Grenze International Journal of Engineering and Technology, June Issue

Indian Sign Language Translation System

Swati Kale[1], Atharva Thenge[2], Pratham Werulkar[3], Kartik Chavan[4], Priyanka Jaiswal[5] and Chetan Puri[6]
[1-5]Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur
Email: {s12_kale@yahoo.com, 22030075, 21070857, 22030104, 21070661}@ycce.in, pgj.ycce@gmail.com
[6]Faculty of Engineering and Technology, Wardha 7Email: chetanp.feat@dmiher.edu.in

***Abstract*— Most dumb and deaf citizens across India communicates in Indian Sign Language (ISL). To further enhance the system's functionality, the proposed system enhances the machine translation feature with text-to-voice feature. Facial expressions or movements of hand, signal, or body language is employed by ISL to convey the required message and emotion. The signs may be understood by the people who are not aware of ISL translation system is very important to reduce the barrier of communication between the community of the deaf and hard of hearing individuals. This research also provides proposed real-time sign language translation system via using computer vision technology adopting RandomForest and Large Language Model (LLM) for accurate sentence creation and OpenCV for definition of image and the elimination of the background noise.**

***Index Words*— RandomForest, LLM (Large Language Model), OpenCV, Hand Landmarks.**

I. INTRODUCTION

Language is one of the critical human activities that need to be performed but people with hearing disorders find it difficult to articulate their selves or even to interact with other people. Though sign language helpful for the deaf and hearing-impaired people it is problematic when communicating to those whom they have not seen using the sign language hence the more the gap. Sign Language Recognition (SLR) technology is trying to solve this problem by converting the hand gestures into text, so that they all can interact effectively. Sign language breaks down into manual signs (hands/fingers/arms) and Non-Manual Signals (Facial and Body). These involve ISL alphabets, which is static hand movement, and dynamic hand movement. ISL is different from ASL, JSL or BSL in that many of the signs involve both hands. Although there has been recent improvement in ASL recognition, the subject areas related to ISL are still scarce and there are no systems that help translate regional speech to ISL. This paper presents an android application design and implementation using smartphone technology's camera features to record ISL gestures and converting them into text, and subsequently into real-time comprehensible sentences through machine learning algorithms.

II. LITERATURE REVIEW

Several attempts at spoken to sign language translation are described in the last few decades. The first systems have relied on a rule-based approach and very few have exhausted the statistical method. That sign languages are the natural languages of the deaf or the hear-impaired person. They had been evolved naturally over a period of time. A sign language is a manual-visual mode for communication and it uses meaningful articulation of fingers, hands, shoulders, and heads.

Grenze ID: 01.GIJET.11.2.140_20

In addition to the manual actions of these articulators non-manual actions or facial expressions are also used in sign language. Jane et al, (2018) presented an Indian Sign Language Interpreter that converts the Indian Sign Language and converted it into text and audio using Artificial Neural Network. Through developing this model it was possible to identify that the model was easy to implement and that it was possibleto use the webcam. Jalal and Mammar developed an American Sign Language Posture Understanding with Deep Neural Networks in 2018. The quality of the model is very good when it does not have to segment the images or pre-process them and by using CNN it has achieved surprisingly good rating for picture classification. The model requires the use of strong resources and high levels of computation[1] . Madan Vasishta first brought out the ISL dictionary in 1981. Ulrik Zeshan continued this with her field work and documentation work on ISL in 2004. However, Samar Sinha has continued the ISL study with his doctoral research study in 2013. They are helpful for further computational analysis and language synthesis as well on his part. Sign languages type discussions have been mainly carried out within the 'Spoken Language Grammatical Framework'. Here we have reviewed the important grammatical properties of 'Marathi' and ISL from the literature[2] . For Sign Language recognition, using CNN for image classification, researchers have got accuracies as high as 93.67% (Tolentino) and 99.91% (Harini). Other CNNs such as YOLOv5 and thermal imaging approaches have also been considered following which accuracies of almost 99.52% have been realized. But these methods are applicable if the signs are static only. For making the signs involving motion dynamic, Recurrent Neural Networks similar to LSTMs are used. CNN-LSTM combined models, like Deep Conv LSTM, obtained 91.1% accurate and still overcome similar signs obstacles. The text read from the image is converted to Indian Sign Language using a Direct Translation technique with reference to index-based character mapping.[3] The Histogram of Oriented Gradient (HOG) is one of the best feature descriptors processes that are essential in image processing for object detection as it provides the best interface to support image categorization based on shapes and structures that the image might be holding. HOG identifies the extent and the steepest slope of incoming images while considering circulation of edge direction for object contour recognition.[4]ANNs are models that imitate human intelligence by emulating the biological neural system within the human brain. These networks handle data through a network of interconnected systems by performing functions of computation like pattern matching the classifying. Due to its flexibility ANN is well suited for the construction of forecasting models and MMDB management.[4] Continuous sign sequence recognition is addressed using the proposed modified LSTM-based model that splits the signs into sub-unit then utilizes neural networks to learn all the possibilities therefore no need to combine the sub-sign in training. Incorporating the Leap Motion data, the approach is trained on 35 isolated sign words and the spatial and temporal features are obtained using a video detection scheme. An off the shelf CNN distinguishes between gestures and pulls out frames for improved recognition.

