Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

Empowering Non-Verbal Communication: Sign Recognition To Multilingual Text And Audio

Dr. Venkata Satya Santhi S[1], Bugatha Surya Teja[2] , Kesamsetty Bhavya Sri[3], Kaicharla Lalith Kumar[4], and Indugubilli Hitha Varshith[5]

[1]Associate Professor, Department of Computer Science and Engineering (AI/ML & DS),
Anil Neerukonda Institute of Technology & Sciences (ANITS), Visakhapatnam-530001, India

[2,3,4,5] UG Students, Department of Computer Science and Engineering (AI/ML & DS),
Anil Neerukonda Institute of Technology & Sciences (ANITS), Visakhapatnam-530001, India

Emails: svssanthi2016@gmail.com[1], suryateja1938102074@gmail.com[2], bhavya.kesamsetty@gmail.com[3], kaicharlalalithkumar@gmail.com[4] , hithavarshith@gmail.com[5]

Corresponding Author Email: suryateja1938102074@gmail.com

Article History:

Received: *21-01-2025*

Revised: *5-03-2025*

Accepted: *12-03-2025*

Abstract

Artificial Neural Networks and Computer Vision have significantly been used to detect hand gestures and facial expressions allowing devices to analyse and recognise images, in turn enhancing the experience of interaction between humans and machines. Existing research in this area primarily focuses on interpreting hand gestures and mapping them to the corresponding alphabets or symbols. In this paper we focus on taking this research forward by introducing an enhancement to the existing model. The proposed model mainly concentrates on improving gesture recognition accuracy in diverse background environments like supporting multiple languages, allowing gestures to be translated into text or speech in different languages. It will utilize a webcam to detect the hand gestures in real time, constructing words based on the recognized gestures and generating sentences. Additionally, the proposed model translates the constructed sentences into a language of one's choice and also allows the text to be converted to audio, providing an audio output. This increases the system's flexibility and facilitates the hard of hearing in a diverse environments and languages. Furthermore, the model showcases its efficiency and novelty by achieving an accuracy of 98.51%.

Keywords: Hand Gesture Recognition; Computer Vision; Artificial Neural Networks; Multilingual Translation.

1.0 INTRODUCTION

The deaf and hard-of-hearing community is over five percent of the global population, according to estimates, and this is why sign languages play a significant role in this community. The difficulties they experience are the necessity to communicate those who are unaware of the use of sign language. Written communication could be used, but it tends to be quite slow and ineffective in day to day Lives. Hiring a professional sign language interpreter is an alternative method, but only in a few cases, it remains possible when engaged in everyday life.

Languages that are conveyed through signs are recognised as fully distinct languages in which area, exhibiting their distinct rules and syntax. For instance, American Sign Language (ASL) becomes difficult to understand by British Sign Language (BSL) users and vice versa. ASL is simple to transcribe, it is the most popular sign language across the world and is

accepted by the deaf community. 30 Each sign language strengthens the global deaf community by focusing on the unique history and culture of its users independently across the world.

Existing sign language recognition methods have several limitations in their efficiency and usability. The lack of background noise detection results in interference during recognition, which is a major issue in gesture recognition. Hand gestures are also usually not dynamically responsive, causing delays in communication with the user experience. The inability to recognise some sign language gestures accurately further degrades system reliability. Most of these solutions are dependent on high-end hardware, thus being expensive and out of reach for a larger population.

However, proposed work has the ability to provide effective background clearance, improving the accuracy of gesture recognition in complex environments. Additionally, these gestures allow interactive and faster communication, engaging the user in a better way. In the proposed system, the model's accuracy was highly boosted in terms of hand gesture recognition by relying on the MediaPipe tool. It connects the dots between the key segments of the hand, helping to detect and track better, which leads to much more reliable feature extraction. By concentrating on the touch points, this tool managed to improve gesture recognition accuracy by providing adequate information about the palm’s position, angle, and other essential parameters

2.0 Literature Survey:

Prasanthi Y et al. (2023) [1], developed a visual Sign Language Recognition system using [9] a Convolutional Neural Networks (CNN) to recognize hand gestures from webcam images. The system processes real-time gestures and displays their meaning as text, facilitating communication for the deaf and hard-of-hearing community. By relying on cameras instead of sensors, the system is cost-effective and robust in gesture recognition, making it an efficient communication tool.

