International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 01 | Jan 2025 www.irjet.net p-ISSN: 2395-0072

HAND GESTURE BASED CURSOR DETECTION

Dr. D. Banumathy[1], Abinaya R[2], Aarthi S[3], Kaviya T[4]

[1]*Professor, Dr. D. Banumathy, Department of Computer Science & Engineering, Paavai Engineering College, Namakkal, Tamilnadu, India*
[2]*Student, Abinaya R, Department of Computer Science & Engineering, Paavai Engineering College, Namakkal, Tamilnadu, India.*
[3]*Student, Aarthi S, Department of Computer Science & Engineering, Paavai Engineering College, Namakkal, Tamilnadu, India.*
[4]*Student, Kaviya T, Department of Computer Science & Engineering, Paavai Engineering College, Namakkal, Tamilnadu, India.*

ABSTRACT: People with speech and hearing problems use sign language as a second language to communicate. Not able to Individuals use nonverbal communication methods, such as these sign language gestures, to express their thoughts and emotions to everyday people. It may be quite difficult to communicate with those who are hard of hearing. When communicating, those who are Deaf or mute must utilise hand gestures, which makes it challenging for others to understand what they are saying. Therefore, it is necessary to have systems that can identify various signs and provide information to the general audience. But because these everyday people have trouble comprehending what they're saying, having proficient sign language abilities is essential for training and educational sessions, as well as for legal and medical consultations. The need for these services has grown within the last several years. Other services, including video remote human interpretation that uses a fast Internet connection, have been created. Despite having serious drawbacks, these programs offer a rudimentary sign language interpreting service that is helpful and may be utilised. To address this problem, apply artificial intelligence technology to determine the user's hand using finger detection. Create the vision-based system for this proposed system in practical environments. A deep learning technique called the Convolutional Neural Network algorithm (CNN) is then used to classify the sign and provide a label pertaining to the identified sign. A Python framework was used to carry out the project's design.

KEYWORDS: Hand image acquisition, Binarization, Region of finger detection, Classification of finger gestures, Sign recognition.

1. INTRODUCTION

The process of translating a user's movements and signs into written text is known as sign language recognition. Those who are unable to interact with the larger community benefit from this approach. Raw photos or videos are converted into readable text by using neural networks and image processing techniques to link the motions with matching text in the training data. People who are mute frequently struggle to communicate with others since most people can only understand a small portion of their body language. These people could thus have trouble interacting with the wider public in an effective manner. Being unable to communicate verbally, those who are deaf or hard of hearing mostly rely on visual communication. For people who are hard of hearing or unable to talk, sign language serves as their primary means of communication. It has its own vocabulary and syntax, much like other languages, but it uses images to communicate. Because most people are generally unaware of these grammatical norms, it can be difficult for deaf or mute persons to express themselves using sign language grammar. As a result, they frequently only communicate with their family members or the deaf community. The growing social acceptability and support for international projects demonstrates the importance of sign language. People who are deaf or hard of hearing are eager for a computer-based solution in this technological age. One step towards achieving this goal is enabling a computer to comprehend human voice, facial emotions, and gestures. Without using words, gestures may convey information. A human may simultaneously produce an endless number of different gestures. Computer vision researchers are particularly interested in human movements since they are experienced through sight. The creation of a human-computer interface that can identify human motions is the aim of this project. To turn these motions into machine-readable code, a complex programming procedure is required. In order to improve output quality, this study mainly focuses on image processing and template matching. The symbols used to represent the letters in sign language format are shown in Figure 1

