A Review on Real-Time Sign Language Recognition System

Anjaiah.K
Assistant Professor, Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Gaddam Gagandeep Yadav
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Elemenate Akshith reddy
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Sabhawat Rahul Naik
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Aamina Arjumand
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Marlapudi Shruthi
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

***Abstract*— Communication problems often isolate people with speech and hearing abilities from the wider society because sign language – an important element for teaching – is still unknown to most people. This paper presents a knowledge base for terminal operations that uses advanced computing and neural network techniques to improve this gap. The proposed method is to recognize gestures and convert them to text or speech, thus enabling interaction between people with disabilities and the public, and combines image processing, deep learning, and neural networks to accurately recognize gestures. The PyQt-based GUI application integrates and provides users with an easy-to-use, real-time experience using live video techniques. Key features include powerful manual search, a pre-programmed pipeline for normalization, and predictive modeling from pre-trained neural network classifiers. This application provides high accuracy, works instantly and can be adapted to many languages. Forcing mechanisms and hybrid autoencoders to extract spatial, temporal and abstract features from hand gestures. This new approach was evaluated using Indo-Russian and Pakistani Sign Language (PSL) data and demonstrated the ability to recognize performance in different contexts. The system also addresses the important challenge of distinguishing between standard and non-standard, which is important for learning and making language learning consistent. educational tools for sign language learning and human-computer development. Through augmentation and accessibility, this research offers solutions to empower people with disabilities, promote effective communication and increase global understanding.**

Keywords—Neural Networks, Image Processing, Gesture-to-Text Conversion, Inclusive Communication, PyQt GUI Application

I. INTRODUCTION

Communication is an important part of human relations, but people with speech and hearing impairments often face serious problems due to limited understanding of sign language in society. This project introduces real-time cognitive processing using advances in computer vision, neural networks, and deep learning to bridge the communication gap. The system captures video and processes hand gestures, converting them into text for seamless interaction. The system ensures that people are accurate and works instantly using technologies such as the "cvzone" library for hand control, pre-trained neural networks for classification, and PyQt for a user-friendly GUI. It normalizes and pre-processes the input images to fit the training model, thereby aiding the prediction process. It supports many languages and sign language formats such as Indo-Russian Sign Language and also has a distinction between standard and non-standard, making it useful for communication and education. The system promotes participation, accessibility and effective communication by encouraging interaction between people with disabilities and the wider community. Its flexibility and ability to be used in education, public services and workplaces further highlights its importance as a revolutionary service technology that provides solutions to support people and improve international understanding. This system addresses the critical need for participation by providing people with the voice and capabilities they need. Its applications extend beyond personal communication to include educational tools, workplace access and even public interaction. The program allows people with disabilities to participate more in social, educational and business life by reducing communication barriers and promoting greater integration. It is a significant step forward in offering new ideas and solutions to bridge the gap between people with disabilities and the wider community. Combining advanced neural networks, real-time video processing, and user experience, the system provides a powerful platform for good communication, learning, and access.

OBJECTIVES AND METHODOLOGY

This project was designed to create an instant navigation system that will facilitate communication for the speech and hearing impaired, allowing them to interact with the public. Using advanced computing, neural networks, and deep learning, the system processes real-time video streams to detect gestures and movements and convert them into text. The system uses the cvzone library for hand tracking and classification, and PyQt for intuitive GUI that provides high accuracy, low latency, and user-friendliness. Gesture recognition is performed with initial steps such as cropping, resizing, and normalizing the image, followed by classification using a pre-trained neural network. The system is designed to support various sign languages, distinguish between standard and non-standard, and is suitable for communication and education. It combines real-time video capture, feature extraction, and graphical output for smooth and responsive operation. The modular system allows for customization and supports multiple languages and gestures. The program offers transformative tools that promote accessibility, inclusivity, and effective communication, empower people with disabilities, and foster global understanding.

