HOG) descriptors for feature extraction, Support Vector Machine (SVM) with different kernels (RBF,	Android-based application capable of real-time translation of 26 ASL alphabets plus a custom 'Space' gesture; designed for	Dataset reduced to 100 images per character, limiting generalization; focuses only on static gestures; lacks dynamic gesture

		Linear) and PCA for dimensionality reduction; implemented in Java with OpenCV libraries.	smartphones to ensure accessibility and affordability; achieves up to 98.82% accuracy with optimized parameters.	processing and integration with multimodal features (e.g., facial expressions, contextual gestures) for complete communication.
Adithya V, Rajesh R. [23]	A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition	Deep learning with CNN, detection and segmentation of hands from image (, three convolution layers along with ReLu and max pooling layers for feature extraction) 1. Input: RGB images of hand posture	Achieved Accuracy (99%),	Focuses only on static hand gesture and lacks dynamic gesture processing and integration with multimodal features (e.g., facial expressions, contextual gestures) for complete communication
Akshit Tayade Arpita Halder [24]	Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning	1. Pre-Processing of Images to get Multi-hand Landmarks using MediaPipe 2. Data cleaning and normalization 3. Prediction using Machine Learning Algorithm and Support	Achieved accuracy (99%), Requirement of less computational power and lightweight	lacks dynamic gesture processing and integration with multimodal features (e.g., facial expressions, contextual gestures) for complete communication

		Vector Machine (SVM) Quantitative Analysis		
Romala Sri Lakshmi Murali, L.D.Ramayya, V. Anil Santosh [25]	Sign Language Recognition System Using Convolutional Neural Network and Computer Vision	1. HSV colour space and background elimination 2. Segmentation Applied 2D CNN model with a tensor flow library	Achieved accuracy of 90%, requirement of low computing power	Sign languages are very broad and differ from country to country in terms of gestures, body language and face expressions. The grammars and structure of a sentence also varies a lot. Some gestures are difficult to reproduce. And it was hard to keep our hands in exact same position when creating our dataset.
Shagun Katoch a, Varsha Singh b, Uma Shanker Tiwary [26]	Indian Sign Language recognition system using SURF with SVM and CNN	1. Dataset collection (manually constructed) 2. Pre-processing: converted into HSV colour space 3. Feature Extraction: n building a bag of visual words	Reverse recognition (text to sign language); Achieved higher accuracy (99%) Real time recognition; Enhanced real-time response	Worked only on Indian Sign Language;

		(BOVW) is to extract descriptors from each image in the dataset. SURF (Speeded Up Robust Features) [27] is used which is a local feature detector and descriptor. 4. Classification: passed the histograms of visual words to the SVM as feature vectors for the classification and recognition of ISL signs. CNNs compare images piece by piece where a filter map slides over the local patches of the image.		
Meenakshi Panwar, Pawan Singh Mehran [27]	Hand Gesture Recognition for Human Computer Interaction	1. Image Enhancement and segmentation (RGB to YCbCr and nose elimination) 2. Orientation detection:	Human computer interaction, Achieved higher accuracy (99%)	Accuracy drastically low if hand gesture is presented diagonally,

		Edge detection and detect either vertical or horizontal 3. Feature Extraction: A. Finding centroid B. Peaks or finger region detection C. Thumbs detection classification		
M.Jerin Jose, V.Priyadharshni, M.Suresh Anand, A.Kumaresan, Dr.N.MohanKumar (2013) [28]	Indian Sign Language translation system for sign language learning	1. Domains include speech to sign translation where voice data is captured. 2. Voice data filtration using spectral subtraction method, weiner filter, least mean square filter and kalman filter. 3. Speech recognition using Hdden Markov model with database integration. 4. Rule based text to sign	1.Usage of customized NLP techniques for speech to text translation. 2.Usage of noise removal techniques to prevent external irrelevant data interference or data loss. 3.Usage of speech to sign translation by sign database. 4.Sign to speech by combining vectorization of hand and face gestures and speech database.	Multilingual Indian languages is a problem for voice recognition and speech to sign/text translation. Database has an extensive use, no usage of cloud/lightweight models. (heavier application) Speech recognition and translation using intelligent NLP is not in use (for understanding and checking the grammatical errors).