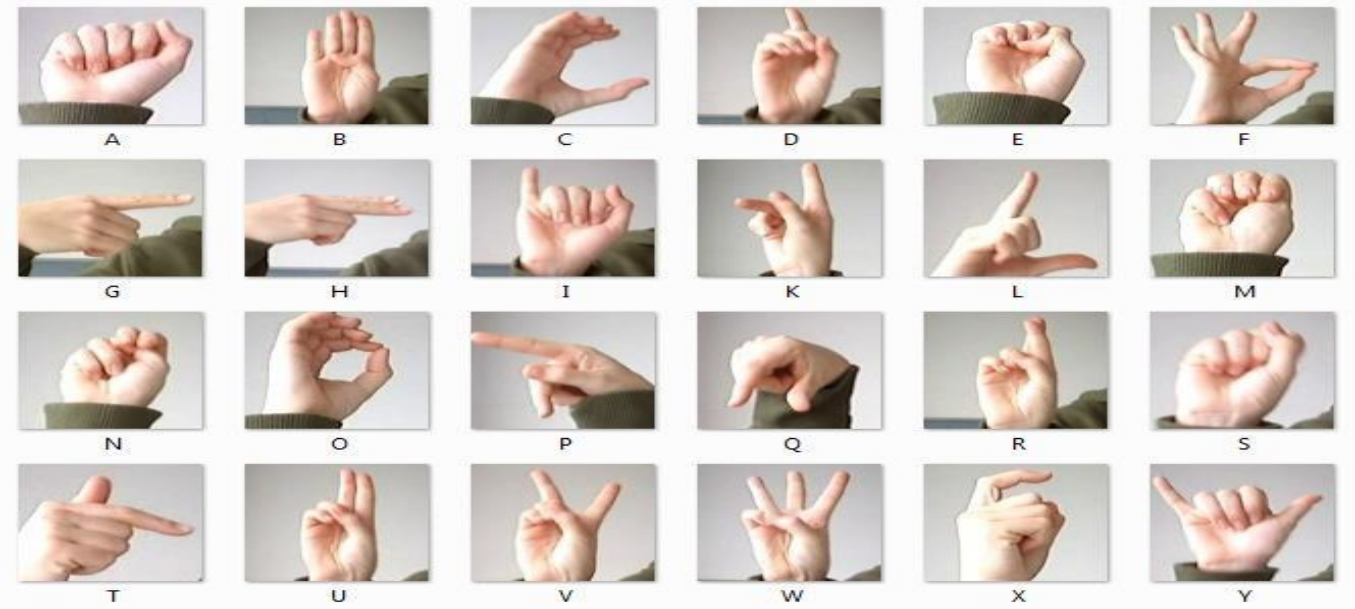

Figure 1: Sign Language for alphabets (refer: https://towardsdatascience.com/sign-language-recognition-with-highly developed-computer-vision-7b74f20f3442)

2. RELATED WORK

Tangfei Tao, Yizhe Zhao, et.al,…[1] The proposed review concentrated on the development, especially in the last few years, of algorithm-driven methods for sign language recognition. These models included both deep learning and conventional techniques, as well as sign language datasets, difficulties, and future directions in sign language recognition. In order to clarify the structure of these methods, the paper examined and explained the fundamental ideas of several approaches from the viewpoints of temporal modelling and feature extraction. The gloss, which denotes a complete sign language word made up of several gestures, is the main informational element in a sign language video presentation. A constant stream of movements is used to deliver the sign language message, and each distinct gesture provides specific meanings in addition to the links between the motions. The main frameworks for recognising sign language, such as Convolutional Neural Networks (CNNs) and Transformers, as well as popular techniques for continuous sign language understanding, were presented in the deep learning section. In addition to these methods, the paper looked at the difficulties in sign language recognition and included assessment metrics and datasets from recent years. It also emphasised the use of feature extraction and temporal modelling in traditional sign language identification methods. Conventional techniques demonstrated sensitivity to occlusion and illumination, and were limited in their ability to extract hand movement-specific characteristics. The manual feature design process was costly and timeconsuming, which resulted in accuracy restrictions. However, more interpretability was possible using older methodologies, which helped researchers appreciate the relevance of certain traits for their study.

Deep R. Kothadiya, et.al,…[2] To improve the precision of identifying isolated signals, the convolution-based mixed Inception architecture was created. The primary contribution was the improvement of InceptionV4 by employing uniform connections for optimum back propagation. The resilience and recognition accuracy of individual sign language recognition systems were further increased by introducing an ensemble learning architecture that made use of several Convolutional Neural Networks. To demonstrate the efficacy of the suggested learning strategies, a benchmark dataset of isolated sign language motions was used. With a recognition accuracy of 98.46 percent and higher robustness, the findings showed that the built ensemble model performed better in sign identification. The suggested deep learning approach helped lower communication obstacles for people with impairments by deciphering sign movements. Significant information was collected from the isolated sign frames by the stem module, the first part of the architecture. In order to reduce the dimensionality of the input frames, it included convolutional, pooling, and normalisation layers. Using ensemble learning techniques, the InceptionV4 model and the top convolutional neural networks were standardised and evaluated. Because the ensemble model had fewer trainable parameters and a lower computing cost, training it was simpler. As a result, the suggested model showed less training complexity and appeared to be well-suited for dynamic sign language recognition. An additional isolated sign language dataset was used to evaluate the created architecture. Larger and more complex structures were produced via ensemble models.