II. LITERATURE SURVEY

There has been much research in the field of linguistics on creating systems that bridge the gap between the hearing and speech impaired and the general population. Traditional language recognition methods, such as electronic devices, have limitations such as high cost, large equipment, and the need for specialized equipment. These systems often require users to wear gloves or sensors to track hand and finger movements, and it is not easy to switch to different languages due to the different graphemes and gesture patterns of the hands. Furthermore, the reliance on these tools also affects mass adoption. Early image processing techniques, such as edge detection, contour analysis, and optical flow correction, had difficulty dealing with lighting changes, hand movement, and background clutter. Recent advances, particularly the use of neural networks (CNNs), have increased the accuracy of gesture recognition. These deep learning models can extract hierarchical features from raw data, thus reducing the reliance on manual extraction. This increases the power of hand and gesture recognition even in different environments. Tools like cvzone's HandTrackingModule are popular for hand detection and tracking in videos. Machine learning models like Support Vector Machines (SVMs) and CNNs are also used to classify gestures into predefined groups. Advanced procedures such as image resizing, cropping, and normalization play an important role in reducing the effects of changing cell size, orientation, and background noise, thus improving the sharing of reality. Many methods are being investigated that combine the visual system with other data, such as depth sensors or electromyography (EMG) signals, in addition to the visual system. This hybrid technique aims to improve recognition accuracy by combining different information. Although multimodal systems provide better performance, their implementation is usually complex and expensive. In addition, good recognition performance can be improved by combining physical features using models such as Long Short-Term Memory Networks (LSTM) or Bi-LSTM, which capture the spatial and physical aspects of movement. Develop learning tools to better teach languages. This tool helps students improve their skills and understand how their movements work with sign language patterns by providing instant feedback on the user's movements. Because there are regional differences in sign language, standards have been developed for specific languages such as Indian Sign Language (ISL), American Sign Language (ASL), and Pakistani Sign Language (PSL). This requires the creation of specific data to train the model on regional sentences to ensure that the system can recognize a variety of gestures. Pre-trained models such as VGG16, ResNet, and InceptionNet allow for the use of well-known methods at the client level, improving the ability to fine-tune feature maps. The reduced computational requirements, combined with improvements in accuracy and immediate performance, make the approach more accessible and efficient for language recognition. Challenges remain, especially when dealing with occlusions, complex movements, and changing lighting. Real-time use of AI remains challenging because these systems need to be reliable and provide immediate feedback in many areas. However, research and technology continue to increase the efficiency and usability of language recognition, helping to provide deaf and mute people with more communication opportunities.

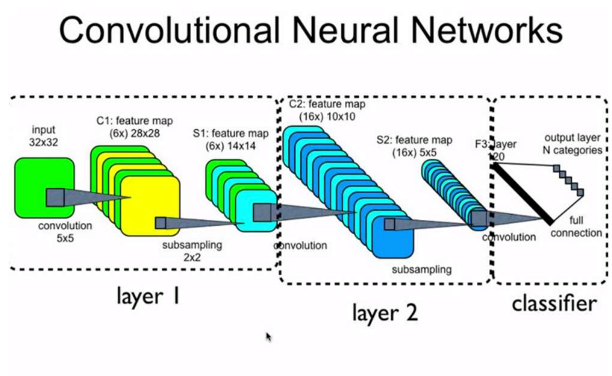


Fig. 1: Convolutional Neural Network

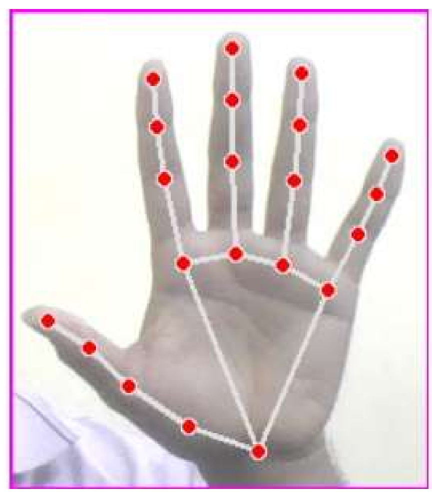


Fig. 2: Hand with Key Points Marked

III. PROPOSED SYSTEM

The goal is to bridge the communication gap between the hearing and speech impaired and the general public by providing easy-to-use, intuitive tools for real-time. Leveraging computer vision, deep learning technology, and real-time video processing, the system can recognize gestures and communicate with language-using humans by translating words into text. Key features of the system include a user-friendly graphical user interface (GUI) built using PyQt5, a pattern recognition function that combines hand recognition and gesture classification, and image processing techniques to normalize and resize gestures to improve classification accuracy.