Kanchan Dabre et al. (2021) [2], developed Image processing based interpretation for Indian Sign Language. It captures hand gestures from webcam footage and converts them into text and speech in real-time without requiring gloves or sensors. The system uses the Haar Cascade Classifier for gesture recognition and a speech synthesizer for audio output, providing an affordable, accessible solution for smoother interaction between non verbal individuals.

Areesha Gul et al. (2020) [3], developed a smart two-way interaction system for Deaf & Dumb individuals, facilitating seamless interaction with normal individuals. Using Leap-Motion-Controller for gesture recognition, Raspberry Pi for hardware integration, and an Android app with Google API for speech-to-text conversion, the system allows real-time communication. Hand gestures are translated into [10] speech and vocalized responses as text. Bluetooth enables data transmission between devices. The system also includes emergency calling and location tracking, making it highly practical for healthcare settings, including dentistry.

Suneetha Mopidevi et al. (2023) [4], developed the HGRSL TV system to enhance communication for Deaf & Dumb individuals by converting hand gestures into text and voice using the Leap Motion Controller, Raspberry Pi, Bluetooth, and Hidden-Markov-Model (HMM). The system operates in real-time, leveraging Google's MediaPipe framework, TensorFlow, OpenCV, and Python. It captures hand gestures, identifies landmarks, and classifies them with 95.7% accuracy for ten gestures like Thumbs Up or Peace. MediaPipe's landmark detection improves speed and accuracy over traditional methods, offering a robust real-time gesture recognition solution.

Mohammedali et al. (2022) [5], developed a real-time system focused on improving interaction with unhearing individuals using ASL. The system influences CNN, specifically the Squeezenet module, for gesture recognition. It achieved 100% accuracy in offline testing and 97.5% in real-time applications, surpassing traditional methods like HOG. With an average response time of 3.3 seconds, the system captures gestures and converts them into text and speech, streamlining communication and enhancing accessibility for the deaf community.

Table 1: Literature Survey

SNO	AUTHORS	TECHNIQUES	LIMITATIONS
1.	Prasanthi .Y et al.[1]	Convolutional Neural Network (CNN),Real-Time Frame Analysis, Indian Sign Language (ISL) Database, Text Display	The system's recognition accuracy is affected by its difficulty in distinguishing gestures overlapping features. It is mainly designed for Indian Sign Language (ISL), which restricts its applicability to other sign languages or gesture systems.
2.	Kanchan Dabre et.al.[2]	Gesture Recognition Using Haar Cascade Classifier, Computer Vision, Text and Speech	Inability of the system to distinguish gestures that have similar feature overlaps, it is negatively impacting its performance. It struggles to recognize significant gestures accurately in dynamic environment.

3.	Areesha Gul et.al.[3]	Leap Motion Controller, Raspberry Pi, Android App with Google API, Bluetooth Emergency Calling and Location	The system depends on the Leap Motion Controller, which may[11] not perform well in low-light or obstructed environments. Problems with Bluetooth connectivity can interrupt the operation of the system especially in situations where accurate and constant communication is needed.
4.	Suneetha Mopidevi et.al[4]	Leap Motion Controller, Raspberry Pi, Bluetooth, Hidden Markov Model (HMM), Google's MediaPipe, TensorFlow and OpenCV, Python	Limited to a certain number of specific languages. Reliance on a pre-trained model but not new or uncommon gestures that were not included in the training dataset.
5.	A.H. Mohammedali et.al[5]	Convolutional Neural Networks (CNN) Offline and Real-Time Testing Text and Speech Conversion Comparison with Traditional Methods	It needs a powerful hardware for real-time processing, creating challenges for the end user. It struggles in noisy or uncleared environments and has trouble recognizing gestures.

3.0 Methodology:

The proposed work for hand gesture recognition [7] in multilingual speech comprises four phases. During the initial phase, input data is preprocessed and hand gestures are detected in real time via a webcam. In the second phase, the pre-processed hand gestures are utilized to identify the alphabets. Third phase involves constructing words and framing sentences by storing the recognized signs in the buffer. Final Phase is to convert the constructed sentences into the desired language and produce an audio output. The proposed work is represented as a diagram in Figure 1.

