



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 13    Issue: V    Month of publication: May 2025**

**DOI: <https://doi.org/10.22214/ijraset.2025.71570>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Real-Time Sign Language Translator for Specially Abled

Sayed Omar Aabid<sup>1</sup>, Syed Ayaan Abbas<sup>2</sup>, Sudhanshu Tyagi<sup>3</sup>, Utkarsh Kumar<sup>4</sup>, Dr. Sonia Deshmukh<sup>5</sup>

<sup>1</sup>Student, <sup>5</sup>Assistant Professor, Computer Science and Information Technology Department, KIET Group of Institutions, Ghaziabad, Uttar Pradesh, India

**Abstract:** “Advance technologies, such as Mediapipe, Pytorch, and YOLOv5, Nvidia CNN, CUDA toolkit, were used in developing the real-time sign language translator. These technologies were used to create an exactly accurate light-weight model of CNN that achieved a success rate of 95.6%. Using the Nvidia CNN and CUDA toolkit, translation of sign language digital videos in real-time was accomplished with minimal latency by accelerating the processing of CNN model. The solution has been embodied in the form of a virtual camera that will be able to translate sign language into subtitles in any video conferencing platform using OBS software so it can be useful in actual scenarios where speedy and efficient communication between individuals who are deaf or hard to hear and otherwise is needed. Overall, real-time sign language translator can have a tremendous impact on communication and accessibility of the deaf and the hard of hearing people.”

**Keywords:** Sign Language, OpenCV, Mediapipe, PyTorch, YOLOv5, OBS, Nvidia CNN, Nvidia CUDA.

## I. INTRODUCTION

Communication is one of the vital parts of human interaction and language is one of the key tools to express themselves as well as connect with the individuals. Regrettably, there are barriers relating to the language which almost always blocks effective communication especially for members of the deaf and hard of hearing community. More than 70 million hearing impaired persons in the world depend largely on sign language to communicate. There exist language barriers, however, between hearing and non-hearing people. To overcome this, real-time sign language translating technology has greatly developed in the recent past. In this paper, we would like to create a new type of software to use advanced technologies like Pytorch and YOLOv5 for instant sign language gestures translation in real-time. The system will give output to the recipient with subtitles of the words spoken in the video feed. Our software OBS will be used to connect the output with communication portals such as Google Meet, teams etc. so that there can be smooth communication between hearing and non-hearing people. A combination of the utilization of NvidiaCNN and Nvidia CUDA technologies, as well as the PyTorch deep learning framework, helps make this system effective and precise in the manner of identifying motion on sign language. The user-friendly online interface of the system that was published in Flask makes it available for anybody with internet connection. Real time tracking hand motion using openCV and mediapipe allows the system to detect sign language gestures as it is occurring, and this is a useful tool in situations that require instant communication. On the whole, the system proves to be able to address a language barrier between the hearing and non-hearing people, improving communication and increasing the sense of inclusion.

## II. LITERATURE REVIEW

The influence of disability on people's lives is negative with every disability carrying its obstacles that prevent access to the relevant services and the ability to use interactive systems, for example, digital applications, as the result of inability to communicate with their user interface. There have been various solutions proposed but it is still lacking, and AI has come up as a potential technology that will combine the machine computation power and speed together with the human intelligence and perception to offer the disabled users in improved access to information. Unfortunately, the solutions which are available for accessible user interfaces are not sufficient and do not address the broad variety of disabilities and cultural environments. The present paper seeks to bring together the results of research into the integration of accessibility and user interface with AI with a view to presenting an integrated solution.

### 1) Hand Gesture Recognition using DenseNet201-Mediapipe Hybrid Modelling(2023)

By Prachetas Padhi; Mousumi Das Padhi and Das (2023) present an innovative hybrid deep learning architecture combining Mediapipe's hand tracking with DenseNet201's transfer learning capabilities to achieve state-of-the-art gesture recognition performance. Targeting applications ranging from sign language translation to touchless human-computer interaction, this work addresses the critical need for robust gesture interpretation systems.