P. V. V. Kishore, et.al,.[3] The development of the multi-resolution convolutional neural network with spatial attention (MRCNNSA) had a major impact on the KL3DISL's ability to generate 3D sign language data. Online 3D datasets and standard deep learning models were used to test the suggested architecture for sign and action detection. Eight motion capture camera sensors and cutting-edge 3D motion capture technologies were used to create this sign language skeleton data. The model was specially customised for our dataset, and a great deal of study went into creating the skeleton shown in Figure 1. All of the Indian sign language symbols may be recognised by this skeleton. Ten distinct signers contributed to the collection, with an average of 550 frames per sign being captured during each capture event. The identification of any missing joints in each sign was made possible by a meticulous reconstruction of the collected data. The Biomechanics and Vision Computing Research Centre at KL University produced the 3D sign language dataset known as KL3DISL. It included 500 indications from different categories that are often used. With 500 3D labels for skeletal signs overall, KL3DISL collected approximately 25,000 samples from 10 subjects who performed each sign five times. The application of optional models to the datasets and the assessment measures were covered in detail in this section. Additionally, benchmark datasets and cutting-edge sign language (action) recognition models were used to compare with a number of common CNN designs as backbone networks. Furthermore, at various feature resolutions, an ablation analysis of the classifier's attention model was shown.

Bashaer Al Abdullah, et.al,…[4] To achieve the research goals, the created approach included formulating research questions, creating queries, choosing studies based on predetermined standards, and retrieving pertinent data. Moreover, enhancing recognition accuracy required the use of non-manual characteristics. To increase system accuracy and usefulness, it was recommended that future research concentrate on improving sophisticated deep learning models and adding non-manual features. For those who mostly utilised sign language, these ongoing developments offered the potential to revolutionise communication and remove barriers. With advancements in sensors found in wearable technology such as data gloves, watches, and bands, significant strides have been made in hardware-based sign language recognition. In order to efficiently record hand motions, orientation, and location, data gloves have been used extensively. On the other hand, electromyography (EMG) sensors tracked electrical muscle activity while signing to identify signals. Over the years, researchers have been trying to make these sensors smaller, which have resulted in the creation of wearable technology that is more comfortable. However, prolonged usage of these devices may result in pain. Furthermore, the costs associated with the research, maintenance, and manufacturing of these sensors continued to be high.

Vasileios Kouvakis, et.al,…[5] By investigating the reconstruction of discrete pictures in the context of American Sign Language (ASL) communication, the suggested approach addressed this problem. By using neural networks to parse feature vectors, the traditional method added needless complexity and inefficiency. Using a 24-QAM variant of the quadrature amplitude modulation (QAM) technology, a novel system model for image-based semantic communications was presented in order to get around these challenges. This modulation technique, which was created by removing eight peripheral symbols from the original 32-QAM collection, has been demonstrated to produce better error performance in ASL applications. Furthermore, a semantic encoder based on convolutional neural networks (CNNs) was shown, efficiently using the ASL alphabet. Key points and red-green-blue landmarks were superimposed on the taken pictures to provide a special dataset that improved the representation of hand locations. The suggested system's training, testing, and communication performance were evaluated using numerical data. This strategy encouraged interesting conversations and highlighted areas for development. To further improve the representation of hand locations, red-green-blue landmarks and important spots were added to the original dataset. Numerical findings that showed the system's performance improvements and the harmony between conventional and semantic communication techniques were used to evaluate the efficacy of the suggested system.

3. LITERATURE SURVEY

S.NO	TITLE	AUTHOR & PUBLISHER	TECHNIQUES	MERITS	DEMERITS
1	Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and	T Tao, Y Zhao, T Liu, J Zhu, 2024 &IEEE ACCESS	Artificial neural network	It increasingly good results on large datasets based on deep learning	It requires great computation al cost.