A. User-Interference Design:

The system is built with a simple yet easy-to-use graphical user interface (GUI) built using PyQt5, which allows for real-time interaction. The user interface has two main elements:

Video location display: The live video from the network camera is displayed in the GUI. This is the main feature of the application, providing a direct view of the captured frames and allowing users to track their movements while the system processes the video in real time. Gestures recognized by the system. The text box cannot be edited to ensure that it is used only to display text representing the recognition of gestures. This flexibility is particularly important to ensure that people, including those with cognitive impairments, can use it effectively to communicate.

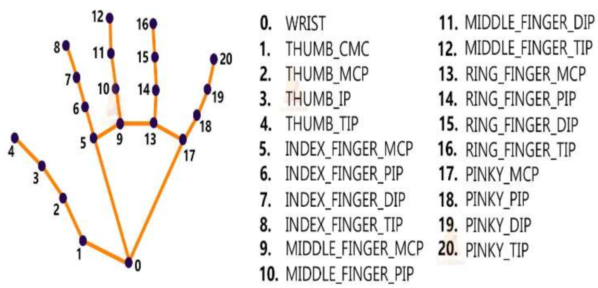


Fig. 3: Hand landmarks

B. Gesture Recognition System:

The basis of the system is a gesture recognition model that is responsible for recognizing gestures and converting them into text representations. The cv zone Hand Tracking Module is used to instantly detect and track hands. The Hand Detector class is particularly useful because it can detect the bounding box of the hand and track its movement in the frame. This class uses state-of-the-art hand detection technology, allowing it to recognize the right hand in different conditions, such as different lighting, distance or angle. Points of interest (POI) on the hand and prior distribution. Gesture images are resized and normalized before being passed to a pre-learning deep model (CNN) for classification. The system confirms the movements by marking them with one of the pre-recorded text messages such as "Hello", "Thank you", "Yes" or "No". The previous description corresponds to the text that appears in the GUI.

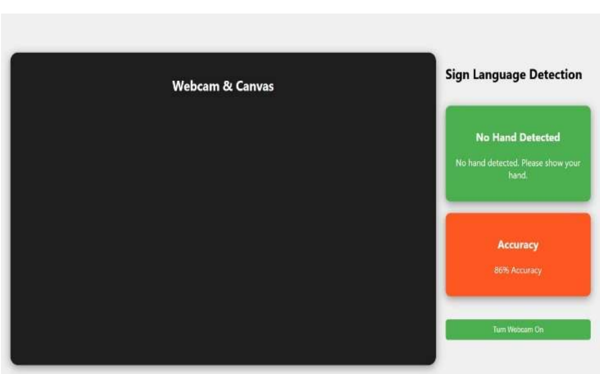


Fig. 4: User Interface

C. Image Processing for Gesture Recognition:

Preliminary images for gesture recognition:
In order to achieve high resolution and reduce external effects, preliminary image techniques are used for frame capture:

Insertion area: When a hand is detected, the system uses the bounding box coordinates to separate the hand from the rest of the frame. This implant eliminates background noise or noise that could interfere with gesture recognition, ensuring that only the movements are processed. In this case, make the neural network take a large change. Image normalization means normalizing the pixel values (scaling the pixel values between 0 and 1), making it easier for neural networks to process the data. aspect ratio. The system detects when the height or width of the hand is larger and adjusts as necessary, ensuring that the definition of the ratio is maintained before entering the model.

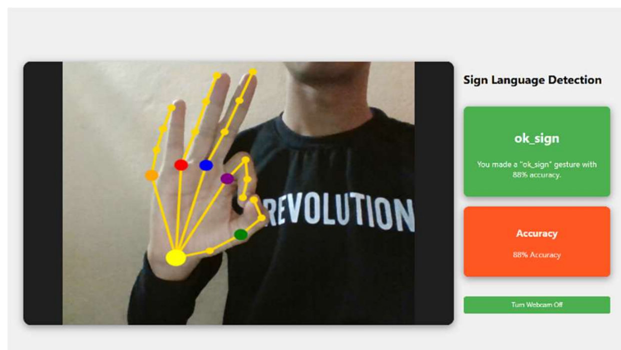


Fig. 5: Real-time recognition of Gestures

D. Real-time video processing:

The system is designed to work in real time and provide regular feedback to users. It captures frames from the web using OpenCV, a widely used computer vision system for video capture and frame processing. The following steps occur during real-time video recording:

Frame capture: The network camera continues to capture images and send them to the standard information grid for analysis. This process is very fast, producing a frame every 20 milliseconds. This allows users to instantly find new video sources with accurate descriptions. Use QtGui. This conversion ensures that colours appear vibrant on the screen.