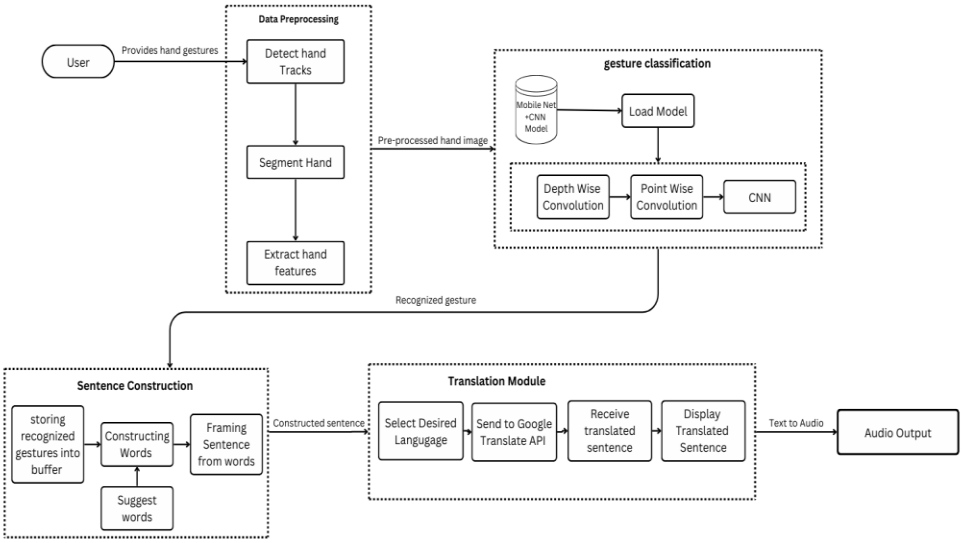

Fig. 1: Architecture of proposed model

3.1 Phase-1: Data Preprocessing of Hand Gesture

In Phase 1, the model captures real-time camera input of the user [8] and the surrounding environment. The input is then preprocessed with the help of CV Zone package to identify hand gestures by eliminating unwanted elements captured through real time. The CV Zone package specifically helps in locating and providing the coordinate points of the detected hand in the input feed. This segmented hand is then processed through MediaPipe to identify key components of the hand such as fingertips, joints, and [8] wrist. Using these components, MediaPipe generates a skeletal model of the hand for further analysis.

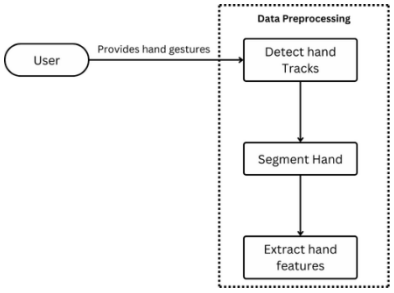

Fig. 2: Data preprocessing stage

3.2 Phase-2: Recognition of Sign from the Preprocessed Hand Gesture:

In Phase 2, the pre-processed hand gesture is fed into the built-in model, i.e. Mobile-Net and Convolution Neural Networks. The input data of size 224x224x3 is fed into Mobile-Net which undergoes depth-wise convolution and point-wise convolution, extracting relevant features, and generating an output of size 7x7x1024. GolabAveragePooling2D is then performed on the obtained output using each channel's average to reduce the feature map to (1, 1024). The output from the GlobalAveragePooling2D is then passed through a dense layer of neurons (512, and 256 neurons) to identify specific features and add dropout to reduce overfitting. The resultant is then passed through the output layer containing 26 neurons to formulate a valid American Sign Language sign prediction.

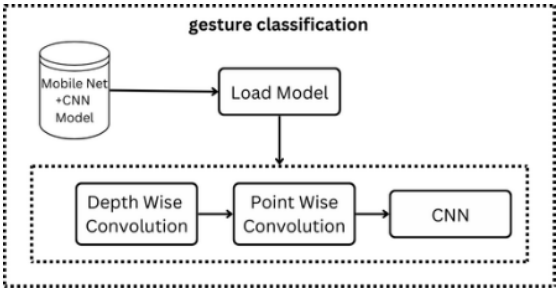

Fig. 3: Gesture classification

3.3 Phase-3: Sentence Construction

In Phase 3, the predicted gesture is interpreted as an alphabet. This new input is then stored in the buffer after confirmation. This process repeats itself generating the next alphabets which are appended to the previous ones resulting in words and in turn sentences. This process is repeated until the end of communication. A word suggestion feature is made available during the word generation process.

