

---

## HAND GESTURE RECOGNITION FOR DEAF AND DUMB USING CNN TECHNIQUES

K. Steephen<sup>1</sup>, N. Devi Sri<sup>2</sup>, C.G. Karunya Shree Nee<sup>3</sup>, K. Kavipriya<sup>4</sup>, R. Keerthana<sup>5</sup>.

<sup>1</sup>Assistant Professor, Department Of Computer Science & Enginerring, Vivekanandha College Of  
Technology For Women, Namakkal, Tamil Nadu, India.

<sup>2,3,4,5</sup>Department Of Computer Science & Enginerring, Vivekanandha College Of Technology For Women,  
Namakkal, Tamil Nadu, India.

DOI: <https://www.doi.org/10.58257/IJPREMS31190>

---

### ABSTRACT

Sign language is a type of communication used by people with hearing and speech impairments. Disabled People utilise these sign language gestures as a form of non-verbal communication to convey their own feelings and ideas to other regular people. It can be quite difficult to communicate with persons who have hearing loss. Since Deaf and Mute persons communicate via hand gestures, normal people have difficulty understanding the signs they make. Systems that can identify various indications and provide information to common people are therefore necessary. However, these common people find it difficult to interpret their expression, thus qualified sign language experts are required during medical and legal appointments, as well as educational and training sessions. The demand for these services has grown during the previous few years. Other types of services have been established, such as video remote human interpretation using a high-speed Internet connection. These services offer a simple sign language interpretation service that can be used and is beneficial, but has significant drawbacks. We can employ artificial intelligence technology to examine the user's hand with finger detection in order to solve this issue. We may construct the vision-based system in real-time situations using the proposed technology. Convolutional neural network algorithm, a deep learning method, is then used to classify the sign and provide a label regarding recognised sign.

**Keywords:** Sign language converter, Artificial Intelligence, Text to speech Conversation, Indian Sign Language, Image Processing.

---

## 1. INTRODUCTION

The process of turning the user's motions and signs into text is known as sign language recognition. It facilitates communication for persons unable to communicate with the wider public. Unprocessed images and videos are converted into text that can be read and comprehended by using image processing techniques and neural networks to map the motion to applicable text in the training data. Dumb people are usually not allowed to have normal conversations with other people in society. Because most individuals only recognise a small number of them, they occasionally struggle to connect with commoners through gestures. Due to their inability to communicate verbally due to hearing loss or deafness, people with these conditions typically employ some form of visual communication. The fundamental type of Local sign language is used. It has syntax and vocabulary much like other languages, however it communicates mostly through visuals.

## 2. RELATED WORKS

### 2.1 Existing Problem

#### 2.1.1 Title: A Study On Arabic Sign Language Recognition For Diferently Abled Using Advanced Machine Learning Classifiers

**Author: Mohammed Mustfa, 2021.** -An automated approach for translating sign language could greatly enhance communication between sign language users and those who might not understand it. Around 70 million people use sign language globally. Sign language refers to nonverbal communication that makes use of additional body components. Face expressions, along with hand, eye, and lip movements, are utilised in sign language communication to convey information. People who have difficulty hearing or speaking frequently use sign language as a means of daily communication. However, the highly complex computer translation of sign language was able to reveal how the hands' uneven size, form, and posture in an image. SLRs can be employed either based on the image or the sensor. The primary benefit of image-based frameworks is that they do not require users to employ expensive equipment. Large computations are required for the pre-processing procedure in any circumstance. Instead of relying just on cameras, sensors frameworks use gloves equipped with sensors Similar to spoken language, sign language is not limited to one place or area. Around the world, it is trained differently (Shin et al. 2019). It is also known as Arabic Sign Language (ArSL), Chinese Sign Language, American Sign Language, and African Sign Language. Contrary to sign languages in Europe and America, India does not have a standardised sign language with significant changes.

---

However, Coimbatore's Vivekananda University has produced an ISL dictionary for the Ramakrishna Missions. In Indian Sign Language (ISL), there are now 2037 signs accessible. Similar to how sensor glove-based and vision-based SLR model types are used. Methods used in recent SLR research can be categorised as contact-based and vision-based. The contact-based technique and clients include physical touch between sensing devices. It frequently makes use of an instrumented glove that measures the position, extension, direction, and angle of the executed sign using electromagnetic, inertial estimation, or electromyography.