III. DATASET CREATION

For dataset, we are collecting the real gesture images using OpenCV. When any new gesture has to be add we have to add the particular gesture images in our system. Below is the explaination how the dataset was gathered and prepared :

- For creation of dataset various images of the gestures were captured through OpenCV. The images are captured from different angles like from front view , side view and close up of a gesture. Total 1000 images were captured for a gesture in a dataset .
- Each gesture has its separate folder consisting of 1000 images in different lightening conditions and in different angles. The system mapped each gesture with a separate folder which further used to classify the gestures .
- For each gesture the image was captured of different hand lengths and different datapoints of the hand. In which all types of hands gestures should detect the relevant gestures while final output.
- After that , the dataset were created from the collected images through a code and data.pickle file was generated. In the pickle file , the datapoints of the hand landmarks were stored which can use anytime when to train the model .
- If a user wants to add any new gesture it will just add new gestures image and create the dataset and will generate the updated data.pickle file.
- After that the pickle file was used for the training of the dataset. In the training if the hand landmarks was not properly detected or the quality of images were poor that images were analyse two times , if it is proper for training then it will get used otherwise it will skipped those images and used remaining images .
- The above process will be done for each one gesture images which is get feeded in the system to maintain the training accuracy and to get the accurate result.

IV. METHODOLOGY

The system diagram appears in Figure 1. Our system uses high-definition cameras to capture sign language videos while advanced vision and learning methods let it detect hand gestures. Through these processes the system makes spoken words from recognized signs. Our proposed system supports real-time hand gesture recognition through Indian Sign Language (ISL) while automatically converting these gestures into textual sentences for users. Our suggested system works by recognizing finger spelling which belongs to the list of hand gesture systems. Our process contains three key procedures: data collection, data preparation and extraction of important features. This research splits its main processes into six distinct sections of work which are shown in figure 1.

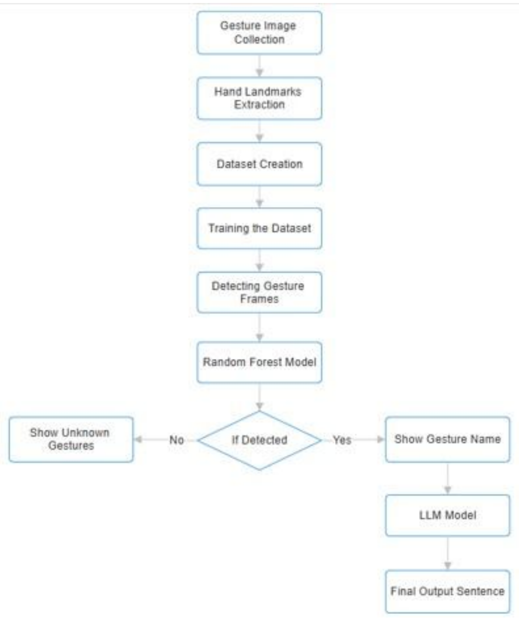

Figure 1: Workflow of System

The system extracts information from webcam images by using hand landmarker models from Mediapipe to find essential elements. The Random Forest model transforms key point detections for dynamic signs. The system converts the extracted key points through an LLM system which creates text sentences. After processing the input the system delivers results both as written text and speech output.