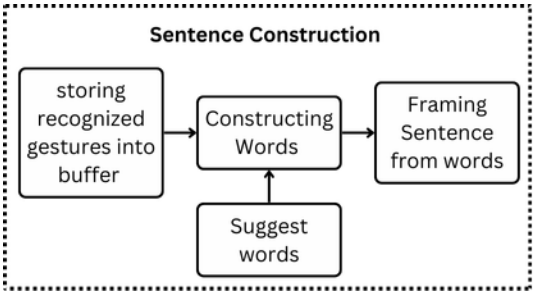

Fig. 4: Sentence construction phase

3.4 Phase-4: Translation Module

In Phase 4, the generated text is converted into any desired language by using Google Translate Application Programming Interface(API). This API with the help of machine translation technology from Google is used to translate the constructed sentences into a desired target language like Telugu, Hindi, Russia, etc., The translated text is then sent back

to the interface which is displayed to the user. Lastly, the translated text is converted into audio by using Google Text To Speech (gTTS) module in the selected language.

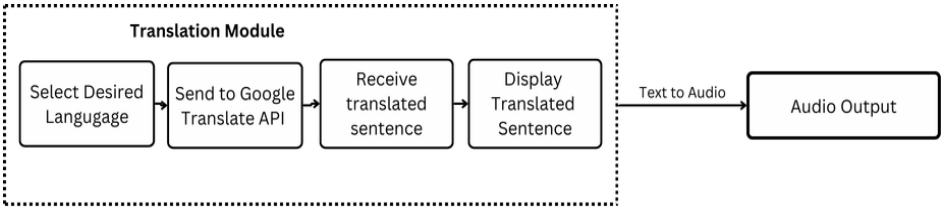

Fig. 5: Translation Module

4.0 Experimental Results:

In the proposed work, the input dataset is yielded by the help of Computer Vision Zone (CVZone) module, which references the ASL to precisely recognize and interpret several hand gestures. The dataset comprises 468 images with over 26 classes corresponding to different ASL signs. The model captures real-time input, and the preprocessing system captures the hand through the CVZone module, as illustrated in Figure 6. The detected hand is then segmented from the given input, and the resulting preprocessed hand gesture is illustrated in Figure 7. Subsequently, the recognized alphabet is presented on the screen, as in Figure 8. With support from recognized alphabets, the model constructs words and sentences, utilizing the word suggestions feature, which is shown in Figure 9. Here, the language selection option is provided to the user as represented in Figure 10. The constructed sentence is then translated into the selected language and is displayed on screen, as shown in Figure 11. Finally, the translated sentence displayed on the screen is converted into audio output, as shown in Figure 12.

