Video Communication Platform For Specially Abled People

Riya Tyagi
Students of Department of CSE
Sharda School of Engineering and Technology
Greater Noida, Knowledge Park-3, U.P., India
line 5: email address or ORCID

Ashutosh Bhardwaj
Students of Department of CSE
Sharda School of Engineering and Technology
Greater Noida, Knowledge Park-3, U.P., India
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 6th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

Abstract**—*****The "Video Communication Platform for Specially Abled People" cross-platform application will enable those suffering from speech and hearing impairments to communicate more effortlessly than ever before. The app converts the hand gestures of Indian Sign Language into text and speech, making daily conversations as well as virtual meetings/virtual video calls much more accessible. Using deep learning with PyTorch, the app recognizes gestures in real time and integrates voice-to-text and text-to-speech features to support two-way communication. This is to bridge the communication gap so that specially abled people can participate more effectively in society.***

Keywords: Video Communication Platform, Specially Abled People, Hearing Disabilities, Speech Disabilities, Indian Sign Language (ISL), Cross-Platform App, Gesture Recognition, Deep Learning, PyTorch, Real-Time Translation, Voice-to-Text, Text-to-Speech

I. Introduction

New research now says that more than 430 million people suffer from much more disabling and debilitating forms of hearing disorders. Of course, more than 1.5 billion worldwide suffer some type of hearing loss, as new statistics now indicate. Millions of them also suffer from speech disorders, which severely limit effective communication. The World Health Organisation says that over 700 million may need hearing restoration treatments by 2050. From studies it is found that approximately 63 million people in India are suffering from speech and hearing deficiencies. The larger number of them use Indian Sign Language as their primary language of communication. Though technical innovations have brought so many opportunities in today's world, communication still continues to lag behind between the integrated and non-integrated population, which includes a person with either speech or hearing deficiency. With more than 300 different sign languages and most of the people not being aware of the sign language, it creates a complicated and inaccessible conversation. In the time of the breakout of the COVID-19 epidemic, video chats and online meetings have been the mainstream, and this isolation only experienced during those times. This implies that the speech and hearing disabled individuals are usually excluded from the significant lives of social, education, or professionalism. Our project shall fill this gap by developing a video communication Android system which interchanges hand gestures into text and vice versa-for effective communication between the hearing population and that of hearing/speech impaired. Advanced machine learning models, image recognition software, and voice-to-text technologies will be the base of this system incorporated into the presentation of easily accessible, instantaneous communication solutions.

We implement this using the incredibly powerful open-source deep learning library, PyTorch. This is suitable for developing and training the neural networks required in image recognition tasks since it provides fast prototyping, adaptive model development, and efficient usage of the GPU. The developed and trained model based on YOLOv8 using PyTorch has been used for hand motion identification from the ISL dataset. The technique preserves the accuracy of good gesture detection while decreasing complexity in computing. In this case, image dimensions are standardized and color images to gray-scale.

Image recognition is done using state-of-the-art object detection models, which include YOLOv5, YOLOv8, and Faster RCNN. These models are highly popular because of the fact that they provide real-time object detection and classification in images. YOLOv8 is particularly notable for its utmost speed and accuracy with which this is accomplished, making it a perfect for applications of real-time processing like recognition of hand gestures during live video chats. The algorithm ensures maximum accuracy as well as efficiency related to the hand movement by locating them inside an image and categorizing the instance accordingly with the help of ISL dataset. These devices translate hand gestures to voice for the mute and conversely translate speaking language into text or speech for the hearing impaired, thus allowing two-way communication. Whether this two-way conversation is

conducted using text messages or video conferencing, such real-time translation creates more inclusive discussions.
Our platform will have integrations of image recognition, deep learning, and natural language processing for an end-to-end communication solution to be applied in ordinary social relationships, work, and education.

Communication for impaired persons shall become as accessible as that to people with no impairment by real-time gesture-to-speech conversion. This is hoped to target an all-inclusive approach by empowering more people with speech and hearing impairments to engage in fairer competition with their louder counterparts in society.