A. Image Collection

Image collection forms the essential base for Indian Sign Language (ISL) to regional language translation systems because the system's outcomes depend heavily on the quality and characteristics of the provided images. Our efforts to collect image data follow specific processes to produce both valid and useful results for machine learning models. We need to control lighting well for effective gesture recognition during our image acquisition phase. Additionally when filming we keep all background objects minimal for increased gesture visibility. The system finds optimal viewing angles to show accurate hand positions and detects all orientation changes. It uses camera vision technology from a computer webcam to capture easy hand positions for comfortable interaction between users and computers. Our feature extractor handles single-hand data accurately since it ignores interactions between two hands. Our method improves how well we recognize hand signals to make interaction smoother. By capturing images under multiple circumstances this system prepares data for its analysis of Indian Sign Language before converting it to local verbal communication.

B. Image Pre-processing

Our project depends on image pre-processing methods to make gestures workable. When we record videos we split each clip into separate frames to encode hand movements and then we split every few frames or increase frame rates to catch tiny details. To simplify the data we select few specific hand position frames that represent all possible movements. Machine learning works better with images that have received noise reduction processing. Our system records 1,000 images from different camera views as it tracks hand positions alongside facial movements. OpenCV helps process images by lowering noise levels and transforming colour spaces while finding object edges and faces to make video stabilization work better.

C. Image (data) Generation

After processing images the system shows users their hand movements as spatial feedback and translates their signs into results. Our technology identifies precise hand and body locations associated with specific signs from Indian Sign Language. Machine learning tools from MediaPipe deliver precise results by finding body landmarks which the gesture recognition system needs to work properly. The system displays images that show where detected gestures match their true meaning for live interaction. The system uses visible signs to improve user interaction while making sure the gestures are understood correctly before next steps.

D. Model Training

During data preparation the result is saved to a data.pickle file. The data.pickle file starts the model training procedure. We load our prepared data to train a machine learning system that detects gestures reliably. As the training progresses the model repeatedly updates itself to reduce detection mistakes from supervised learning approaches. When training completes the model achieves adequate accuracy the trained parameters get saved as model.pickle for reuse. The real-time gesture predictions depend on this serialized model design to work smoothly and reliably. During training the model generates visual displays showing how accurate predictions changed with each learning phase of the process as shown in Figure 2. The system shows both the training accuracy graph and line interface results after training ends. Our dual method shows the model's performance progress at total glance during its training phase. This graph tracks how well the model learns to identify gestures which demonstrates that the training succeeded.

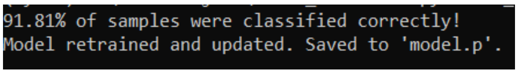
91.81% of samples were classified correctly!
Model retrained and updated. Saved to 'model.p'.

Figure 2: Training Accuracy of Model

V. RESULTS

The feasibility of the proposed ISL system was tested on a wide variety of hand movements and each position of both hands was examined. The model is successful in eliminating noise in the signs thus increasing the probability of identifying features. Some difficulties come with new sign presented in different position or with repeated movements because similar signs share the same label although they look alike. The same pre-processed dataset was employed for the training as well as the test phase; the quality of images and expanding the colour space seemed to enhance its performance. Variations in lighting conditions, backgrounds and specificity of users were addressed in the dataset used. While there is recognition of other movements not in the list, it is labelled “unknown” and one is allowed to edit the dataset to add new gestures. The system had a 10% false detection rate for gestures meaning it was capable of pinpointing invalid gestures while correctly differentiating valid ones. Correct gestures provided the right output as depicted in table 1 which made the system functional in all the conditions tested. Real time Implementation In real time image data is analysed and the results of gesture recognition are given without any time delay. As it was seen during testing the ability of the system to distinguish gestures took place in milliseconds allowing for smooth and continuous interaction. This makes the system ideal for specific real time applications due to the low latency which arises from optimized processing. This tool provides an advanced and efficient application for the integration of real-time with techniques that make it allow personal communication from any perspective.

