International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Hand Sign Language Detection Using Deep Learning

Dhanasankar M, Mr. M. Jayakumar, M.Sc., M.Phil., Mr. K. Aneesh, M.Sc.

M. Sc, Data Science and Business Analysis
Rathinam college of Arts and Science, Coimbatore, India
dhanasankar2002@gmail.com

ABSTRACT

For those who are hard of hearing or deaf, sign language is a vital means of communication that helps them interact with the external environment. The ability to automatically recognize sign language gestures can be a highly helpful tool in helping the cultures of the deaf and hearing communicate with one another. Convolutional neural networks (CNNs), a potent deep learning architecture well known for its efficacy in image classification tasks, are use in the work to propose a novel approach for the sign language detection using VGG16 Architecture. Starting with a data collection and preprocessing, the suggested approach is a thorough pipeline for the identification of sign language. To improve the model's resilience, a dataset of pictures captioned in sign language is selected, and image augmentation methods are used. A CNN with a VGG16 architecture that is specifically intended for sign language recognition is then fed the preprocessed data. In order to take the advantage of pre-trained models on large image datasets, the CNN model is trained by using a combination of labelled images through in use of transfer learning techniques. By capturing the spatial dependencies and temporal features inherent in sign sequences, the architecture has been adjusted to fit the unique properties of sign language movements. The suggested CNN model is a viable option for real-world application due to its efficiency and robustness, this will enable hearingimpaired individuals to engage and become more accessible in a range of contexts.

Keywords Gesture recognition, Hand pose estimation, Sign language interpretation, CNN, Deep Learning, Sign language detection, Visual recognition of sign language, Hand movement analysis, Hand gesture classification

1.Introduction

Hand sign language is an essential means of communication for those who are deaf or hard of hearing, allowing them to convey their ideas, feelings, and thoughts. Recent advances in deep learning techniques have revolutionized the fields of computer vision and pattern recognition, making it feasible to develop systems for real-time hand sign detection and interpretation. The general population and people who use sign language may be able to communicate more effectively as a result of this technology. Due to its potential to enhance accessibility, communication, and inclusion, deep learning based on hand sign language detection has attracted the lot of the attention. We investigate the many deep learning techniques and approaches used in the fascinating field of hand sign language detection. We'll talk about the importance of this technology, as well its uses, drawbacks, and providing opportunities to enhance the lives of the hard of hearing. We will also highlight current developments, best practices, and upcoming paths in this cutting-edge field of study. The area of hand sign language detection, which emphasizes the recognition and interpretation of manual gestures for communication, is revolutionizing computer vision and machine learning. To increase accessibility for the deaf and hard of hearing, it is imperative to be able to interpret sign language into written or spoken language. By utilizing deep learning models and computer vision techniques, hand sign language recognition systems are able to recognize and comprehend certain hand signs, movements, and locations associated with different messages. In practical applications, real-time processing to communicate effectively, you need to have certain talents. This technology has a number of uses, such as enhancing communication accessibility, supporting sign language instruction, and promoting computerhuman connection. Providing extensive annotated datasets, managing variations in signing styles, and guaranteeing cultural and linguistic appropriateness are a few of the difficulties. Further research into hand sign language recognition has the potential to bring about significant improvements in communication accessibility, education, and assistive technology all of which might result in more friendly encounters across diverse communities.

2.Objective

The main goal of hand sign language detection is to create trustworthy and accurate systems that can identify and decipher manual motions, making Facilitating communication for the hard of hearing. The purpose of the technology is to improve communication between the hearing and the deaf/hard-of-hearing communities. The objective is to provide context-aware and real-time hand sign language identification using cutting-edge computer vision and machine learning techniques, enabling users to interact naturally and intuitively. To provide inclusion in a variety of linguistic and cultural contexts, these systems should also be flexible enough to accommodate a wide range of dialects and variants of sign language. Aiming to improve communication opportunities for those with hearing impairments in many contexts, hand sign language recognition is being integrated into various platforms and devices, such as wearables, tablets, and smartphones. One major area of issue is accessibility. An inclusive society is intended to be created by reducing barriers to communication and encouraging a stronger sense of understanding and solidarity among people with different language ability through the development of hand sign language detection devices.