II. Prior Work

Recently, initiatives have been taken to develop communication technologies in favor of the hearing-impaired community. More than that, much has been done for ISL, the Indian Sign Language. A video calling application exclusively developed for sign language users will present a singular space that might be efficiently utilized for immediate information exchange among users, with greater possibilities of inclusion. Such applications also enable carrying out conversations while at the same time creating awareness of, and knowledge about ISL in a larger population [1]. Research has shown that the implementation of visual communication tools enabled users with an avenue to better communicate themselves [2]. The proposed app is supposed to be an effective communication bridge which allows one-way and two-way interaction between deaf and mute people with add-on features of speech-to-text and text-to-speech that support accessibility for various abilities.[3]

Moreover, applications with regional sign languages, such as ISL, are also essential to achieve communication effectively. With localized features that help in accommodating the linguistic and cultural functionalities of ISL, it becomes much easier for users to interact and engage [4]. This will allow the users to personalize hand gestures easily that make communication much easier and provide a higher accuracy rate up to a near figure of 98% if sophisticated algorithms are used [5]. This kind of personalization not only facilitates the user but also promotes ISL usage in the society [6]. As the application will cater to the cultural norms of the users, it will provide a sense of belonging and promote active engagement in communication.

To make the video call application in sign language very effective, dynamic bidirectional translation systems should be incorporated. Recent experiments suggest that the use of ML-injected approaches greatly improves both accuracy and efficiency in translation [7]. Approaches such as YOLOv8, Faster R-CNN, and YOLOv5 can perform very accurate hand gesture classifications. For instance, YOLOv8 will reach an average IoU of 0.941. hence making it very efficient for real-time applications[8]. These methods classify the hand sign very fast and accurately, therefore not causing any delay in conversation between users.

Real-time processing allows for seamless communication between users, and technology acts as an extension of themselves in real life. Real-time is crucial where the information to be exchanged requires urgency-for instance, during consultations, or even in academic environments [9]. The ability to transmit thoughts and emotions in real-time significantly improves the experience by the user and thus makes more people use the app.

Moreover, using hybrid models for instance, the Enhanced InceptionNet architecture for isolated sign language recognition will allow fine-tuning of the application. It has been deduced by research that such architectures enhance isolated sign recognition rates, which further result in better accuracy in communication, and thus improves the quality of communication [10].Since capturing and interpreting a wide span of signs with high precision is ensured, timely feedback is provided that is necessary for effective communication in dynamic environments [11]. Along with facilitating user interaction, enhancement in identification acts as a stepping stone for adoption of sign language as the preferred channel of communication.

Methodologically, faster processing time is required. Innovative methods generally make video processing easier and hence facilitate quick translation of sign language into simple English sentences. The architecture of the application takes advantage of optimized object detection algorithms for minimal latency. Thus, YOLOv8 processing time could be as low as 0.0091 seconds, which is necessary in keeping a flow of conversation . Therefore, it becomes highly valuable for emergency or healthcare purposes because clear communication may influence outcomes in that time [12]. Minimizing the delay of processing helps the developers ensure an application that would be more responsive, thus increasing user satisfaction and accessibility .

Add to that the interface and functionality should be user-friendly and adopted for different needs of users. Providing visual cues, for instance, emojis or pictograms will add an additional advantage towards straightforward communication between users who may not be fluently signed [13]. This flexibility means that an application can easily be inclusive for all users since it caters to each user's preference of personal communication.

The designing of sign language recognition and translation models manages to bridge-building of communication barriers between the hearing-impaired[14]. Advanced technologies can be used for identification and translation with high accuracy. Such translation systems build an inclusive environment, allowing individuals to communicate freely in their preferred language [15]. The application can support many users who are normal, deaf, or mute, thus ensuring that communication is uninterrupted across different skills and preferences [16]. It is considerably important in various contexts such as education and public services where effective communication leads to better support and outcomes for hearing-impaired individuals [17].

Such applications will need multi-modal data sources to enhance the correctness of sign language detection. As such applications increase recognition features by leveraging visual, textual, and auditory information [18], it makes systems adaptive to multiple users' signing styles and preference, thereby making the system robust over time and enhancing algorithms with user feedback [19]. Such applications will accommodate the diversity range of sign language users and result in more accurate personalized interactions.

Despite recent strides in sign language recognition technology, several limitations persist. Notable among them is the fact that well-annotated high-quality datasets are needed to adequately train machine learning models [20]. In the absence of such data, biases could develop in the recognition systems which will generally lead to degradation in performance across various populations and geographies [21]. The inherent complexity of sign language - something that varies from one regional dialect to another, and even as presented by an individual - adds further challenges to the establishment of accurate recognition systems. To overcome these shortcomings, it is necessary that there must be a continuous and systematic effort toward constant improvement of the methods of training and evaluation of the model. In this way, systems for recognition of sign language can become stronger.