	Challenges				
2	Hybrid InceptionNet Based Enhanced Architecture for Isolated Sign Language Recognition	DR Kothadiya, CM Bhatt, H Kharwa, F Albu,2024 & IEEE ACCESS	InceptionNet framework	The proposed study can reduce the size of the model to minimize computational time.	The biggest drawback of these approaches is the dependency on gloves
3	Joint Motion Affinity Maps (JMAM) and Their Impact on Deep Learning Models for 3D Sign Language Recognition	PVV Kishore, DA Kumar, RC Tanguturi, 2024 & IEEE ACCESS	Joint motion model	The major advantage is the joint information coverage in 3D pose estimations using the 3D motion capture data, which makes it a reliable system for real time operation	This disadvantage gets further magnifies during the training process which gives more attention to retracting joint features than the contracting ones.
4	Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects	B Al Abdullah, G Amoudi, H Alghamdi ,2024 & IEEE ACCESS	Recurrent neural networks	This technique seldom attains good accuracies	The corpora for gestures in sign language are limited
5	Semantic Communication s for Image-Based Sign Language Transmission	V Kouvakis, SE Trevlakis, Boulogeorgos,2024 &IEEE ACCESS	Quadrature amplitude modulation (QAM)	This approach not only enhances regularization and reduces the processing	It created computation al errors.

TABLE 1: SURVEY PAPERS WITH MERITS AND DEMERITS

4. EXISTING METHODOLOGIES

Assessors examined the interface and assessed whether or not it met usability criteria using heuristic evaluation with usability professionals. The discovered usability problems led to a revision of the system interfaces. Following that, deaf and hard-of-hearing users participated in usability testing, which assessed the program's most important features using both objective and subjective approaches. Indian sign language is being employed as part of an ongoing endeavour to close the communication gap between the general population and the deaf and dumb. It might accelerate the development of autonomous systems that can understand and help the deaf and dumb, as well as making it simpler and faster for them to interact with the outside world, if this program is successfully expanded to incorporate words and common phrases. The absence of standardised databases causes Indian Sign Language research to lag behind that of the United States.

5. PROPOSED METHODOLOGIES

Sign Language uses hand movements, hand compass reading, and facial gestures in place of auditory sound patterns. There are erratic patterns in this language that differ from person to person and are not universal. However, since most people are not familiar with sign language, Deaf-mute people are finding it more difficult to communicate without some sort of translation. They feel as though they are being avoided. A commonly used method for interacting with the deaf-mute is

sign language recognition. There are two kinds of recognition models: sensor-based systems and computer vision-based systems. Computer vision-based sign recognition uses the camera as an input source, and motions entered are first image processed before being recognised. A number of methods, including the region of notice algorithm and the neural network approach, are then used to identify the processed movements. The main drawback of a vision-based system for sign language identification is that its picture collection process is sensitive to several environmental factors, including background conditions, camera orientation, and lightning sensitivity. However, it is less costly and more useful than using a camera and tracker to gather data. However, camera data is integrated into neural network techniques like the Convolutional Neural Network to increase accuracy.

CNN ALGORITHM: To develop a CNN (Convolutional Neural Network) method for sign language identification, you would typically do the following: Information Gathering: Create a substantial library of sign language films or images. Make sure the dataset includes a broad range of sign gestures and variants in order to improve the model's generalisation. Preprocessing data: Preprocess the collected data to enhance the model's learning process. Typical preparation steps include splitting the dataset into training and testing sets, resizing the images to a consistent size, and levelling the pixel values. Enhancement of Data: To increase the model's generalisation, use data augmentation techniques like flipping, scaling, and rotation to artificially increase the dataset's size. If you have a tiny initial data collection, this phase is quite beneficial. Model Architecture: Create your CNN model's architecture. It typically consists of convolutional, pooling, and fully linked layers. The convolutional layers extract relevant characteristics from the input images, which the fully connected layers utilise to do classification. Training: Train your CNN model using the given dataset. In order to provide correct predictions, the model learns to minimise a selected loss function, like categorical cross-entropy, while optimising its internal parameters during training. Evaluation: Use the testing dataset to assess the trained model's performance. Correctness, accuracy, recall, and F1 score are commonly used evaluation metrics for classification tasks. Fine-tuning: Consider altering the architecture, hyper parameters, or preprocessing techniques if the model's output is subpar. You may also consider techniques like transfer learning, which is applying a previously trained model from a large dataset (such as Image Net) to a sign language dataset. Deployment: If you are satisfied with the model's performance, use it for real-world applications. This can entail incorporating it into a web service or mobile application that can take in information (such pictures or videos) and make predictions. To guarantee the accuracy and dependability of your CNN algorithm for identifying sign language, don't forget to annotate the dataset with the appropriate labels, include enough variations in sign motions, and carry out thorough testing and validation.