3.Literature Survey

3.1 Deep Learning-Based Recognition of Sign Language:

Mohamed Mahyoub, instructor; United Kingdom; Frisk Natalia; Sud Sudirman, Jamila Mustafina. [1] Sign language recognition is one kind of action recognition issue. Automated translation of one sign language into another is the aim of this kind of technology. Despite a great deal of work, there are still many topics in the SLR domain that require more research due to its vast reach. In this work, we examine whether speech recognition and classification in several sign languages can be achieved from video frames using deep learning techniques. The American, Indian, and Turkish sign languages are taken into consideration. Five distinct, varying degrees of complexity deep learning models are used in our methodology. Some of these are a shallow four-layer convolutional neural network, an inflated 3D model with the same, a basic VGG16 model, a VGG16 model with an attention mechanism, and a Transformer Encoder and Gated Recurrent Units-based Decoder. To test how effectively the models could identify and categorize words in movies, they were trained and assessed on three distinct sign language datasets. We found that there is a high association between model complexity and performance, with the inflated 3D model working best. Additionally, out of all the datasets, we discovered that the American Sign Language dataset presents the greatest challenge for word recognition across all methods.

3.2 Deep Learning-Based Sign Language Recognition System

Sudhanshu Tripathi, Nidhi Malik, Leena Singh, and Anupriya Chandrasekhar Since they are unable to hear or speak with the outside world, those who are deaf or dumb are often singled out for attention in an effort to set them apart from the general populace. distinct gestures in [2] indicate distinct letters, numbers, or sentences. Most communication is done in either Indian Sign Language (ISL) or American Sign Language (ASL). Not everyone who is deaf uses sign language; some make use of assistive technology, such hearing aids, however not everyone who is deaf can afford them. Sign language addresses each of these issues. To ensure that create software that is both skilled and well-known, this article seeks to improve existing systems and recognize sign movements using deep learning.

3.3 Recognition of Hand Signs Using Machine Learning

Greeshma Pala, Jagruti Bhagwan Jethwa, Satish Shivaji Kumbhar, and Shruti Dilip Patil the most common problem faced by the deaf and voiceless people while utilizing sign language to interact with others is that not everyone in the proximity may be able to comprehend sign language. The main objective of this method is to open up communication channels amongst communities so that people in the silent community may engage and speak with one another. The linearity problem is brought up in [3] because different people make different hand motions in different shapes and orientations. Recent systems have created an range of tactics and algorithms to build this system and solve the problem. The hand gesture movements were previously decoded using algorithms like K Nearest Neighbors (KNN), Multiclass Super Vector Machine (SVM), and glove experiments. Algorithms such as Multiclass Super Vector Machine (SVM), K Nearest Neighbors (KNN), and glove experiments were previously used to decipher hand gesture motions. Identifying the algorithm that would provide the best degree of accuracy, this study examines the KNN, SVM, and CNN algorithms. After preprocessing 29,000 images, the accuracy of KNN, SVM, and CNN models were 93.83%, 88.89%, and 98.49%, respectively. Two sets of pictures were created: test and the train data.

3.4 Recognizing American Sign Language using Computer Vision and Deep Learning

Bantupalli Kshitij; Xie Ying A disability that affects a person's capacity for hearing and speaking is called speech impairment. In [4] Affected individuals communicate through alternative means, such as sign language. Even though sign language is used widely these days, non-speakers still find it challenging to connect with signers or language users. Recent advancements in these sectors have led to promising advances in the areas of motion and gesture identification utilizing deep learning and computer vision-based approaches Developing a vision-based program that converts sign language to text is the goal of this project., facilitating communication between signers and non-signers. The proposed approach extracts temporal and spatial characteristics from video sequences. Next, we use a CNN named Inception to identify spatial characteristics. Next, we use temporal characteristics to train an RNN. This dataset is called the American Sign Language Dataset.