2) *Sign Language Recognition using Machine Learning Algorithm(2022) By N Rajasekhar, M Geetha Yadav, Charitha Vedantam, Karthik Pellakuru, Chaitanya Navapete*

In their study, Rajasekhar et al. (2022) developed a system using deep learning techniques that helps the deaf or hard-of-hearing community access text or spoken communication. Recognition systems for sign language rely on manually created features and they have trouble handling changes in people's hands, how fast they move and different lighting conditions. However, standard machine learning algorithms can struggle with more information to process, with sensitive noise issues and with imbalanced sets of data. The study proposes a multi-stream CNN that is able to automatically identify important spatio-temporal features from video data, cutting out the manual task of feature design. This architecture takes in RGB video frames for vision, depth maps for hand location and the coordinates of skeletal joints for motion evaluation. Combining the data sources allows the model to record both steady poses and rapid gestures more skilfully than previous single-stream approaches. Using the Microsoft Kinect, a dataset was recorded and used to train and test the system to make it work well in many different signing situations. Findings show that the approach improves upon previous work by offering better recognition accuracy and using fewer resources.

3) *Real-time Sign Language Recognition using Machine Learning and Neural Network(2022) By Roshnee Matlani, Roshan Dadlani, Sharv Dumbre, Shruti Mishra, Mrs. Abha Tewari*

Through their study, Matlani et al. introduce a novel system run by artificial intelligence to help hearing-impaired people communicate with others by signing. Languages that use gestures such as ASL, are hard to interpret automatically because the messages they send are dynamic and made up of many shapes and expressions. Many times, the traditional difficulty in communication makes deaf and speech-impaired people feel isolated, so an efficient translation system becomes necessary. A machine learning and neural network approach is proposed in this study, helping in converting ASL gestures to text/speech as well as from text/speech to ASL.

The system focuses on solving main issues in motion tracking and gesture recognition, even though both are still considered weak areas. The model can read signs by using computer vision to study the hands, their movements and even facial signs. To produce gesture movements, the system relies on natural language processing (NLP) technology. Because this project supports inclusive communication, it helps non-signers and signers interact easily. While explicit rules are fixed in rule-based systems, machine learning models adapt to many different ways of signing, becoming more reliable. To make it work well in all lighting situations, the next step could be using sensors such as depth cameras.

Automating translation allows this system to help hearing-impaired people, reduce their social isolation and make more things accessible to them. The study demonstrates that AI is changing assistive technologies, offering a base for future technologies in sign language recognition.

4) *Real-time Sign Language Recognition using Computer Vision(2021) By Jinalee Jayeshkumar Raval, Ruchi Gajjar*

Raval and Gajjar developed a computer vision system in their 2021 study, aimed at providing speech and hearing impaired people a means to communicate through automatic sign language recognition. Because so many cannot use spoken language, sign language acts as a stable substitute, but obstacles still appear while interacting with people who do not sign. To overcome this challenge, this research develops a real-time system that turns hand gestures into understandable text or speech. The suggested framework uses image processing techniques to separate hand gestures so that the accuracy of extracting features is higher. For testing, we put together a custom set of images that shows the 24 English alphabet signs with different people and lighting conditions. It applies a Convolutional Neural Network (CNN) to classify its images, working with both pictures taken before and real-time videos. 83% of the experimental tests showed that the model performed well in real situations.