III. Proposed Methodology

The proposed application serves as a communication bridge, which helps in two-way interaction between individuals who are deaf or mute. In this context, a normal person refers to one who has no hearing or vocal impairment. The Key points of proposed application are:

Normal and Mute Communication : This feature will allow someone who speaks to converse with one who cannot. The speech of a talking person is converted to text and voice message that a mute person can read and understand easily. In addition, when a mute uses hand gestures, the platform converts them into text and speech so the speaking person can understand too. It's basically having a real-time translator.

Normal and Deaf Communication : This platform can translate what the hearing and talking person says into text, which a deaf person can read. The deaf person can respond to him through hand gestures. The application will translate what the deaf person is saying in hand gestures to text or speech for the hearing person. This way, they can communicate easily even though the other person is deaf and cannot hear.

Deaf and Mute Communication : This part deals with making it easier for those who cannot hear or speak to communicate back and forth with each other. They may use hand gestures and the app can translate this into text talk. This makes it really easy for them to have a conversation without needing to type or write out their words.

Normal to Normal Communication : Furthermore, it functions similarly to any other common communication application for those with the ability to hear and speak. It accommodates both normal voice and text messaging and thus is easy to use for anyone.

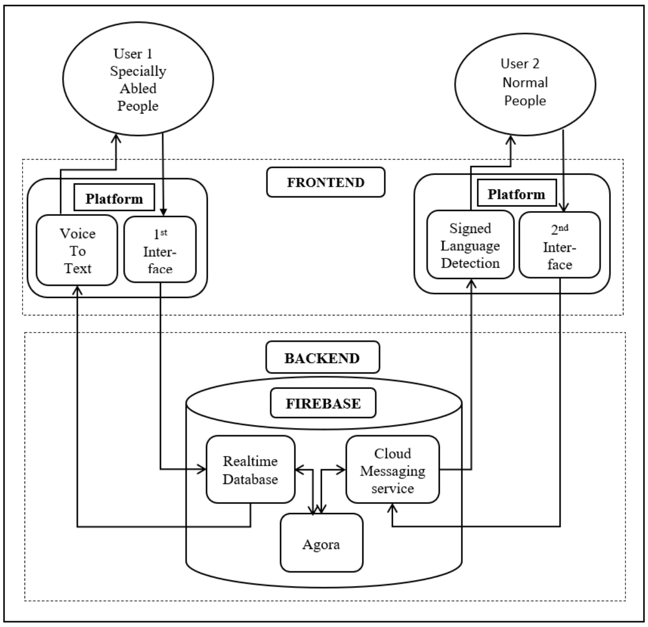

Fig 1. System Flow Diagram

Each feature is designed to make communication between people with different abilities feel natural and smooth. The app makes it easy for everyone to connect, understand each other, and chat without barriers!

System Architecture

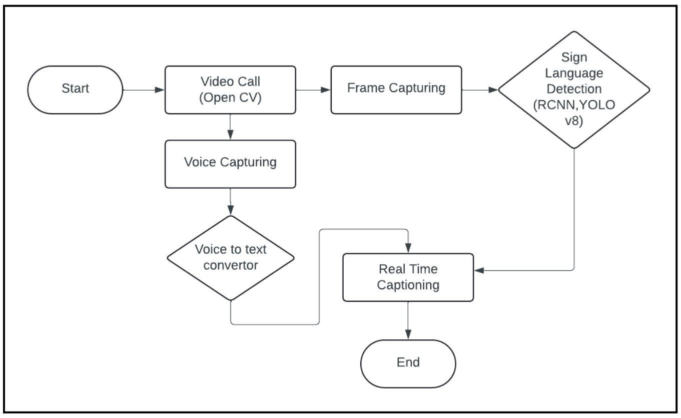

Fig 2. System Architecture Diagram

Dataset Preparation

Indian Sign Language is a dataset that can be experimented with on ideas about multi-class classification based on the technologies you are comfortable with. Curated for CNN, the multi-class classification result will appear to be close to 98% in its accuracy if it has a good algorithm.