		matching using natural language translation(with error rates- SER, PER, BLEU and NIST) 5. Weighted sign confidence measure using a value between 0.0 to 1.0.		
Ashok Kumar Sahoo1, Gouri sankar Mishra2 & Pervez Ahmed (2012) [29]	A proposed system for ISL recognition and translation to English language using Interlingua - based architecture. System is based for education of hard of hearing students in India.	1. Data collection from different deaf schools, to track the pattern and mapping of objects to signs. 2. Feature extraction process includes statistical methods for training the ML models. 3. Testing of data and validation for tracking system performance.	Usage of several methodologies for pattern recognition. These include Template matching, statistical method, structural method, syntactic method, and artificial neural network.	Proposed system only, there is no proper mention of machine learning and ANN models. Usage of database, complex stats feature extractions, and ML models leads to heavier system.
Poornima B.V., Srinath S. (2023) [30]	A comprehensive review on ISL recognition	1. SLR process involves video frame capturing, preprocessin	A review paper of all the recent papers on sign language recognition and	Technical specifications/ methodologies are not mentioned with

	system using vision-based approaches	g, segmentation, feature extraction, & classification. 2.Usage of KNN algorithm for hand pose recognition, CNN for feature extraction of ISL, RNN for temporal relevant data training, LSTM for dynamic gesture & various other deep learning models.	translation, containing all the current approaches of various papers and systems.	comprehensive studies to know about the approaches in technologies being used.
Kumud Tripathi, Neha Baranwal and G. C. Nandi [31]	Gesture Recognition and Sentence Formation	1.Video frame capture, feature extraction, frame preprocessing, key frame extraction by gradient method, organisation histogram, recognition leads to output as text/audio. 2.Models include SVM and HMM.	Principal component analysis for data patterns. Classification of data includes methodologies like Euclidean distance and Mahalanobis distance, city block distance, chess board distance, cosine & correlation distance.	Comprehensive analysis of feature extraction but classification and mapping for sign to text translation is missing.