E. System Workflow:

Initialization: The application is launched, initializing the webcam and loading the model. The GUI components (video feed and gesture text box) are displayed to the user.

Frame Capture: The webcam continuously captures frames, which are displayed in the video feed.
Hand Detection: The system detects the hand within each frame and isolates the region of interest (ROI).

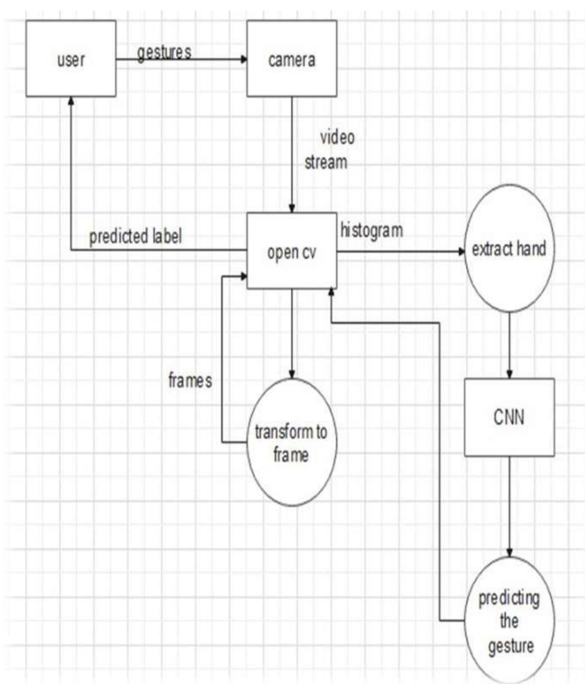
Gesture Classification: The cropped hand image is passed through the trained deep learning model for classification.
Display Prediction: The predicted gesture (as text) is displayed on the GUI.

Real-Time Updates: The system continuously updates the video feed and gesture predictions every 20 milliseconds.

IV. IMPLEMENTATION

Gesture recognition combines computer vision and machine learning to interpret human gestures. The system is designed to detect hands in video frames and classify them by descriptors before using a learning model. This involves preprocessing the hand drawing, resizing it to size, and feeding it into a separate chamber for prediction. These systems can be used for human-computer interaction, language recognition, and interface management. Training data consists of gesture images organized into clusters and augmented with data augmentation to increase power. CNN extracts features from multiple convolutional processes, places them across all layers, and predicts movements. After training, the model is saved for immediate use in applications. This session displays live video from the webcam and provides an estimate of the detection time. It has a handheld sensor that separates regions of interest and makes them suitable for classification. Users can view familiar movements in the application, making the application intuitive and interactive. Such systems can be integrated into assistive devices, robots, games, and virtual reality environments. Gesture recognition is paving the way for innovative, user-friendly technology by leveraging advanced machine learning technology.

A. Architecture Diagram:



This design represents the cognitive performance of an aircraft using computer vision and neural network (CNN).

The user makes gestures that are recorded by the camera as a video stream. The video images are processed using OpenCV to extract histograms and separate the cell region from the background. The extracted hands are then converted to a suitable format (resized or normalized) and fed to the CNN model that predicts the gestures. The predictions are parsed into text and sent back to the users or used in applications such as virtual interaction or gesture interpretation. The system enables human-computer interaction through real-time gesture recognition.

B. Test Cases:

Test Case Id	**Description**	**Input**	**Expected Output**	**Remarks**
TC1	Initialization of Gesture Recognition class	Valid model path and label path	Object is initialized successfully without errors.	Verify paths are correct.
TC2	Hand detection and bounding box extraction	Frame containing a visible hand	Hand is detected, and bounding box is drawn around the hand	Requires proper lighting.
TC3	No hand detection	Frame without a visible hand	"No Hand Detected" is displayed	Test with an empty frame
TC4	Gesture classification for known gestures	Frame with a gesture (e.g., "Hello")	Predicted gesture matches the expected label	Requires labeled gestures.
TC5	Gesture classification for unknown gestures	Frame with a gesture not in the training data	Reasonable fallback or "No Hand Detected"	For gestures not in labels.