3.5 Deep Learning-Based Recognition of Human Sign Language

D. Femi, Y. V. Rakesh Reddy, S. Thylashri, D. Santhakumar, and V. Jawahar the deaf and dumb impairment is among the biggest problems confronting humanity.[5] There exist several methods that technologies for the identification of the deaf and dumb handicapped are produced. To enable users of sign language to communicate with the general population, it is essential to learn how to understand sign language. Machine learning and picture classification algorithms enable computers to read sign language, which humans can subsequently decipher. To ensure that detect motions in sign language, this article employed convolutional neural networks to recognize the symbolic hand gestures used by the deaf and dumb challenged. Sign language is primarily unknown to the general population, although it is learned by those who are deaf or dumb. Normal individuals find it challenging to converse with deaf and dumb impaired persons because of this. The main objective of our work is to predict the hand motions of the dumb and deaf. In this study, we apply the CNN Algorithm to predict with a certain accuracy level by recognizing different hand gestures as input. The model is trained on a pre-existing dataset.

4.Methodology

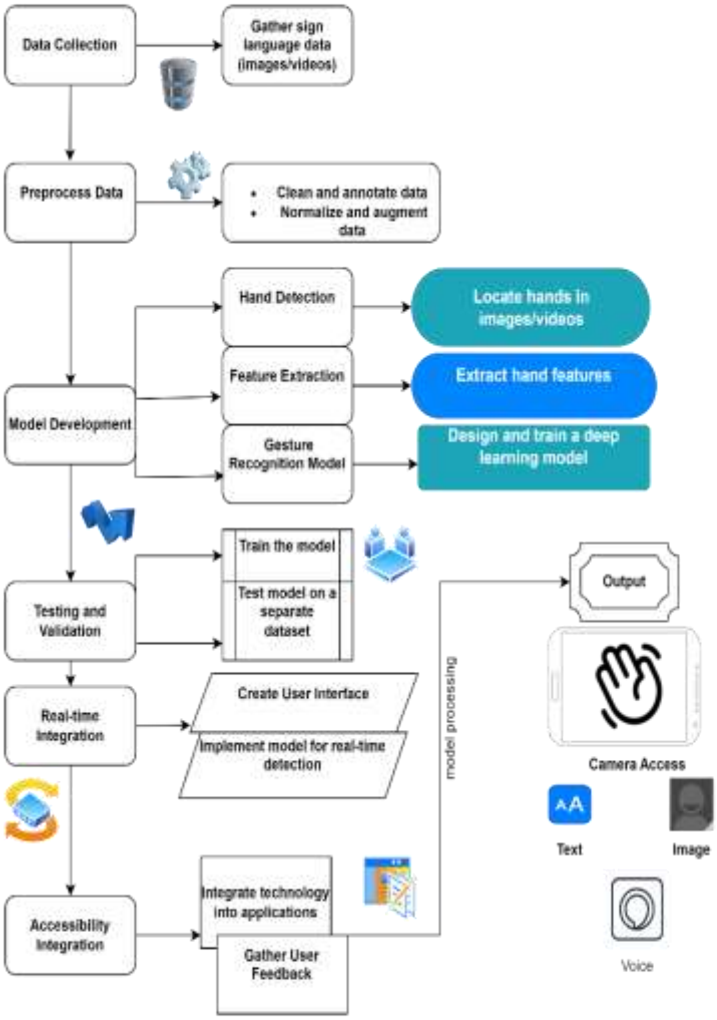
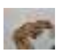
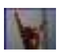
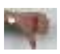
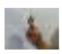
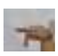
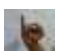
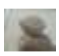
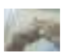
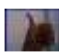