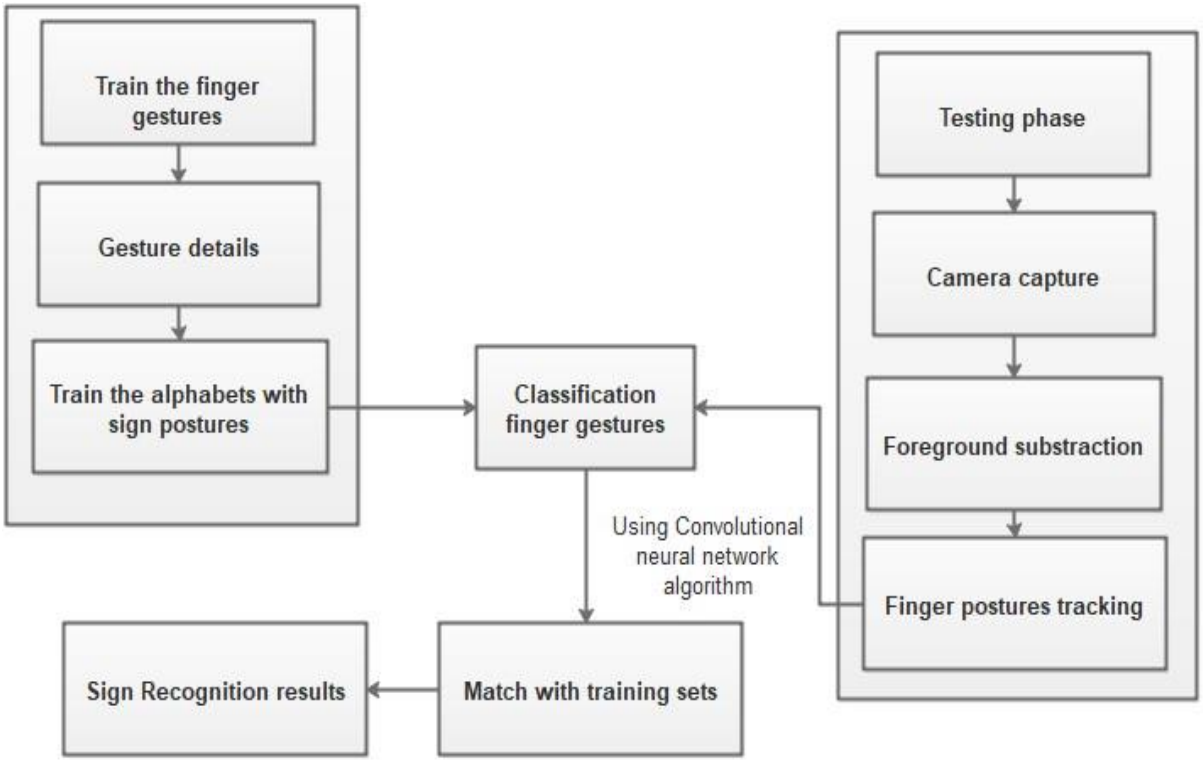

Fig 3: Proposed architecture

6. EXPERIMENTAL FOUNDATIONS

Sign language, a gesture-based communication method that varies by area and society, is used by deafmute people. Communication challenges arise because the general public is not widely familiar with sign language. Your objective is to create an SLR system that fills this gap by concentrating on vision-based recognition techniques that make use of neural networks and camera data.