TABLE I: SHOWING FINAL SENTENCES AFTER CAPTURING ALL FRAMES

Gesture Detected	Final Output Sentence
 	डोके दुखी आहे. मला ताप आहे.

A. Real-time Recognition

Our system provides instant analysis of image data to detect gestures as users see them on their screens. Our tests confirmed the system's ability to read hand gestures in less than a millisecond which succeeded to keep communication flowing without interruptions. The optimized processing system provides fast results which make it well-suited for live human interaction. The software merges today's best technology with real-time functions to deliver an adaptable system that boosts inclusive interaction.

VI. Future Work

Sign language translation and recognition show endless use with today's advanced technology. Artificial intelligence tools and machine learning work better today to recognize signs because they identify more of them accurately. Our objective is to detect sign languages of various cultures to help states communicate with each other. Under the system users will communicate through text-based, gesture, speaking, and image formats for seamless integration. Research teams aim to optimize the proposed mobile and edge device system through various strategies that reduce model size and perform lightweight computations. Support groups for the deaf and hard-of-hearing community will help build this system by showing us how to fix what is currently breaking online connections between nations.

VII. Conclusion

This is where this project comes in, and fill the need for real time communication support by providing interpreted sign language and translating it into different regional languages. Speaking in the second language, in a country where different states are having different languages it provides equality to everyone. Sign language, as much as it is a one provocative use of hands, face and overall body movement, has strict grammar and syntax meaning it is not just a language but a complete one. Sign language translating technology gives the deaf and the hard of hearing improved chance not only in the health sector, education, and employment opportunities. However, personalization in terms of hand motion and facial expressions, the lack of familiarity with different cultural sign languages also remain a problem which needs to be resolved to increase the efficacy of the existing system. When it comes to sign language translation, there is definitely a level of technology that one could imagine would make a very much positive difference within this population group.

References

[1] Jalal MA, Chen R, Moore RK & Mihaylova L.(2019) American Sign Language Posture Understanding with Deep Neural Networks.

[2] U. Zeshan and others, "Sign Language in Indo-Pakistan," A description of a signed language. Amsterdam: Benjamins, 2000.

[3] S. X. Thong, E. L. Tan and C. P. Goh, "Sign Language to Text Translation with Computer Vision: Bridging the Communication Gap," 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024

[4] M. Rajmohan, C. Srinivasan, O. R. Babu, S. Murugan and B. S. Kumar Reddy, "Efficient Indian Sign Language Interpreter For Hearing Impaired," 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon), Singapore, Singapore, 2023

[5] K. B. Prathap, G. D. Swaroop, B. P. Kumar, V. Kamble and M. Parate, "ISLR: Indian Sign Language Recognition," 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems,Machine Learning and Signal Processing (PCEMS), Nagpur, India, 2023

[6] A. D. Shetty, J. Shetty, K. K, Rakshitha and S. S. B, "Real-Time Translation of Sign Language for Speech Impaired," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC),Erode, India, 2023

[7] K. S. Sindhu, Mehnaaz, B. Nikitha, P. L. Varma and C. Uddagiri, "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired," 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India, 2024

[8] S. Sivamohan, S. Anslam Sibi, T. R. Divakar and S. Jagan, "Hand Gesture Recognition and Translation for International Sign Language Communication using Convolutional Neural Networks," 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2024

[9] Jaiswal P. G., et al. (2023). A stacking ensemble learning model for rainfall prediction based on Indian climate. Proceedings of the 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 1–6.

[10] S. G. Kale et al., "Identification of Ayurvedic Leaves using Deep Learning," 2023 International Conference on Communication, Circuits, and Systems (IC3S), BHUBANESWAR, India, 2023, pp. 1-6, doi: 10.1109/IC3S57698.2023.10169388.