Fig 1: The model's workflow

4.1 DATA COLLECTION

Gather data from several sources to get a thorough representation of sign language gestures. Various people, backgrounds, lighting settings, and camera angles can all contribute to this diversity. To guarantee inclusivity and authenticity, work with communities or sign language specialists. Add several different sign language gestures to our dataset. Hand movements and shapes can vary amongst gestures. We should able to recognize the vocabulary in the dataset. Put the appropriate sign language label on each image by hand. This is noting which sign or gesture appears in each picture. For the model to be successfully trained, make sure that labeling is accurate and consistent. Get a big enough dataset so that it can used to train a reliable model. The intricacy of sign language and the variety of gestures determine the dataset's size.

 Brush	 Call	 Deer	 Dislike	 friend
 Handicap	 I	 Ill	 Include	 Like

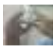
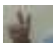
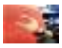
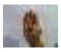
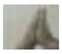
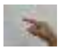
 More	 Nice	 Night	 Ok	 parrot
 pay	 peace	 scissors	 small	 stone
 Stop	 Stress	 We	 Welcome	 you

Fig 2: ISL Sample dataset

4.2 DATA PREPROCESSING

Hand Region Detection

Utilize image processing techniques to detect and localize the hand region within an image. Common methods include skin color detection, background subtraction, or contour analysis.

Gesture Segmentation

Develop algorithms to segment the hand from the background and other objects in the image.

4.4 MODEL TRAINING

Training the VGG16 Model

The dataset is split, and the training set is used to train the model. The model learns to map input data to the appropriate output labels during training. During this procedure, the internal parameters of the model are adjusted to minimize the difference between the model's predictions and the actual labels of the training set. On the training set of data, the model's performance is continuously tracked. A crucial phase in creating a machine learning model, like a VGG16-based hand sign language identification system, is monitoring training metrics like accuracy and loss on the validation set. This process ensures that the model generalizes to new data correctly and helps identify potential issues such as under- or overfitting at an early stage. The model minimizes the loss function during training. a method of measuring the distinction between expected and actual values by altering the weights and biases of the training dataset and learning from it. However, the model needs to work correctly on both new, untested data and the train set of data. In this instance, the validation set is used. The validation set comprises an additional subset of the information not shown to the model during training. Regularly, generally after each training cycle, the model's performance is assessed on the validation set. The VGG16 Architecture