### 2.1.2 Title: Sign Language Transformers: Joint End-To-End Sign Language Recognition And Translation.

**Author: Necati Chian Camg"Oz, 2021.**-According to past studies on sign language translation, the translation is enhanced by having a mid-level sign gloss representation that effectively distinguishes the numerous signs. Performance considerably in fact, for the state-of-the-art in translation to work, gloss level tokenization is required. We outline a novel architecture based on transformers that can simultaneously learn Continuous Sign Language Recognition and Translation and be trained from beginning to end. The recognition and translation problems are combined into a single, unified architecture using a Connectionist Temporal Classification (CTC) loss to achieve this. With no need for ground-truth timing data, our cooperative strategy solves two related sequence-to-sequence learning challenges while also yielding significant performance benefits. This includes both manual and non-manual traits including head, shoulder, and torso movement as well as manual traits like hand shape, movement, and posture. The goal of sign language translation is to either translate written text into a movie of signs or to translate written text into an equivalent spoken language sentence. A video of the continuous sign being performed. The term "articulators" is used by linguists to describe these channels in the context of computer vision, where a significant percentage of the latter work is conducted. Instead of offering a comprehensive translation into a spoken language counterpart (Sign Language Translation, or SLT), word embedding with spatial embedding has focused on comprehending the order of sign glosses. This distinction is essential because the grammatical structures of spoken and sign languages differ greatly. Just a few of these variants include word order variations, the use of many channels to communicate information simultaneously, and the use of direction and space to denote the relationships between objects.

### 2.1.3 Title: Sign Language Recognition Systems: A Decade Systematic Literature Review.

**Author: Ankita Wadhawan, 2020.**-Speak languages use the "vocal-auditory" channel since they are uttered with the lips and heard with the ear. Additionally, all writing systems derive from or are representations of spoken languages. Sign languages (SLs) are distinctive because they make use of the "corporal-visual" channel, which is produced by the body and seen by the eyes. Although SLs are frequently utilised in the deaf communities, they are not recognised worldwide. Because deaf people can spontaneously come together and interact with one another everywhere, they are regarded as natural languages. The vocabulary and grammatical structures of SLs are distinct from those of spoken languages. The internal structure of the signals that the deaf employ is identical to that of spoken words. The sounds that make up SL signs are not as numerous as those that make up the hundreds of thousands of English words. A predetermined assortment of gestural traits. As a result, signals are not fully formed motions but rather can be studied as a group of significant language features. These features are combined to form a gloss, which is the fundamental part of an SL and the most accurate depiction of a sign's meaning. Similar to spoken languages, sign languages (SLs) have a set of grammatically flexible rules that are applicable to both manual and non-manual parts. To compose phrases in an SL, signers use both of them simultaneously (and frequently with a flexible temporal structure). Depending on the situation, a specific element could be the most crucial factor to take into account while reading a gloss. It can alter the meaning of a verb, offer spatial and chronological context, and discriminate between objects and living creatures. Sign language recognition (SLR), a technique, can extract a signer's glosses from video recordings. Despite the fact that much work has been done, there has been very little thorough experimental research on the topic of SLR. The majority of articles also don't provide their source code or include results from all of the available datasets. As a result, it is difficult to replicate and evaluate experimental results in the field of SL.

### 2.1.4 Title: A Comprehensive Study On Sign Language Recognition Methods.

**Author: Nikolas Adaloglou, 2020.**-Sign language is widely used by those who are deaf-dumb to communicate. Just a variety of hand gestures made with different hand shapes, movements, and orientations, along with facial expressions, make up a sign language. Children make for 34 million of the 466 million individuals worldwide who have hearing loss. People who identify as "deaf" are hearing-impaired or completely deaf. They utilise sign language to converse. People all across the world utilise a variety of sign languages. When compared to spoken languages, they are relatively few in number. The paucity of datasets in the current system, combined with regional variations in sign language, has limited efforts in finger motion detection. The current study intends to take the first fundamental step towards employing Indian sign language to overcome the communication gap between hearing-impaired and deaf persons. The successful expansion of this effort to words and everyday expressions may not only help the deaf and

dumb connect with the outside world more quickly and easily, but it may also aid in the development of autonomous systems that can understand them and provide assistance. Due to a dearth of standardised datasets, Indian Sign Language falls behind its American Counterpart in study. A process known as sign language recognition (SLR) can be used to infer a signer's glosses from video recordings in addition to the inherent difficulties of human motion analysis (such as variations in the participants' appearances, the characteristics of the human silhouette, and the execution of the repetition of operations, the presence of obstructions, etc.). Despite the fact that much work has been done, there has been very little thorough experimental research on the topic of SLR. The majority of articles also don't provide their source code or include results from all of the available datasets.