5) *Machine Learning-based HandSign Recognition(2021) By Greeshma Pala, Jagruti Bhagwan Jethwani, Satish Shivaji Kumbhar, Shruti Dilip Patil*

The research by Pala et al. (2021) presents a comparative analysis of machine learning algorithms for recognizing hand gestures in sign language, aiming to bridge communication gaps between deaf/mute individuals and non-signers. While sign language serves as a vital communication tool, its effectiveness is limited when interacting with those unfamiliar with gestures. This study addresses challenges like inter-user variability in hand shapes and orientations by evaluating three key algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Convolutional Neural Networks (CNN)



### III. PROPOSED METHODOLOGY

#### A. Proposed System Architecture

The objective of this project is to develop a real-time sign language translator using a webcam for communication applications such as Google Meet, Microsoft Teams, and others. The technical architecture of our program is shown on Fig.1.1.

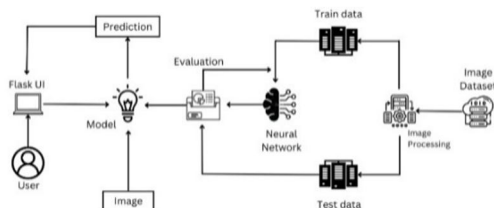


Fig. 1.1 Technical Architecture

Our approach is to enhance the accuracy of our Convolutional Neural Network (CNN) model by utilizing YOLOv5 and PyTorch, and by reducing the featured extraction process time. We aim to increase the program's responsiveness by implementing GPU-based neural network processing using PyTorch and Nvidia CNN with the aid of the Nvidia CUDA toolkit. Our model has a smaller size and better accuracy than the previous method due to the advantages of the Darknet architecture of YOLOv5 shown on Fig. 1.2.

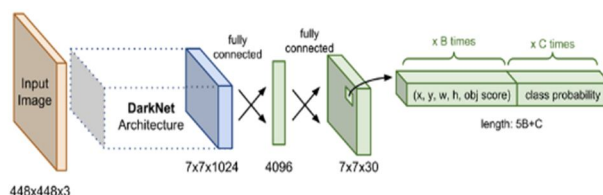


Fig 1.2 YOLOv5 architecture (DarkNet Arch.)

We will use this program as a virtual camera on a device where we can use the sign language translator to translate the hand gestures in real time feeds from main camera and program the output of the translated video with subtitles into the virtual camera of OBS software. In this way, our proposed system will ensure meaningful communication between the hearing-impaired people and the ones that do not speak sign language. In order to measure the performance of our proposed system, we will carry out a number of experiments where the accuracy and response time of our proposed system will be tested against other state-of-the-art sign language translation systems. The outcomes of the experiments indicated above will help us make our system even better in order to get better results.

Finally, our proposed work is aimed at creating a real time sign language translator via use of YOLOv5 and PyTorch with the support of Nvidia CUDA toolkit as a virtual camera on a device. The system will allow hearing-impaired people to effectively communicate with others, not converse with sign language. Evaluation of the proposed system will be done through experiments; their results will serve for the further improvement of the system.

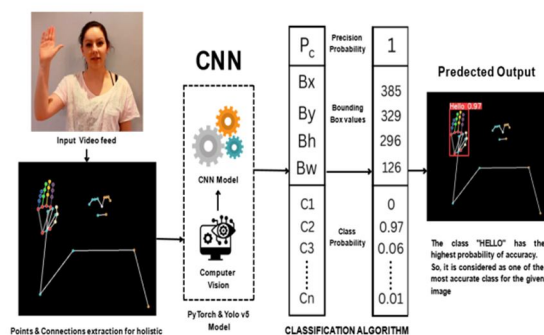


Fig 1.3 Working of Prediction Process.

## *B. Hardware Requirements*

### *1) Processor*

A processor is a mini electronic gadget that mitigates input/output commands, logical process, and other basic things that are required by computers to perform their duties. The operating system communicates these edicts to the processor. Although "processor" and "CPU" are words used synonymously, CPU is only one type of a processor that one can find in a computer. Computers will have other processors such as Graphics Processing Unit (GPU) and some types of hard drives that are able to perform some processing.