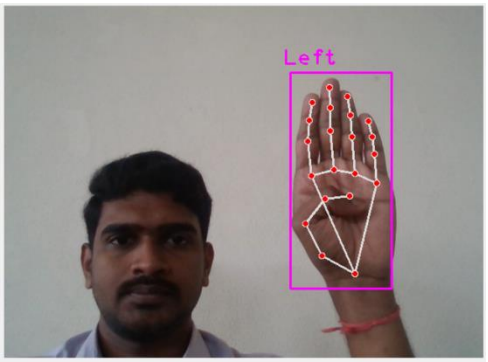

Fig 6: Real-time input through webcam

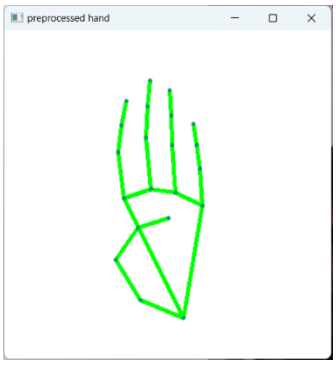

Fig 7: Segmented hand with tracks

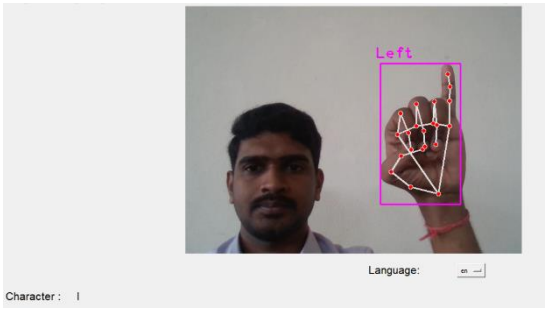

Fig 8: ASL sign recognition

Fig 9: Sentence framing along with word suggestions

Fig 10: Language Selection

Fig 11: Translated Sentence into selected language

Fig 12: Audio Output

5.0 Result Analysis:

The proposed model predicts the signs accurately by grouping similar signs on the basis of highest probability. The Accuracy scores represented in Figure 13 show how reliable the model improves with rise in number of epochs and is graphically represented in Figures 14 and 15. The resulting accuracy of the introduced model is 98.51%. This demonstrates the increased efficiency of the suggested model.

```
Found 3744 images belonging to 26 classes.
Found 937 images belonging to 26 classes.
C:\Users\SURYA\anaconda3\Lib\site-packages\keras\src\trainers\data_adapters\py_dataset_adapter.py:121: UserWarning: Your `PyDataset` class should call `s
uper().__init__(**kwargs)` in its constructor. `**kwargs` can include `workers`, `use_multiprocessing`, `max_queue_size`. Do not pass these arguments to
`fit()`, as they will be ignored.
  self._warn_if_super_not_called()
Epoch 1/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.2520 - loss: 2.6916
C:\Users\SURYA\anaconda3\Lib\site-packages\keras\src\trainers\data_adapters\py_dataset_adapter.py:121: UserWarning: Your `PyDataset` class should call `s
uper().__init__(**kwargs)` in its constructor. `**kwargs` can include `workers`, `use_multiprocessing`, `max_queue_size`. Do not pass these arguments to
`fit()`, as they will be ignored.
  self._warn_if_super_not_called()
117/117 ━━━━━━━━━━━━━━━━━━━━ 248s 2s/step - accuracy: 0.2533 - loss: 2.6865 - val_accuracy: 0.9246 - val_loss: 0.7045
Epoch 2/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 172s 1s/step - accuracy: 0.6656 - loss: 1.1032 - val_accuracy: 0.9526 - val_loss: 0.3322
Epoch 3/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 173s 1s/step - accuracy: 0.8032 - loss: 0.6909 - val_accuracy: 0.9623 - val_loss: 0.2065
Epoch 4/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 174s 1s/step - accuracy: 0.8437 - loss: 0.5459 - val_accuracy: 0.9763 - val_loss: 0.1525
Epoch 5/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 173s 1s/step - accuracy: 0.8726 - loss: 0.4310 - val_accuracy: 0.9709 - val_loss: 0.1381
Epoch 6/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 175s 1s/step - accuracy: 0.8831 - loss: 0.3862 - val_accuracy: 0.9774 - val_loss: 0.1079
Epoch 7/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 175s 1s/step - accuracy: 0.9135 - loss: 0.3147 - val_accuracy: 0.9849 - val_loss: 0.0909
Epoch 8/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 172s 1s/step - accuracy: 0.9154 - loss: 0.2954 - val_accuracy: 0.9871 - val_loss: 0.0809
Epoch 9/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 171s 1s/step - accuracy: 0.9254 - loss: 0.2741 - val_accuracy: 0.9860 - val_loss: 0.0777
Epoch 10/10
117/117 ━━━━━━━━━━━━━━━━━━━━ 173s 1s/step - accuracy: 0.9242 - loss: 0.2471 - val_accuracy: 0.9849 - val_loss: 0.0812
30/30 ━━━━━━━━━━━━━━━━━━━━ 18s 598ms/step - accuracy: 0.9854 - loss: 0.0846
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. We
recommend using instead the native Keras format, e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_model.keras')`.
Validation Accuracy: 98.51%
```