### 2.1.5 Title: Transferring Cross-Domain Knowledge For Video Sign Language Recognition

**Author: Dongxu Li, 2020-**Word-level sign language recognition (WSLR), a core goal of sign language interpretation, attempts to facilitate communication for the deaf. However, WSLR is exceedingly challenging since it calls for fast body motions, expressive face expressions, and intricate, subtle hand gestures. Deep learning techniques have recently been used to demonstrate Isolated Sign Words Web News Sign Words Localizer. In order to transmit information from web news signs to WSLR models, our model learns domain-invariant properties. The example frames in the picture are recognised by our model as the signature that best conveys the gesture and benefits of the WSLR task. Even though the largest current datasets only contain a few instances—for example, 10 to 50 on average per word—annotating WSLR datasets calls for domain-specific expertise. Compared to common video datasets for action learning and recognition, for instance, this is considerably less. The insufficient training data for the sign recognition challenge may lead to overfitting or negatively impact WSLR's performance in other ways. models in realistic conditions. On the other hand, WSLR could benefit from the abundance of easily accessible news videos with subtitles that are accessible online. Despite the presence of sign news videos, it can be challenging to apply this information to WSLR. First, there are only flimsy labels for the existence of indications in the subtitles rather than any annotations of chronological location or categories. These labels are also quite loud. In order to improve the performance of WSLR models, we present a method in this paper for transferring cross-domain knowledge from news signs to them. To be more precise, using a base WSLR model and a sliding window technique, we first create a sign word localizer to extract sign words. Following that, we recommend jointly coarse-aligning two domains. developing a classifier using both isolated and recent indicators. After receiving the representations of the coarsely-aligned new words, we compute and store the centroid of each class in a separate memory known as the prototype memory.

**2.2 Problem Statement Definition-**Sign language is a popular form of communication for people who are deaf-dumb. Simply put, a sign language is made up of a variety of hand gestures that are made using different hand shapes, movements, and orientations as well as facial expressions. Children make up 34 million of the world's 466 million hearing-impaired persons. Those who consider themselves "deaf" are hearing-impaired or completely deaf. Using sign language, they converse. Numerous sign languages are used by people all over the world. Comparing them to spoken languages, they are extremely rare. The paucity of datasets in the current system, combined with regional variations in sign language, has limited efforts in finger motion detection. The current study intends to take the first fundamental step towards employing Indian sign language to overcome the communication gap between hearing-impaired and deaf persons. The successful expansion of this effort to words and everyday expressions may not only help the deaf and dumb connect with the outside world more quickly and easily, but it may also aid in the development of autonomous systems that can understand them and provide assistance. Due to a dearth of standardised datasets, Indian Sign Language falls behind its American Counterpart in study.

## 3. PROPOSED SOLUTION

The camera is utilised as an input in computer vision-based gesture recognition, and image processing is done before recognition. After that, neural network approaches and the region of interest algorithm are used to recognise the processed motions. A basic flaw in a vision-based sign language identification system is that the process of capturing images is delicate to several environmental elements, such as camera placement, background conditions, and lightning sensitivity. In contrast, it is a more practical and cost-effective method of information gathering than utilising a camera and tracker. However, for greater accuracy, camera data is combined with neural network methods like the Hidden Markov Model.

**3.1. Problem Solution Fit -**Deaf-dumb people frequently use sign language as a form of communication. Simply put, a sign language is a collection of different hand motions made with different hand shapes, movements, and orientations together with facial expressions. Children make up 34 million of the 466 million persons with hearing loss in the globe. Those who consider themselves to be "deaf" have very little to no hearing. They use sign language to communicate. People utilise a variety of sign languages all throughout the world. When compared to spoken languages, they are extremely rare. The camera is utilised as an input in computer vision-based gesture recognition,

and image processing is done before recognition. After that, neural network approaches and the region of interest algorithm are used to recognise the processed motions. A basic flaw in a vision-based sign language identification system is that the process of capturing images is delicate to several environmental elements, such as camera placement, background conditions, and lightning sensitivity. In contrast, it is a more practical and cost-effective method of information gathering than utilising a camera and tracker. In contrast, camera data is combined with neural network methods like the Hidden Markov Model for increased.