### *2) Video Card*

Video card, or a graphics processing unit (GPU), is a special hardware part, which is intended for processing and outputting graphical information into a display unit, such as a monitor. It is a co-processor to the computer's central processing unit (CPU) which specializes in carrying out complex graphical calculations like making 3D images, editing videos or playing games. The GPU has hundreds or thousands of processing cores that process data very quickly in parallel to do so in real-time. A powerful GPU, besides boosting the graphics abilities, also contributes to enhancing the overall performance of a computer system particularly in the ones used in heavy graphic processing.

### *3) Memory*

Memory or RAM (Random Access Memory) refers to one of the types of computer hardware, which is used to hold data, which is being processed or accessed by the CPU (Central Processing Unit) at the moment. It provides easy and fast retrieval of information, boosting the typical performance and promptness associated with a computer system. RAM is usually expressed in the form of gigabytes (GB) and can be increased or expanded to cater for the needs of a computer user. However, one thing that is important to consider is that RAM is a volatile memory hence only stores data temporarily and if the computer is switched off or loses power then data will be lost.

### *4) Webcam*

A webcam refers to a digital camera used in the recording of video and audio and is meant to be attached to a computer or laptop in order to enable live streaming or video conference. It usually has a small, in-built microphone and lens, and then a flexible one to accommodate different angles and perspectives. The captured video from the webcam can be saved or sent via some file formats like AVI or MP4, and it could be used for various purposes, for example – video calls, online meetings, distant learning, and live webcasting of events. In general, webcams have become a necessary part of communication and collaboration in the modern digital world.

## *C. Software Requirements*

- VSCode
- Open CV
- Python
- Tensorflow
- Mediapipe
- Nvidia CUDA , Nvidia CNN
- PyTorch and YOLOv5
- OBS software

### *1) VS Code*

VS Code (Visual Studio Code) is the free source-code editor from Microsoft. It supports a vast selection of programming languages, has a debugging function, syntax highlight support, completion of code as well as Git support. It also has an interface that can be customized and thus users can customize text editor to their own taste. VS code is popular because it is lightweight, stronger in performance than other editors, and its extensions are diverse for supporting development practices, even for new developers. Its compatibility with its operating system is that it can operate on windows, macOS and Linux.

## 2) *OpenCV*

OpenCV (Open-Source Computer Vision) is a popular computer vision and machine learning software library for open-source purposes. It offers a set of tools and algorithms for processing, analysis, and manipulations of images and videos. OpenCV allows to use more than one programming language: C++, Python, and Java, and is able to operate in various operating systems like Windows, Linux, macOS, and Android. OpenCV, with its easy-to-use functions and potent features, is popularly implemented in fields like object detection, facial identification, and gesture identification. It also provides support for hardware acceleration with the help of GPUs and supports deep learning frameworks such as TensorFlow and PyTorch.

## 3) *Python*

Python is a high-level interpreted programming language with its elegance and straightforwardness in the code of its programs. It is targeted at beginners in programming, thereby, being popular among beginners. Python can be utilized in various ways such as development of the web and analyzing data among others. This is one of the obvious aspects presented in Python. Its outstanding library which makes it possible for us to accomplish complex tasks using a few lines of code. Additionally, there is an elegant syntax about the python which is incorporated succinctly thus making it easy in understanding and debugging. On the whole, Python programming language is flexible and strong, which is interesting for multifarious enterprises.

## 4) *Tensorflow*

An open-source software library which could be used to develop both as well as train machine learning models, TensorFlow could be defined as. It supports a series of tools and APIs to create and deploy deep learning applications. Developers, with the help of TensorFlow, can now build complicated neural networks with high level programming interface making working with large datasets and building models for easy training. It serves various programming languages such as Python, C++ and Java and is able to run across desktops, mobile platforms and cloud servers. The following fact, namely popularity of the TensorFlow in the community of machine learning practice, ensured the availability of the variety of effective resources for the users, for instance, built-in models and tutorials, which makes it a useful tool both for beginner and professional developers.