ISLDataset-
https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl

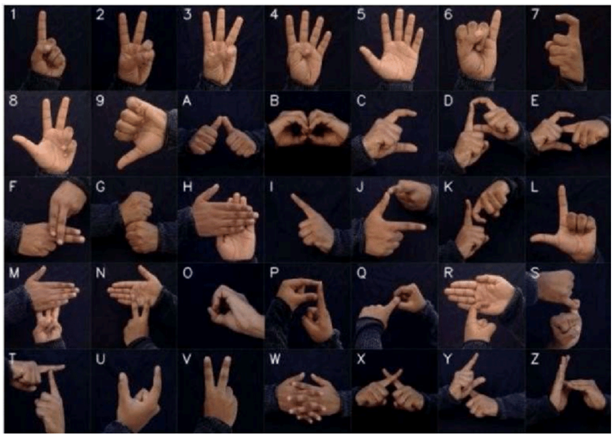

Fig 3. Sample Dataset[22]

A.Pre-processing

To reduce calculation complexity, convert the color photos into grayscale Image standardization: In CNN, it's critical to scale all of the images in your dataset to the same size.

B.Training and Testing Datasets

Training data is prepared using the labeled Indian Sign Language dataset. These models are designed and trained to identify hand gestures.. They identify photos using methods like YOLOv8, Faster RCNN, and YOLOv5.The accuracy and speed efficiency of these models are then evaluated by comparing them to a certain test dataset. Mean Average Precision (mAP), Intersection over Union (IoU), and testing time processing are a few of the most often used metrics.

C. Object Detection:

Object detection is all about determining where an object is in a picture, while object classification is related to the determination of what the object is. In this project, we are using object detection to identify hand gestures and classify them according to ISL.

YOLO v5: This incredibly quick method can identify several things in a single picture. It divides a picture into a grid, which resembles a chessboard, and then searches each square for its objects. It surrounds things it detects with boxes for each box and gives a "score" indicating the likelihood that the object was located. The box with the highest score is selected as the correct one.. Because of its speed, the YOLO v5 version is ideal for applications that require real-time results.

YOLO v8: This model is the new version of the YOLO v5 model. Considering that various enhancements have been made to its design, it is also in some ways more accurate and efficient. In fact, it outperforms its counterpart model in object identification, achieving 12% faster and 10% greater accuracy when processing more features in roughly one GPU.

Region-based Convolutional Neural Network is abbreviated as the faster RCNN

Faster R-CNN is another smart method to detect objects, however it differs a little from YOLO. It views an image, determines regions where it believes there may be some objects, and then presents those regions to a CNN (Convolutional Neural Network) to determine each region's object of interest. This two-step approach takes a bit longer, but it is highly accurate. Faster R-CNN is ideal for projects where accuracy matters most but not much speed does.

ResNet: This is interesting in that it allows for a huge number of layers to stack up without losing any details due to "residual connections." It skips some layers at times to reuse previously learned details and thus avoids problems like data loss during the learning process.

AlexNet: This is one of the first CNNs, consisting of eight layers in total and was novel since it incorporated new ideas such as ReLU activation and overlapping pooling. AlexNet demonstrated CNNs could be useful for plenty beyond image recognition, such as language or medical images.

VGG: VGG uses tiny filters (3x3 pixels), small but strong! By having such tiny filters, VGG picks the details by piling up layers. This thus makes VGG good at picking up the parts of an image into fine detail.

Rendering

After recognizing the gesture, the system converts it into text or speech. Therefore, people who are either not able to listen or even speak can join any conversation.

IV. Experimental results

We have experimented with hand detection using YOLO v8, Faster R-CNN, and YOLO v5 on the ISL dataset. In this result, our experiments depict that YOLO v8 is leading way above all others in terms of accuracy and running time; it turns out to be significantly more accurate for gesture recognition to be faster than all other models, thus being highly suitable for real-time hand gesture recognition systems.

Table 1. Model Performance Results

Sign Language Detection Method	Avg IoU	Avg Time
YOLO v8	0.941	0.0091
Faster RCNN	0.916	2.128
YOLO v5	0.93	0.0189

YOLOv8:

Data labeling: We used bounded box annotations for data labeling. It's a very crucial procedure of a supervised machine learning operation.YOLOv4 is trained on the Darknet framework, utilizing a GPU with 12GB of RAM. yolov3_training_last.weights is created.For testing, spyder IDE is utilized and YOLO is tested for datasets for testing.

Similarly, out of three algorithms we will select one for Image Classification

V. Conclusion

We may have a chance to get closer to a society in which each person will be free to communicate, be it speech or hearing impairment, with the help of this project, "Video Communication Platform for Specially Abled People." The program flows in such a way that it bridges deaf and mute from the hearing world through two major models: one is hand gesture recognition, and the other is speech to text conversion. This really makes real-time communications more inclusive and accessible with its use in video conferencing, messaging, or just regular encounters. The tools are making everyone feel a little more connected.