#### 4. METHODOLOGIES

**4.1.1 Hand Image Acquisition:** In daily life, the hand gesture is a natural form of communication that is typically exclusively employed by those who have some difficulty hearing or speaking. However, there are numerous application situations for a gesture-based human computer interface system. We can input hand photos captured by a real-time camera in this module. The system can be connected to the built-in camera. Since many years ago, gesture recognition has gained popularity. To accomplish gesture recognition today, essentially two methods are used. One is based on electromagnetic devices that can be worn while working, such as customised gloves. Computer vision is used as an alternative. The former is mostly employed in the motion picture business. Although it works effectively, it is expensive and not suitable for all environments. Image processing is involved in the latter. However, the ability to recognise gestures on the basis of features derived by image processing is only moderately effective shot with a web camera is a hand image. Web cameras are used to record human-generated hand gestures and preserve the resulting image in memory. For storing images in memory, utilise the Python Framework package.

**4.1.2 Binarization-**One of the key problems in the area of computer vision and image processing with the goal of detecting changes in image sequences is background subtraction. Any technique that enables the foreground of an image to be extracted for subsequent processing (object recognition, etc.) is known as background removal. Because the areas of interest in an image are typically the objects in the foreground of the image (people, automobiles, text, etc.), many applications just need to know the information about changes in the scene. This technique may be used for object localisation after the stage of picture pre-processing (which may include image denoising, post-processing like morphology, etc.). To distinguish these changes occurring in the foreground from the backdrop, one must detect the foreground. It is a collection of methods that frequently performs real-time analysis on video sequences that are captured by fixed cameras. Every detection method relies on modelling the image's background, or setting the background and looking for changes. When the background incorporates forms, shadows, and moving things, it can be exceedingly challenging to define it. It is expected while defining the background that the stationary items' colour and intensity may change over time. There are many different scenarios where these strategies can be used. Sequences with a great deal of variation may include pictures with drastically diverse lighting, interiors, exteriors, quality, and sounds. Systems must be able to process in real time as well as be adaptable to these changes. They use methods to separate the foreground from the background of an image. To assign the values to the background and foreground, use the binarization technique. Real-time settings allow for the identification of foreground pixels.

**4.1.3 Region Of Finger Detection** -The division of a digital image into various pieces is referred to as segmentation. In other terms, segmentation refers to the division of pixels into various groups. More specifically, image segmentation is the process of assigning labels to each pixel in an image so that pixels with the same label share specific visual characteristics. Image analysis, object representation, visualisation, and many other image processing activities frequently need the segmentation of a picture into meaningful components. However, segmenting a satellite image into sections (groups) with various textures is a challenging challenge. One cannot know in advance what ways a satellite image contains textures, how many there are, or which areas have which textures. Both supervised and unsupervised segmentation strategies can be used to carry out the monitoring task. A region of interest (ROI) is a portion of an image or dataset that has been selected for a specific objective. To put it another way, a region of interest (ROI) is a section of a picture that has to be filtered or subjected to another process.

**4.1.4 Classification Of Finger Gestures** -Artificial Neural Networks (ANNs) have the ability to learn, which allows them to be trained to spot patterns, locate answers, predict the future, and classify data. It is commonly known that CNN is utilised for activities involving traffic. The connections between each of the many computing components in a neural network as well as the strength or weight of these connections determine how the network learns and behaves. By training the network in accordance with a predetermined learning rule until it successfully completes the required task, these weights can be automatically modified. CNN is a machine learning technique that employs supervised learning and known data, commonly referred to as training data. CNN uses these well-known properties to aid with prediction. The core elements of a training dataset are input data and their response values. The optimal approach is to employ larger training datasets in order to have stronger prediction power and the capacity to generalise for numerous new datasets. With the use of a convolutional neural network technique, the fingers may be categorised. A popular



technique for training artificial neural networks to reduce the goal function is CNN. It is a generalisation of the delta rule and a supervised learning technique. A dataset representing the desired outcome for several inputs, serving as the training set, is necessary. It operates well in feed-forward networks (networks without feedback or simply without loop connections).