TC 6	Gesture resizing and normalization for different aspect ratios	Frame with varying hand sizes and positions	Cropped and resized hand image matches expected size (300x300) and is normalized.	Validate resizing logic.
TC 7	Model training process	Valid training and validation datasets	Model trains successfully without errors; accuracy and loss metrics improve over epochs	Requires proper dataset setup.
TC 8	Model prediction for a single gesture image	Preprocessed image of a known gesture	Predicted label matches the actual label	Validate against labels.
TC 9	Prediction text area update	Frame containing a known gesture	Predicted gesture is displayed in the text area	Ensure synchronization.

V. DISCUSSION

A. Usability and User Experience:

The two vital components for a successful gesture recognition system are usability and user experience. The user interface of the system is very intuitive, designed simply and accessible for the users; it features clear, real-time feedback on layout. The ease with which one can use this system lies in the performance of hand gestures to be translated quickly into text predictions. It assures real-time responsiveness, meaning quick results for immediate engagement and better user satisfaction. However, users have to make sure their hands are in the camera's view for detection, which might be a little challenging for some to learn. Visual cues that help users position their hands correctly could be added to enhance usability. The simplicity of this gesture-based input method makes it suitable for people with disabilities who use sign language to communicate. Gesture recognition can be used as a hands-free alternative when traditional input methods such as keyboards or touchscreens are impractical. However, performance may differ under various environmental conditions, such as lighting or background distractions. A seamless user experience is contingent upon the ability to provide consistent accuracy and ease of interaction, thus encouraging long-term use. Continuous testing and iteration will be crucial in optimizing both usability and overall user satisfaction.

B. Image Pre-processing and Normalization:

When a hand is detected, a bounding box (bbox) is used to slice the hand region. The image is then resized and normalized before being fed into the motion model. Normalization is an important step because it ensures that the model is working properly by measuring pixel values between 0 and 1. The aspect ratio ensures that the input image maintains the same properties for the distribution. This process is important to ensure that the model generalizes well to different shapes and sizes. In this case, data enhancement techniques (e.g. rotation, transformation) can be used to improve the quality of the model.

C. Applications in Real World Scenarios:

The application of gesture recognition in real-life scenarios covers a wide range of industries and offers innovative solutions for individuals and businesses. In Assistive Technology, the system can convert hand gestures into text or speech, closing the communication gap for people with hearing or speech impairments and enabling greater interaction. In healthcare, gesture recognition can assist patients with limited mobility, such as in wheelchairs, smart home devices or hospitals with home automation, thus promoting greater independence and improving quality of life. In the field of Virtual Reality and Augmented Reality, guidance can provide effective control procedures, eliminate the need for a controller, and allow users to interact in a virtual environment. In Human Computer Interaction, gesture recognition supports touchless interfaces, which is especially useful in environments where using common sense is difficult or impossible, such as public kiosks, ATMs, or smart devices Interactive entertainment applications are also useful because users can control media with simple gestures, leading to increased media consumption. In the Smart Home, navigation information can enhance the user experience by controlling lighting, heating, and appliances without the need for physical interaction with the device. Finally, Robot-scan incorporate cognitive guidance to control robots in a variety of areas, from manufacturing to self-service. These real-world applications show how gesture recognition can streamline interaction, improve accessibility, and enable innovative solutions across multiple platforms.

VI. CONCLUSION AND FUTURESCOPE

The Gesture Recognition Project has successfully developed a functional and interactive system that can instantly recognize and interpret hand gestures. The system uses computer vision and deep learning models to accurately identify and classify various gestures, providing users with an alternative to manual interaction. The integration of the PyQt5 graphical user interface (GUI) provides interaction and rapid feedback, improving the overall user experience. The system's ability to translate gestures such as "hello," "thank you," and "I love you" is an important step towards creating easy-to-use communication tools, especially for people with hearing or speech impairments. Although this system has great promise, issues such as lighting, background noise, and hand placement still affect the accuracy and reliability of directional information. This technology has established a solid foundation for cognitive navigation, which has been used in a variety of applications, including healthcare, gaming, education, and augmented reality. The system now allows for seamless integration into a conference environment, making it ideal for future assistive technology, security, and IoT devices. The ability to use gestures to control devices without physical contact makes the body particularly useful in medical facilities where inputs like keyboards or screens are normally used, or for people with disabilities where grey touch is not a good fit. There are many areas where the project can be improved and expanded going forward. Firstly, the gesture recognition model can be expanded to recognize more descriptive and detailed information, allowing the system to meet communication needs. This could include multi-hand detection, which can recognize gestures involving both hands simultaneously. Furthermore, solving environmental issues such as lighting changes and complex backgrounds will increase the accuracy and reliability of the system. Exploring the use of depth sensors or infrared cameras could help solve these problems and provide consistent performance across different situations. or wearable devices. This would increase the mobility of the technology, making the job of the viewer easier. Integration with IoT devices and smart home devices could also be interesting, such as using gestures to control lights, appliances or security. You could also look into gesture-based security, which would provide a new and secure way to authenticate users. This could be done by optimizing the underlying machine learning model or using faster hardware such as GPUs or edge computing to make instructions more efficient. The system could also benefit from expanding its gestures to support more languages and gestures, such as different languages, making the device more usable and accessible to an international audience. Integrating technology with interactions such as voice or eye tracking will lead to the development of multimodal systems that provide more information and adapt to the user. By combining gestures with voice commands or visual cues, the system could provide greater understanding and flexibility, especially in environments where protection or non-invasive control is not important. The project has made significant progress in gesture recognition, with many opportunities for improvement and expansion. This system has the potential to change the way we interact with technology, making it easier, more accessible, and adaptable to the needs of different users in many ways by addressing current limitations and exploring new applications.