Efficiency=∑(Totalnumberoffavourableconditiononthebasicfeatures)/(Totalnumberofconditions)

Increased efficiency leads to more precise and effective results. The presentation char till us trading this association can be seen in the following graph

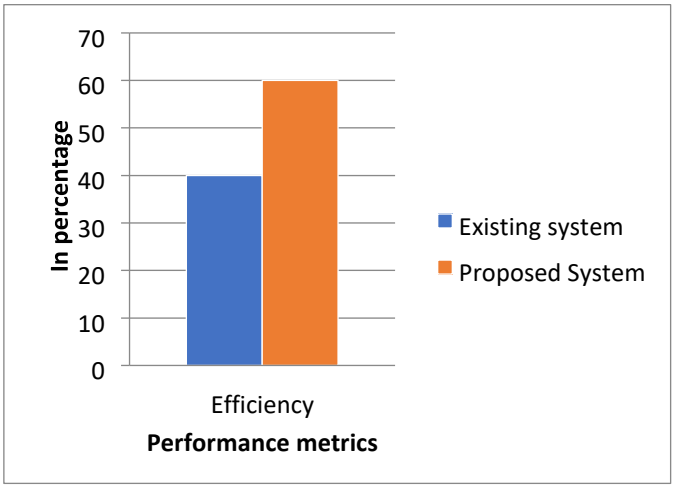

Fig4: Performancechart

7. CONCLUSION

Being able to see, hear, talk, and react correctly to situations was one of the most precious blessings a person could have. However, some unfortunate people are not afforded this chance. When people communicate their opinions, thoughts, and experiences with others around them, they come to know each other better. The best method for doing this is to use "Speech." Everyone has excellent persuasive speech communication abilities and is understanding of one another. Our idea seeks to help blind individuals communicate by introducing a lowcost computer into the mix. This will make it possible to record, recognise, and translate sign language into spoken language. This article uses an image processing approach to identify the handmade motions. This application is an illustration of a contemporary integrated system designed for people with hearing loss. The camera-based zone of notice could make data collection easier for the user. Each action will have meaning in and of itself.

REFERENCES

[1] Tao, Tangfei, et al. "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges." IEEE Access (2024).

[2] Kothadiya, Deep R., et al. "Hybrid InceptionNet based Enhanced Architecture for Isolated Sign Language Recognition." IEEE Access (2024).

[3] Kishore, P. V. V., et al. "Joint Motion Affinity Maps (JMAM) and their impact on deep learning models for 3D sign language recognition." IEEE Access (2024).

[4] Al Abdullah, Bashaer, Ghada Amoudi, and Hanan Alghamdi. "Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects." IEEE Access (2024).

[5] Kouvakis, Vasileios, Stylianos E. Trevlakis, and Alexandros-Apostolos A. Boulogeorgos. "Semantic communications for image-based sign language transmission." IEEE Open Journal of the Communications Society (2024).

[6] Gu, Yutong, Hiromasa Oku, and Masahiro Todoh. "American Sign Language Recognition and Translation Using Perception Neuron Wearable Inertial Motion Capture System." Sensors 24.2 (2024): 453.

[7] Teran-Quezada, Alvaro A., et al. "Sign-to-Text Translation from Panamanian Sign Language to Spanish in Continuous Capture Mode with Deep Neural Networks." Big Data and Cognitive Computing 8.3 (2024): 25.

[8] Renjith, S., Manazhy Rashmi, and Sumi Suresh. "Sign language recognition by using spatio-temporal features." Procedia Computer Science 233 (2024): 353-362.

[9] Arooj, Sadia, et al. "Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language." Journal of King Saud University-Computer and Information Sciences 36.2 (2024): 101934.

[10] Al Khuzayem, Lama, et al. "Efhamni: A Deep Learning-Based Saudi Sign Language Recognition Application." Sensors 24.10 (2024): 3112.

[11] Shah, Farman, et al. "Sign language recognition using multiple kernel learning: A case study of pakistan sign language." Ieee Access 9 (2021): 67548-67558.

[12] A. Wadhawan and P. Kumar, ‘‘Deep learning-based sign language recognition system for static signs,’’ Neural Comput. Appl., vol. 32, no. 12, pp. 7957–7968, Jun. 2020.

[13] Sharma, Shikhar, and Krishan Kumar. "ASL-3DCNN: American sign language recognition technique using 3D convolutional neural networks." Multimedia Tools and Applications 80.17 (2021): 26319-26331.