**4.1.5 Sign Recognition** -Every motion in sign language has a specific meaning, making it a well-organized kind of gesture coding. For those who are deaf, sign language is their only form of communication. Many approaches have been created as a result of science and technology advancements, both to lessen the difficulty of deaf people and to apply it in various industries. Label the signs with a higher accuracy rate using the classification of sign features.

**4.2 Non-Functional Requirements-Usability**

Users must be able to access the system using a web application on a computer. The system's user interface is a web application. The system is simple because it is user-friendly.

**Availability**

The system is used around-the-clock, 365 days a year, and is completely accessible to the user. The system must function seven days a week, twenty-four hours a day.

**Scalability**-Scalability is the ability of a system to adjust in response to changes in application and system processing demands without significantly affecting its performance or cost.

**Security** -A security requirement is a declaration of essential security functionality that guarantees the fulfilment of one of numerous security attributes of software.

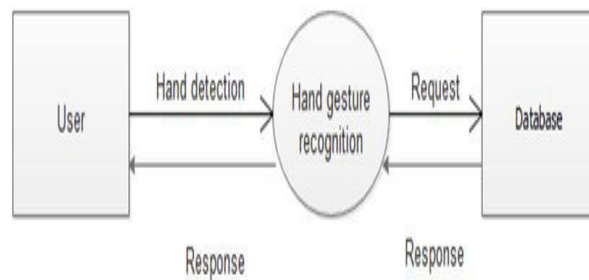
**Performance** -Depending on whether or not the application has undergone any updates, the information is updated. Within two seconds of the member's request submission, the system must react to them. When processing big amounts of data, the system should be permitted to take extra time. Responses to information requests must display on the screen within 5 seconds.

**Reliability** -Due to the value of data and the harm that inaccurate or incomplete data can cause, the system must be completely dependable. It will function every day of the week. Every single day.

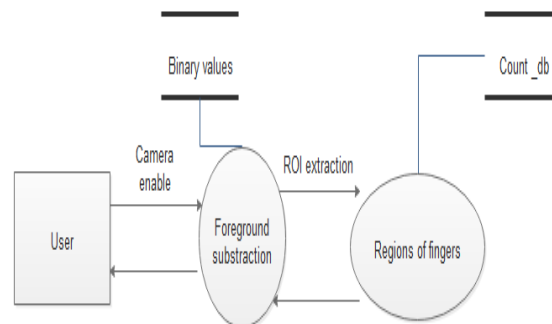
**5. PROJECT DESIGN**

**5.1 Data Flow Diagrams** -A two-dimensional diagram known as a data flow diagram demonstrates how data is transmitted and processed in a system. Each data source is identified, along with how it interacts with other data sources to produce a common result, in the graphical representation. To create a data flow diagram, a person has to know what the external inputs and outputs are, how they link to one another, and how to visually depict these connections and the outcomes they produce.

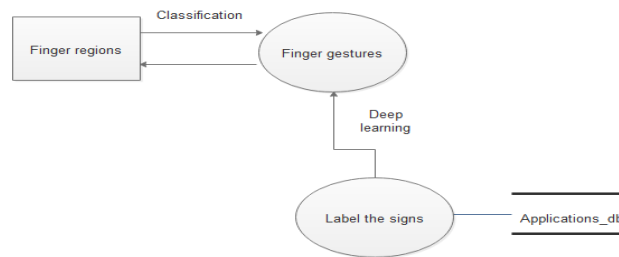
**Level 0:**



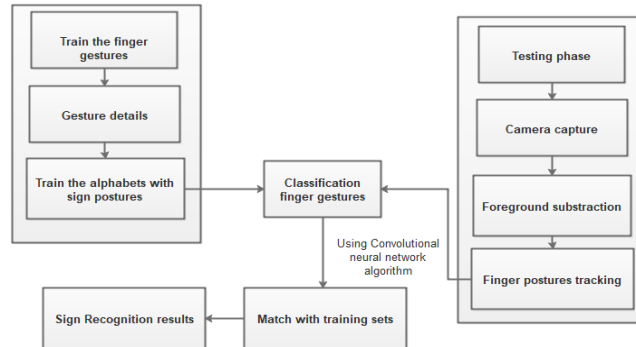
**Level 1:**



**Level 2:**



**5.2 Technical Architecture**



**6. TESTING**

**6.1 testcase:**In order to ascertain whether a feature of an application is functioning properly, a test case has components that explain input, action, and an expected result. A test case is a collection of guidelines on "HOW" to validate a specific test objective or aim, which when followed will reveal whether or not the system's expected behaviour has been satisfied.