VII. REFERENCES

[1] Y. Saleh and G. F. Issa, Arabic sign language recognition through deep neural networks ne-tuning, Int. J. Online Biomed. Eng., vol. 16, no. 5,pp. 7183, 2020.

[2] X. Jiang, M. Lu, and S.-H. Wang, An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for ngerspelling recognition of Chinese sign language, Multimedia Tools Appl., vol. 79, nos. 2122, pp. 1569715715, Jun. 2020.

[3] O. Sevli and N. Kemalo lu, Turkish sign language digits classification with CNN using different optimizers, Int. Adv. Researches Eng. J., vol. 4, no. 3, pp. 200207, Dec. 2020.

[4] K. Wangchuk, K. Wangchuk, and P. Riyamongkol, Bhutanese sign language hand-shaped alphabets and digits detection and recognition, Ph.D. dissertation, Dept. Comput. Eng., Naresuan Univ., Phitsanulok, Thailand, 2020.

[5] R. Elakkiya and E. Rajalakshmi. ISLAN. Mendeley Data. Accessed:Jan.8,2021.[Online].Available:https://data.mendeley.com/datasets/rc349j45m5/1

[6] W. Sandler, The phonological organization of sign languages, Lang. Linguistics Compass, vol. 6, no. 3, pp. 162182, Mar. 2012.

[7] R. Elakkiya, Retraction note to: Machine learning based sign language recognition: A review and its research frontier, J. Ambient Intell. Humanized Comput., vol. 12, no. 7, pp. 72057224, Jul. 2022.

[8] S.DiwakarandA.Basu, Amultilingual multimedia Indian sign language dictionary tool, in Proc. IJCNLP, 2008, p. 57.

[9] S. K.Liddell and R. E. Johnson, American sign language: The phonological base, Sign Lang. Stud., vol. 1064, no. 1, pp. 195277, 1989.

[10]P.EccariusandD.Brentari,Symmetryanddominance:Across-linguistic study of signs and classifier constructions, Lingua, vol. 117, no. 7, pp. 11691201, Jul. 2007.

[11] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, Robust part-based hand gesture recognition using Kinect sensor, IEEE Trans. Multimedia, vol. 15, no. 5, pp. 11101120, Aug. 2013.

[12] C. Wang, Z. Liu, and S.-C. Chan, Superpixel-based hand gesture recognition with Kinect depth camera, IEEE Trans. Multimedia, vol. 17, no. 1, pp. 2939, Jan. 2015.

[13]Z.Liu,X.Chai,Z.Liu,andX.Chen,Continuousgesturerecognitionwith hand-oriented spatiotemporal feature, in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 30563064.

[14] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, Online detection and classi cation of dynamic hand gestures with recurrent 3D convolutional neural networks, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 42074215.

[15] R. Agarwal, Bayesian K-nearest neighbour based redundancy removal and hand gesture recognition in isolated Indian sign language without materials support, IOP Conf. Ser., Mater. Sci. Eng., vol. 1116, no. 1, 2021,Art. no. 012126.

[16] O. Aran, SignTutor: An interactive system for sign language tutoring, IEEE MultiMedia vol. 16, no. 1, pp. 8193, Mar. 2009.