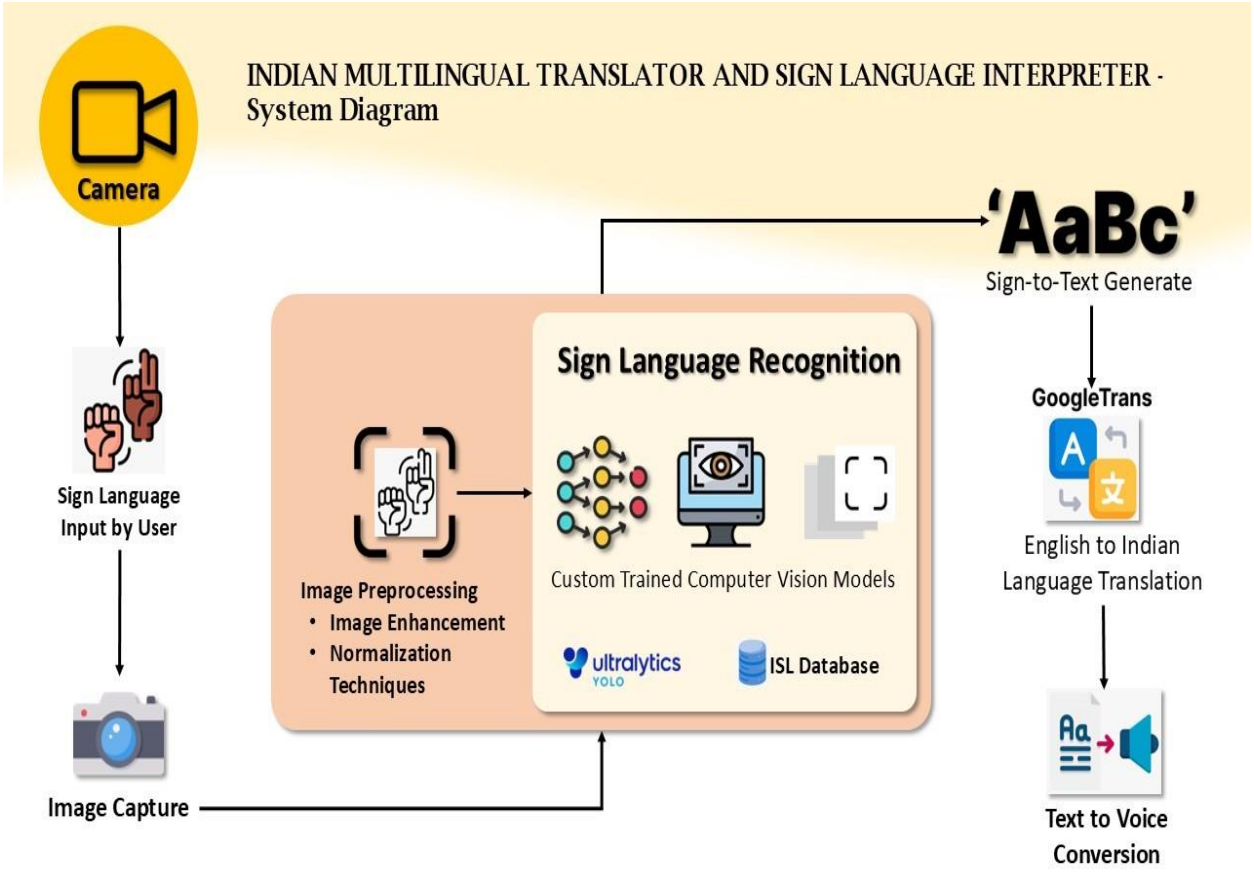
3. PROPOSED WORK

4. EXPERIMENTS AND ANALYSIS

5. CONCLUSION

The "Indian Multilingual Translator and Sign Language Interpreter" project leverages advanced deep learning techniques, including YOLOv11, to bridge communication gaps for the hearing-impaired community. The system offers real-time gesture recognition and translation, with future scope to incorporate text and speech conversion into animated sign language for enhanced interactivity. It holds significant potential in education, healthcare, public services, and corporate accessibility. Revenue can be generated through SaaS models, licensing, and subscriptions. By combining societal impact with commercial scalability, this solution can empower millions while fostering inclusivity and accessibility in diverse real-world scenarios.

6. ANNEXURE:



6.1 System Overview

The project implements the "Indian Multilingual Translator and Sign Language Interpreter" using a modular design. The attached **System Diagram** explains the flow:

1. **Input:**

a. Users provide **real-time ISL gestures** via a camera or upload gesture images.

2. **Image Capture and Preprocessing:**
   a. Captured images undergo preprocessing, including **image enhancement** (to handle varying lighting conditions) and **normalization** (to standardize dimensions and pixel intensity).
3. **Image Processing and Recognition:**
   a. **Custom-trained YOLOv11 models** are employed for gesture recognition.
   b. Models are trained on a diverse ISL dataset and fine-tuned for Indian-specific gestures.
   c. The recognized gesture is translated into text through the model's prediction output.
4. **Multilingual Translation:**
   a. Using **GoogleTrans API**, the recognized text is translated into multiple Indian languages for accessibility.
5. **Text-to-Speech Conversion:**
   a. Translated text is converted to voice output using Text-to-Speech (TTS) services such as **PlaSound** or equivalent libraries, enabling communication for non-sign language users.

6.2 Dataset Details

The dataset used for this project was sourced from **Roboflow**, a platform offering high-quality annotated datasets optimized for computer vision tasks. The dataset contains Indian Sign Language (ISL) gestures, with each image annotated in the YOLO format, including class IDs and normalized bounding box coordinates. To improve robustness and generalization, data augmentation techniques such as rotation, scaling, and brightness adjustments were applied. The dataset annotations were reviewed and verified to ensure accuracy and consistency. The model training was conducted on **Kaggle**, utilizing its high-performance GPU environment to efficiently process the data and achieve optimal results.

6.3 Technological Advancements

- Utilization of **YOLOv11** for fast and accurate real-time gesture detection.
- Deployment-ready, lightweight architecture optimized for edge devices, ensuring low-cost accessibility.