References

[1] Natarajan, B. (2021). Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation.https://ieeexplore.ieee.org/document/9905589

[2] Sosa Jimenez, C. O. (2022). A Prototype for Mexican Sign Language Recognition and Synthesis in Support of a Primary Care Physician.https://ieeexplore.ieee.org/document/9826586

[3] Aggarwal, R., Meena, A., & Kaur, N. (2020). A comprehensive review of sign language recognition: Different types, modalities, and datasets. https://ieeexplore.ieee.org/document/10696449

[4] Somya Jain,Shikha Diwakar, & Neha Yadav (2023). Dynamic Bidirectional Translation for Sign Language by Using Machine Learning-Infused Approach with Integrated Computer Vision. https://ieeexplore.ieee.org/document/10489440

[5] Deep R. Kothadiya, Chintan M. Bhatt (2021). Hybrid InceptionNet Based Enhanced Architecture for Isolated Sign Language Recognition.https://ieeexplore.ieee.org/document/10577129

[6] **:** Pramod, S., & Kumar, V. (2022). An Integrated Healthcare System Using Sign Language Recognition and IoT for Remote Monitoring.https://www.mdpi.com/1424-8220/23/15/6760

[7] Chen, L., & Zhu, Y. (2022). Multi-Modal Framework for Real-Time Sign Language Recognition Using AI and Vision Sensors.https://www.mdpi.com/2079-9292/12/23/4827

[8] Zhang, Y., & Liu, H. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness.https://docs.ultralytics.com/models/yolov8/

[9] Jacob, A., Koshy, M., & Nisha, K. K. (2021) Real Time Static and Dynamic Hand Gestures Cognizance for Human Computer Interaction. https://ieeexplore.ieee.org/document/9708249

[10] Cui, R., Liu, H., & Zhang, C. (2017). Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. https://ieeexplore.ieee.org/document/8099658

[11] Al Abdullah, B. A., Amoudi, G. A., & Alghamdi, H. S. (2024). Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects. https://ieeexplore.ieee.org/document/10670380

[12] Lai, Zhuiwen, & Huang, Z. (2024). Enhancing Sign Language Recognition and Accessibility for the Deaf Community in China.https://www.tandfonline.com/doi/full/10.1080/09687599.2024.2412271#abstract

[13] Shofia Priyadharshini, D., Anandraj, R., Ganesh Prasath, K. R., & Franklin Manogar, S. A. (2024). A Comprehensive Application for Sign Language Alphabet and World Recognition, Text-to-Action Conversion for Learners, Multi-Language Support and Integrated Voice Output Functionality.https://ieeexplore.ieee.org/document/10561024

[14] Sonsare, P., Gupta, D., Sayyed, J., Nandha, P., & Lairawrite, S. (2023). Sign Language to Text Conversion Using Deep Learning Techniques. https://ieeexplore.ieee.org/document/10486568

[15] Joksimoski, B., Zdravevski, E., Lameski, P., Pires, I. M., Melero, F. J., & Puebla Martinez, T. (2023). Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities.https://ieeexplore.ieee.org/abstract/document/9739689\

[16] Anwar Mohammad Alzghoul(2024) Implementation of Adaptive Technology Tools and Applications for Accessible Physics Education with Deaf and Handicapped Students. https://nano-ntp.com/index.php/nano/article/view/812

[17] Falvo, V., Scatalon, L. P., & Barbosa, E. F. (2022). The Role of Technology in Teaching and Learning Sign Languages: A Systematic Mapping.https://ieeexplore.ieee.org/document/9274169

[18] Smith, J. & Patel, K. (2023). Sign Language Recognition and Translation: A Multi-Modal Approach Using Computer Vision and Natural Language Processing.https://aclanthology.org/2023.ranlp-1.71/

[19] Dorrington, P., Wilkinson, C. R., Tasker, L., & Walters, A. (2016). User-Centered Design Method for the Design of Assistive Switch Devices to Improve User Experience, Accessibility, and Independence.https://uxpajournal.org/ucd-method-assistive-switch-devices-accessibility/

[20] Joksimoski, B., Zdravevski, E., Lameski, P., Pires, I. M., Melero, F. J., & Puebla Martinez, T. (2023). Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities. https://ieeexplore.ieee.org/document/10526274.

[21] Yadav, R., & Sharma, K. (2022). Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges.https://ieeexplore.ieee.org/document/10526274

[22] Chhajed, R. R., & Parmar, K. P. (2021). Messaging and Video Calling Application for Specially Abled People using Hand Gesture Recognition.https://ieeexplore.ieee.org/document/9417924