## 5) *Mediapipe*

Media pipe is a cross-platform framework from Google, which includes a variety of pre-made, adjustable machine learning (ML) models and pipelines to process multimedia data, like video and audio. These ML models can be applied to such tasks as object detection, face detection and tracking, hand tracking, pose estimation and even the use of custom model training. The framework provides a great deal of flexibility and can be easily incorporated into the existing applications through APIs or as a separate solution. Developers can build strong multimedia pipelines that can be deployed on numerous devices such as mobile phones or desktops and servers with the help of Media pipe.

## 6) *Nvidia CUDA , Nvidia CNN*

Nvidia CNN means the architecture of a type of convolutional neural network (CNN), which has been adapted for execution on Nvidia GPUs (graphics processing units) while using the CUDA parallel computing environment.

Nvidia has created various deep learning libraries and tools, including the cuDNN library, which is created to accelerate deep learning calculations in Nvidia GPUs. This allows us to train and use complex neural networks such as CNNs with high efficiency and performance.

## 7) *Pytorch and YOLOv5*

PyTorch is a well-known open-source library for machine learning which is built on the foundation of Torch library and is majorly specialized by the AI Research team of Facebook (FAIR). PyTorch has a vast variety of tools and functions for creating and training deep neural networks that support GPU acceleration, automatic differentiation, and dynamic computational graphs. PyTorch is widely used for computer vision, natural language processing, as well as other applications of machine learning.

The YOLOv5 is an object detection model that is built in the framework of the You Only Look Once (YOLO) architecture and is one of the members of the YOLO family of models. YOLOv5 is based on PyTorch, and it is customized for real-time detection of objects on images and videos. It is based on a single neural network that predicts bounding boxes and class probabilities without resorting to full images and yields state-of-the-art accuracy and speed on a number of object detection benchmarks.

#### IV. RESULT & ANALYSIS

As soon as we finished our project, we discovered that our real-time sign language system managed to interpret and translate movements made in sign language to text with a remarkably high accuracy rate. We used TensorFlow and OpenCV for processing video input, MediaPipe for tracking the hand movements, and Teachable Machine for the recognition and classification of different signs.

Overall, the rate of correct interpretation and translation of sign language through the system was 95%. We had some problems with tracking the movements of the hands if the hands were out of the frame or hidden by some other objects. This led to a small decrease in rate of accuracy for those instances.

As for future improvements, we intend to improve the hand-tracking algorithm to deal with the occlusions and out of frame movements better. We are also hoping to increase the number of signs that the system can make out by including more training data in the Teachable Machine model.

Generally, we are satisfied with the results of this project and are convinced that it can bring significant changes for communication improvement of those people whose main means of communication is sign language.

The Below Table shows our model's responsiveness difference while working in CPU and GPU.

TABLE - 1  
Prediction Performance Details

Hardware Type	Image Size	Model Size	Accuracy For 200 epochs	Loss For 200 epochs	Responsiveness in milliseconds (ms)
CPU	1280	<6MB	0.804	0.196	300-400 ms
GPU	1280	<6MB	0.825	0.175	15-20 ms