6.4 Future Scope

Ensemble Approach:

- Combine YOLOv11 with CNNs, RNNs, or LSTMs for enhanced static gesture recognition by leveraging complementary strengths of different deep learning architectures.

Dataset Expansion:

- Develop diverse datasets reflecting real-world conditions, including variations in lighting, backgrounds, and regional ISL dialects, to improve robustness and generalization.

Multilingual Capabilities:

- Enhance support for more Indian languages to cater to the diverse linguistic needs of users.

User-Friendly Interfaces:

- Create intuitive and accessible interfaces to make the system usable by non-technical users, fostering inclusivity.

Dynamic Gesture Recognition:

- Incorporate the ability to interpret sequences of gestures for recognizing dynamic sign language.

Integration of Generative AI (GenAI):

- Use GenAI to synthesize diverse datasets, generate advanced gesture features, and enable real-time contextual translation for seamless communication.

Text and Speech to Animated Sign Language:

- Add functionality to translate text and speech into animated sign language, creating a fully bidirectional communication system for greater inclusivity.

7. REFERENCES

[1] The Constitution of India [https://legislative.gov.in/constitution-of-india/]
[2] Indian Sign Language Research and Training Centre [https://islrtc.nic.in/]
[3] Prof. Mrs. Maheshwari Chitampalli, Dnyaneshwari Takalkar, Gaytri Pillai, Pradnya Gaykar, Sanya Khubchandani (2023) [REAL TIME SIGN LANGUAGE DETECTION]

[4] Deep Kothadiya ,Chintan Bhatt ,Krenil Sapariya ,Kevin Patel ,Ana-Belén Gil-González and Juan M. Corchado (2022) [https://doi.org/10.3390/electronics11111780]

[5] Hope Orovwode, Ibukun Deborah Oduntan, John Abubakar(2023) [https://doi.org/10.1109/icABCD59051.2023.10220456]

[6] Iftikhar Alam, Abdul Hameed and Riaz Ahmad Ziar(2024) [https://doi.org/10.1155/2024/1487500]

[7] Sachin Tripathi , Priyanka Verma , Lakshay Malhotra , Priya Verma , Dr. Preety Verma Dhaka (2021) [SIGN LANGUAGE RECOGNITION APPLICATION-SANKET]

[8]Rajesh B. Mapari, Dr. G.U.Kharat [https://www.academia.edu/28423478/Indian_Sign_Language_Alpha_Numeric_Character_Classification_using_Neural_Network]

[9] Rosemarry Antony, Stephy Paul, Silji Simon C, Assistant Prof. Scaria Alex (2020) [https://www.ijert.org/sign-language-converter-for-deaf-and-dumb-people-in-two-way-communication-for-regional-languages]

[10] Nishchal Gowda H S, Preksha V L, Prathibha M, Prathiksha N V [http://ijariie.com/AdminUploadPdf/A_Review_on_Sign_Language_Conversion_Into_Text_and_Audio_ijariie16512.pdf?srsltid=AfmBOor5bm1Hf9-i9G9uJ6nl00P8-UhJqB6iIGhHRef8uvpM8ITG7shO]

[11]V.Valarmathi,S.Sowmiya,Viswanathan,M.Viswanathan [https://www.technoarete.org/common_abstract/pdf/IJERCSE/v9/i7/Ext_34810.pdf]

[12] Aaditya. C. Punekar1, Saranya. A. Nambiar2, Raunak Singh Laliya3, Prof. Saumya. R. Salian

[https://www.researchgate.net/publication/342565986_A_Translator_for_Indian_Sign_Language_to_Text_and_Speech]

[13] Vashisth, H.K.; Tarafder, T.;Aziz, R.; Arora, M.; A. Hand Gesture Recognition in Indian Sign Language; Using Deep Learning. Eng. Proc. 2023,59, 96. [https://doi.org/10.3390/engproc2023059096]

[14] L. LATHA, M. KAVIYA; International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-7 Issue-4s2, December 2019

[15] V. Valarmathi,S. Sowmiya, M.Viswanathan; EquiSign Dynamic Sign Language Translator;