Fig 13: Accuracy Scores

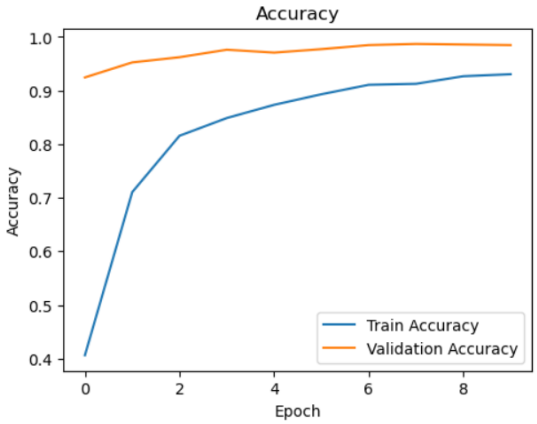

Fig 14: Graphical Visualization of Accuracy

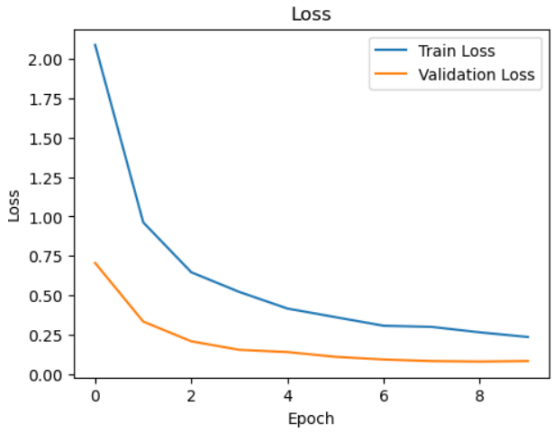

Fig 15: Graphical Visualization of Losses

6.0 CONCLUSION:

This work exhibits a real-time gesture-based recognition system, which marks a significant advancement in enhancing communication accessibility for the deaf and hard-of-hearing community. This model preprocesses data by isolating the hand from the environment and ensuring accurate detection even under noisy and unclear conditions. MediaPipe frameworks are leveraged to create a skeletal representation of the hand to enable extraction of hand features for further processing. Additionally, it contains a word suggestion feature for construction of meaningful sentences. Further, the constructed sentence can be translated into multilingual text and converted to audio which facilitates easy communication between non-verbal and verbal individuals. By implementing the discussed features, the proposed model demonstrates to be efficient over the existing models. Notably, it yeilds an accuracy of 98.51%, outperforming the existing models that have nearly precision rate of 96%, which in turn, proves to acts as a novel method.

References:

[1] Prasanthi Y., B. Keerthana, P. Jahnavi, K. Venkata Rao, and C. Raj Kumar, "Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb," *Int. J. Exp. Res. Rev. (IJERR)*, Special Vol.34, pp. 26-35, 2023.

[2] Kanchan Dabre, Surekha Dholay, "Machine learning model for Sign Language Interpretation using webcam images*," IEEE International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, pp. 317–321, June 2014.

[3] Suneetha Mopidevi, Shivananda Biradhar, Neha Bobberla, Kiran Sai Buddati, "Hand gesture recognition and voice conversion for deaf and Dumb*," E3S Web of Conferences*, vol. 391, Art. No. 01060, June 2023

[4] A. H. Mohammedali, H. H. Abbas, H. I. Shahadi (2022)." Real-time sign language recognition system," *International Journal of Health Sciences*, Vol. 6 No. S4, pp. 10384–10407, 2022. (https://doi.org/10.53730/ijhs.v6nS4.12206)

[5] Areesha Gul, Batool Zehra, Sadia Shah, Nazish Javed, Muhammad Imran Saleem*," Two-way Smart Communication System for Deaf & Dumb and Normal People," 2020 International Conference on Information Science and Communication Technology (ICISCT)*, April 2020.

[6] V. K. R. Vadisena, N. V. S. B. Kalluri, U. H. Ketepalli, C. S. N. T. Dathi, and S. K. Jonnakuti "Hand Gesture Recognition Using Convolutional Neural Networks and Computer Vision," *International Journal of Scientific & Engineering Research*, Vol. 8 No. 1, pp. 2229-5518, July 2021.

[7] Zhibo Wang, Tengda Zhao, Jinxin Ma, Hongkai Chen, Kaixin Liu, Huajie Shao, "Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System, " *IEEE Transactions on Mobile Computing*, Vol. 21, Issue. 7, pp. 2398 – 2410, November 2020.

[8] Uttam Patole , Rutuja Wable , Shubhangi Pawar , Sakshi Shewale , Siddheshwar Kadam ," Gesture-Based Virtual Mouse and Keyboard for Human-Computer Interaction *," International Journal of Creative Research Thoughts (IJCRT),* Vol.12 Issue. 11,pp. b452-b458, Nov 2024.

[9] Shweta Sonajirao Shinde, Dr. R.M. Autee, (2021),"Real-time hand gesture recognition and voice conversion system for deaf and dumb person based on image processing ," *JournalNX - A Multidisciplinary Peer Reviewed Journal*, Vol. 2 Issue. 09, pp. 39–43, Feb 2021.

[10] Ayan Ghosh, Daniel Alonso Paredes Soto, Sandor M Veres and Anthony Rossiter, "Human Robot Interaction for Future Remote Manipulations in Industry 4.0, " *IFAC-PapersOnLine. 21st IFAC World Congress 2020*, Vol.53 Art.2, pp. 10223–10228, Jul 2020.

[11] Jia Min Yen and Jeong Hoon Lim *"*A Clinical Perspective on Bespoke Sensing Mechanisms for Remote Monitoring and Rehabilitation of Neurological Diseases: Scoping Review, " *Sensors (MDPI),* Vol.23 Issue.1, 536, January 2023.

[12] Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation, " IEEE*Access*, Vol.8, pp. 192527 – 192542, October 2020.