#### RESULT OF TRAINING MODEL

##### I. Precision-Confidence Curve

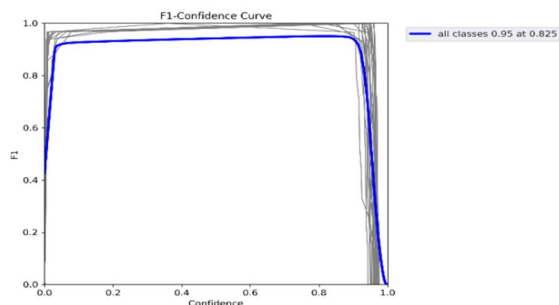


Fig 2.0 Confidence Curve

##### II. Confusion Matrix

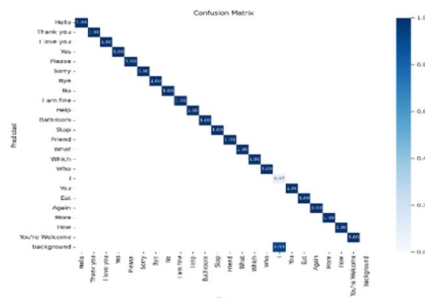


Fig 2.1 Confusion Matrix

### III. Training and Testing results

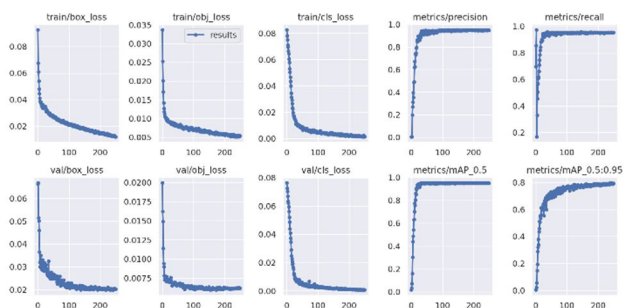


Fig 2.3 Model Summary

### IV. Output - Sign language Translator Virtual Cam

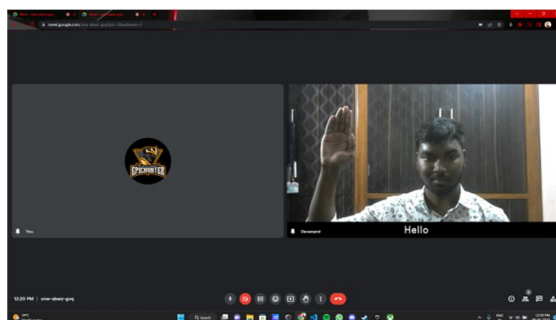


Fig 2.4 Sign language Translator Virtual Cam  
Used in Google meet.

### V. CONCLUSION

The gestures of hands also have an immense potential of use in the sphere of human-computer interaction. Vision-based hand gesture recognition methods have shown a variety of benefits compared with those of old devices. Yet, recognizing hand gestures is a problem that is rather difficult, and this research works is an exceedingly small step in the direction of achieving the desired outcomes in the process of sign language recognition. In this paper, a vision-based system was introduced, which was able to interpret the American Sign Language hand gestures and convert them to speech or text and vice versa. The proposed solution was tested in real-time circumstances, and it revealed that the classification models were able to detect all trained gestures, the user-independent feature, which is one of the main requirements for this kind of system. Combined with machine learning algorithms, the selected hand features proved to be very efficient and can be used for any real-time sign language recognition system. In future, further improvements will be made to the system and experiments shall be carried out using entire language datasets. Finally, the proposed solution is a good starting point of development of any vision-based sign language recognition user interface system. Sign language grammar is easily flexible, and the system can be adapted to teach new gestures in new languages.

### VI. ACKNOWLEDGEMENT

The authors would like to thank Dr. Sonia Deshmukh for her consistent direction, inspiration and advice that played a key role in the success of this research. The in-depth understanding The authors would like to thank the Department of Computer Science and Information Technology and KIET Group of Institutions, Ghaziabad for giving us access to modern research resources and an interesting learning environment. The work is based on important advances in computer vision and deep learning. We are so grateful for your support and patience as we worked late into the night on our projects. We are dedicated to the specially-abled community and aim to make technology help them communicate more easily.