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 9, Issue 7, July 2022

[16] Rosemarry Antony, Stephy Paul, Silji Simon C, Assistant Prof. Scaria Alex; Sign Language Translation to Voice For Dumb People

[International Journal of Scientific Research & Engineering Trends, Volume 6, Issue 2, Mar-Apr-2020, ISSN (Online): 2395-566X]

[17] Elakkiya R, Selvamani K, Kanimozhi S, Velumadhava Rao, Kannan A; Intelligent System for Human Computer Interface UsingHand Gesture Recognition [https://www.researchgate.net/publication/271617331]

[18] Sharvani Srivastava, Amisha Gangwar, Richa Mishra, Sudhakar Singh. (2021). *Sign Language Recognition System using TensorFlow Object Detection API*. Proceedings of the International Conference on Advanced Network Technologies and Intelligent Computing (ANTIC-2021), part of the book series 'Communications in Computer and Information Science (CCIS)', Springer. https://doi.org/10.1007/978-3-030-96040-7_48.

[19] S.G. Mundada, K. Khurana, A. Bagora. (2019). *Real Time Conversion of Hand Gestures to Speech using Vision Based Technique*. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(9), 3184–3190. https://doi.org/10.35940/ijitee.I8748.078919.

[20] Beena M.V., & Agnisarman Namboodiri, M.N. (2017). *Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis*. International Journal of Pure and Applied Mathematics, 117(20), 9–15. http://www.ijpam.eu.

[21] V. Adithya, & R. Rajesh. (2020). *Hand gestures for emergency situations: A video dataset based on words from Indian sign language*. https://doi.org/10.1016/j.dib.2020.106016.

[22] Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, & Indra R. (2020). *Real-time Conversion of Sign Language to Text and Speech*. (ICIRCA-2020), IEEE, 346–351. https://doi.org/10.1109/ICIRCA48905.2020.9182877 .

[23] Adithya V, Rajesh R (2020). A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition. Third International Conference on Computing and Network Communications (CoCoNet'19), ELSEVIER, Procedia Computer Science 171 (2020) 2353–2361

[24] Akshit Tayade, Arpita Halder (2021). Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. International Journal of Research Publication and Reviews Vol (2) Issue (5) (2021) Page 9-17. https://www.researchgate.net/publication/369945035

[25] Romala Sri Lakshmi Murali, L.D. Ramayya, V. Anil Santosh (2022). Sign Language Recognition System Using Convolutional Neural Network And Computer Vision. International Journal of Engineering Innovations in Advanced Technology ISSN: 2582-1431 (Online), Volume-4 Issue-4, December 2022

[26] Shagun Katoch, Varsha Singh, Uma Shanker Tiwary (2022). Indian Sign Language recognition system using SURF with SVM and CNN. ELSEVIER, Array 14 (2022) 100141.

[27] Meenakshi Panwar, Pawan Singh Mehran (2011). Hand Gesture Recognition for Human Computer Interaction. Hand gesture recognition for human computer interaction | IEEE Conference Publication | IEEE Xplore

[28] M.Jerin Jose, V.Priyadharshni, M.Suresh Anand, A.Kumaresan, Dr.N.MohanKumar

(2013). Indian Sign Language translation system for sign language learning. https://www.academia.edu/67849732/Indian_Sign_Language_ISL_Translation_System_For_Sign_Language_Learning

[29] Ashok Kumar Sahoo1, Gouri sankar Mishra2 & Pervez Ahmed (2012). A proposed framework for Indian Sign Language Recognition. https://www.researchgate.net/publication/263506784_A_proposed_framework_for_India

n_Sign_Language_Recognition

[30] Poornima B.V. , Srinath S.(2023) A comprehensive review on ISL recognition system using vision-based approaches. https://www.ijcaonline.org/archives/volume184/number43/32601-2023922548/

[31] Kumud Tripathi, Neha Baranwal and G. C. Nandi. Continuous Indian Sign Language Gesture Recognition and Sentence Formation(2015). https://www.sciencedirect.com/science/article/pii/S1877050915013848?via%3Dihub