Characteristics of a good test case:

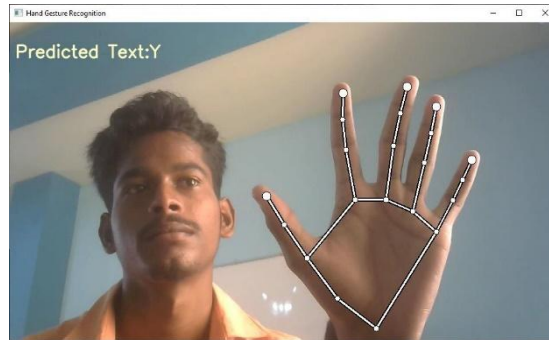
- Accurate: Exacts the purpose.
- Economical: No unnecessary steps or words.
- Traceable: Ability to be tracked back to specifications.
- Repeatable: Enables repeated administration of the test.
- Reusable: Can be reused if necessary

S.NO	FUNCTION	DESCRIPTION	EXPECTED OUTPUT	ACTUAL OUTPUT	STATUS
1.	Framework construction	Generate the GUI for admin and user	Individual page for admin and user	Individual page for admin and user	Success
2.	Read the comments	Comments Analysis	Comments in text format	Comments in text format	Success
3.	Classification	Classify the Datasets	Finger Gestures	Finger Gestures	Success
4.	Rules implementation	Block the comments and friends	Block the users	Block the users	Success

**6.2 User Acceptance Testing**

A straightforward definition of acceptance testing is that it is successful when the programme performs as the customer may reasonably expect it to. One of the two potential conditions persists once the acceptance test is finished. Whether the database or other validations accept the inputs or not, this is OK. Accept only numeric values in the numeric field, for instance, and date-formatted data in the date field. Additionally, the not-null fields' null check. Show the error messages if any are encountered. Performance attributes that meet specifications are accepted. A deficiency

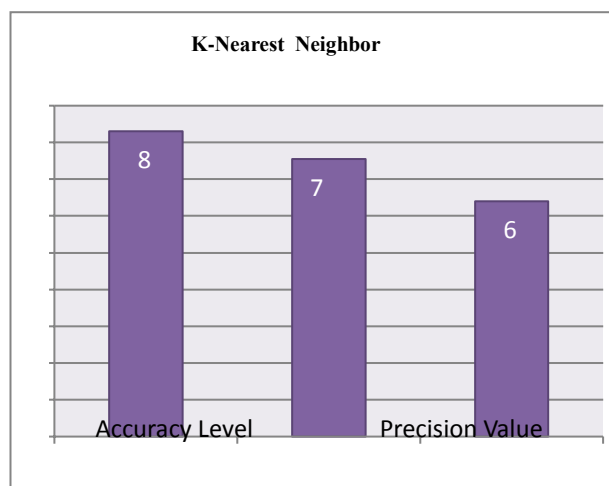
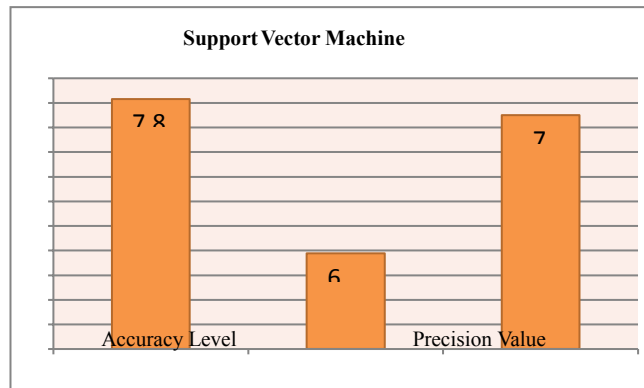
list is made after a specification variation is discovered. User acceptance testing, a crucial stage of any project, necessitates active user involvement. Additionally, it makes sure the system satisfies the functional specifications.