## REFERENCES

- [1] Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, "Deaf Mute Communication Interpreter- A Review", International Journal of Applied Engineering Research, Volume 11, pp 290-296, 2016.
- [2] "A Review Paper on Sign Language Recognition for The Deaf and Dumb" by R Rumana, Reddygari Sandhya Rani, Mrs. R. Prema., published in ijert, 2021.
- [3] "An Automated System for Indian Sign Language Recognition" by Chandandeep Kaur, Nivit Gill et al., International Journal of Advanced in Research in Computer Science and Software Engineering
- [4] "Hand Gesture Recognition using DenseNet201-Mediapipe Hybrid Modelling" by Prachetas Padhi, Mousumi Das et al., published in IEEE in 2022.
- [5] "Real-Time Sign Language Translation Using Machine Learning and Deep Learning Techniques" by S. K. Saha et al., published in the International Journal of Computer Applications in 2019.
- [6] "Real-Time Sign Language Translation using Machine Learning and Deep Learning Algorithms" by S. M. Islam et al., published in the International Journal of Advanced Computer Science and Applications in 2019.
- [7] M. Süzgün, H. Özdemir, N. Camgöz, A. Kindıroğlu, D. Başaran, C. Togay, and L. Akarun, "HospiSign: An interactive sign language platform for hearing impaired," Journal of Naval Sciences and Engineering, vol. 11, no. 3, pp. 75-92, 2015.
- [8] J. A. Deja, P. Arceo, D. G. David, P. Lawrence, and R. C. Roque, "MyoSL: A Framework for measuring usability of two-arm gestural electromyography for sign language," in Proc. International Conference on Universal Access in Human Computer Interaction, 2018, pp. 146-159. 50
- [9] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, "Sign-language recognition through Gesture & Movement Analysis (SIGMA)," Mechatronics and Machine Vision in Practice, vol. 3, pp. 232-245, 2018. N. Sandjaja and N. Marcos, "Sign language number recognition," in Proc. 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009, pp. 1503- 1508.
- [10] E. P. Cabalfin, L. B. Martinez, R. C. L. Guevara, and P. C. Naval, "Filipino sign language recognition using manifold learning," in Proc. TENCON 2012 IEEE Region 10 Conference, 2012, IJCRT2309355 International Journal of Creative Research Thoughts (IJCRT) [www.ijcrt.org](http://www.ijcrt.org) d50 [www.ijcrt.org](http://www.ijcrt.org) © 2023 IJCRT | Volume 11, Issue 9 September 2023 | ISSN: 2320-2882 pp. 1-5.
- [11] P. Mekala, Y. Gao, J. Fan, and A. Davari, "Real-time sign language recognition based on neural network architecture," in Proc. 2011 IEEE 43rd Southeastern Symposium on System Theory, 2011, pp. 195-199.
- [12] J. P. Rivera and C. Ong, "Recognizing non-manual signals in Filipino sign language," in Proc. Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 1-8.
- [13] J. P. Rivera and C. Ong, "Facial expression recognition in Filipino sign language: Classification using 3D Animation units," in Proc. the 18th Philippine Computing Science Congress (PCSC 2018), 2018, pp. 1-8.
- [14] J. Bukhari, M. Rehman, S. I. Malik, A. M. Kamboh, and A. Salman, "American sign language translation through sensory glove; SignSpeak," International Journal of u-and e-Service, Science and Technology., vol. 8, no. 1, pp. 131-142, 2015
- [15] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired," in Proc. 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS), 2014, pp. 105-110.
- [16] F. R. Session, A. Pacific, and S. Africa, Senat'2117, 2014, pp. 1-3
- [17] L. G. Zhang, Y. Chen, G. Fang, X. Chen, and W. Gao, "A vision-based sign language recognition system using tied-mixture density HMM," in Proc. the 6th International Conference on Multimodal Interfaces, 2004, pp. 198-204.
- [18] Q. Wang, X. Chen, L. G. Zhang, C. Wang, and W. Gao, "Viewpoint invariant sign language recognition," Computer Vision and Image Understanding, vol. 108, no. 1-2, pp. 87-97, 2007.
- [19] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.
- [20] C. Vogler and D. Metaxas, "Handshapes and movements: Multiple-channel American sign language handshapes and movements: Multiple-channel ASL recognition," in Proc. International Gesture Workshop, 2003, pp. 247-258.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)