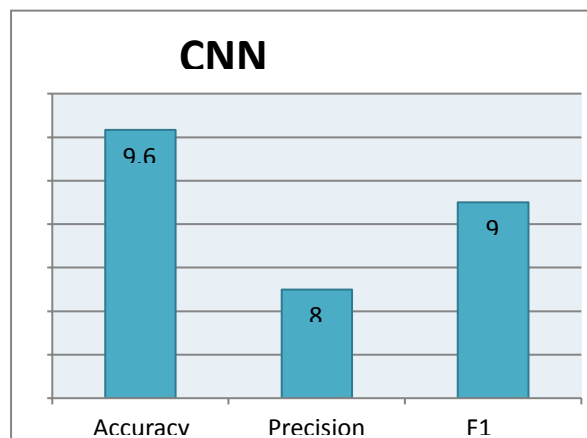
## 7. RESULTS

### 7.1. Performance Metrics

#### Existing Algorithm



#### Proposed Algorithm



## 8. CONCLUSION

One of the most precious gifts a person can possess is the capacity to look, listen, speak, and react to situations appropriately. Some unfortunate folks, nevertheless, are not given this chance. Sharing ideas, thoughts, and experiences with those around them helps people come to know one another. There are a few methods to do this, but the best one is to use your talent of "Speech." Everyone is quite good at communicating their ideas through speaking and understanding one another. By integrating a low-cost computer into the communication process, our programme aims to close the communication gap for the benefit of blind people by capturing, recognising, and translating sign language into voice. This paper uses an image processing technique to identify the handcrafted movements.

## 9. FUTURE SCOPE

Our system is still well-matched with the existing systems despite its average accuracy because it can execute recognition at the specified accuracy with wider vocabularies and without a tool like gloves or hand marks. The framework can be expanded in the future to incorporate different deep learning algorithms for sign recognition and real-time applications.

## 10. REFERENCES

- [1] Mohammed Mustafa, "A study on arabic sign language recognition for differently abled using advanced machine learning classifiers", 2020.
- [2] Necati Cihan Camgözü, "Sign language transformers: joint end-to-end sign language recognition and translation", 2021.
- [3] Ankita Wadhawan, "Sign language recognition systems: a decade systematic literature review", 2020
- [4] Nikolas Adaloglou "a comprehensive study on sign language recognition methods", 2020.
- [5] Dongxu li, "transferring cross-domain knowledge for video sign language recognition", 2020
- [6] Z. Lei, Z. H. Gan, M. Jiang, and K. Dong, "Artificial robot navigation based on gesture and speech recognition," Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Wuhan, 2014, pp. 323-327.
- [7] H. S. K., Rai S, S., Pal, S., Sulthana K, U., & Chakma, S. Development of Device for Gesture to Speech Conversion for the Mute Community. 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), IEEE explore(2018).
- [8] S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," IEEE Sensors Journal, vol. 15, no. 3, pp. 1321-1330, March 2015. doi: 10.1109/JSEN.2014.2370945.
- [9] R. R. Itkarkar and A. V. Nandi, "Hand gesture to speech conversion using MATLAB," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-4. doi: 10.1109/ICCCNT.2013.6726505.
- [10] Ahmed, Syed Faiz & Ali, Syed & Munawwar, Saqib. (2010). "Electronic Speaking Glove for Speechless Patients, A Tongue to a Dumb". Sustainable Utilization and Development in Engineering and Technology, IEEE Conference on. 56 - 60. 10.1109.
- [11] Safayet Ahmed, Rafiqul Islam, Md.Saniat Rahman Zishan, Mohammed Rabiul Hasan, Md.Nahian Islam, "Electronic speaking system for speech impaired people: Speak up", Electrical Engineering and Information Communication Technology (ICEEICT) 2015 International Conference on, pp. 1-4, 2015.
- [12] S.K. Imam Basha, S. Ramasubba Reddy, "Speaking system to mute people using hand gestures", International Research Journal of Engineering and Technology, Volume: 05 Issue: 09, Sep 2018.
- [13] Preetham, C., Ramakrishnan, G., Kumar, S., Tamse, A. and Krishnapura, N., 2013, April. Hand talk-implementation of a gesture recognizing glove. In 2013 Texas Instruments India Educators' Conference (pp. 328-331). IEEE.
- [14] Aiswarya, V., Raju, N.N., Joy, S.S.J., Nagarajan, T. and Vijayalakshmi, P., 2018, March. Hidden Markov Model-Based Sign Language to Speech Conversion System in TAMIL. In 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII) (pp. 206-212). IEEE.
- [15] P. Vamsi Praveen, K. Satya Prasad, Electronic Voice to Deaf & Dumb People, Using Flex Sensor International Journal of Innovative Research in Computer and Communication Engineering Vol. 4, Issue 8, August 2016, ISSN(Online): 2320-9801