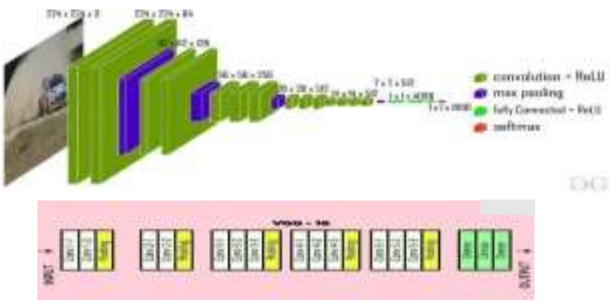


Fig 3: Architecture of VGG16

Figure:3 provides an outline of Among artificial neural networks, the CNN is a framework that performs particularly well in image identification applications. In 2014, the well-liked CNN architecture known as VGG16 was unveiled. It consists of three completely linked layers after sixteen convolutional layers. The completely interconnected layers categorize the characteristics derived from the input picture by the convolutional layers. At the network's end, the layer of SoftMax outputs the likelihoods that the input picture falls into each of the potential classes. The specific image you sent me shows the input image, which is 224 pixels wide by 224 pixels high and has three color channels (red, green, and blue), on the far left of the diagram. The image passes through successive convolutional layers, represented by the blue boxes in the diagram. on extract features from the input image, each convolutional layer applies a sequence of filters on the image. Beneath it is the matching box in the diagram for each convolutional layer's output size. One convolutional layer, for instance, contains 64 filters and produces a 64-channel picture with 224 pixels for width and 224 pixels for height. The image passes through several pooling layers after the convolutional layers; these are represented in the graphic as green boxes. Through pooling layers, the image's size is reduced through subsampling. The first pooling layer in the VGG16 architecture, for example, minimizes the dimensions of the image by a factor of two in each dimension. By doing this, the processing load on the network is reduced, which may improve network efficiency. The series of fully linked layers that the image passes through after the pooling layers is embodied by the yellow boxes in the diagram. Like in classical neural networks, all of the neurons in one layer are connected to every other layer's neuron through fully connected layers. The fully connected layers in the VGG16 architecture learn to classify the features that the convolutional layers extract. The final layer in the network is called SoftMax, and the manner in which it is shown is by the orange box in the diagram. The SoftMax layer outputs the probability that the input image belongs to each of the potential classes. The final layer in the network is called SoftMax, and This is exemplified by the orange box in the diagram. The SoftMax layer outputs the probability that the input image belongs to each of the possible classes. If the input image shows a friend sign, the SoftMax layer might generate a probability of 0.9 for the "friend" class and a probability of 0.1 for all of the other classes.

Validation Performance

Evaluating the model's performance in the validation set is essential to guaranteeing that it works well when applied to new data. The set of validation helps optimize model performance by adjusting hyperparameters such batch size and learning rate. Adjustments to prevent overfitting or underfitting can be made by monitoring the model's performance on this independent dataset.

Optimization with hyperparameters

Many aspects of a model's performance are controlled by its hyperparameters. These consist of regularization parameters, batch size, and learning rate. The validation set serves as a benchmark for assessing various hyperparameter combinations.

Final Model Evaluation:

Training sets and validation data assist in training and optimizing the model. The model's ultimate efficacy is then ascertained using the test set. In complete contrast to the training and validation sets, the test set offers an accurate analysis of the model's data generalization.

4.5 GUI CREATION

Creating a user-friendly interface for graphical users (GUI) is essential for developing systems that need collaboration from users. This is particularly valid for programs that use gesture recognition to collect sign language during engagement. As an intermediary between the user and the underlying technology, the GUI enhances accessibility and usability. This is a detailed breakdown in the process.

User Interface Design: Begin by designing the graphical elements of the interface, considering user experience and accessibility. This involves creating a visually appealing layout that intuitively guides users through the interaction process.

Capture Device Integration: This system involves capturing sign language gestures through a camera, integrate the camera functionality seamlessly into the GUI.

4.6 DESCRIPTION FOR OUTPUT

Image-to-text transcription: Image words are translated into text using real-time image recognition in this manner. Despite not having anything to do with sign language it is applicable to conjunction with a sign language interpreter to foster an atmosphere that is more welcoming to both hearing and deaf people.

Speech recognition using sign language: Systems that can recognize spoken words or brief phrases based on signs that match is being developed. Essential components would allow deaf people to "speak" and interact with non-signers more effectively by translating their signs into spoken words.

5.Implementation

Have a real-time video capture module implemented to be able to continuously capture frames from a webcam or other video source. utilizing the VGG16 model, that has already undergone hand sign language categorization training. Preprocessing involves resizing each frame to the input dimensions predicted by the model and normalizing the pixel values. To ascertain the expected class label for the hand sign, feed the pre-processed frame into the VGG16 model. Match the text (such as "friend") to the expected class label. Real-time text recognition should be displayed on the display. Provide a user interaction method that will enable the system to reset or adjust to new signs. This may entail setting up gestures to initiate particular operations, such rebooting the system or teaching it new signals. Update the video stream continuously, predicting each frame and presenting the recognized text as appropriate. For an efficient and simple user experience, include error management and make sure the UI is user-friendly. This solution offers a useful application for hand sign language identification by enabling users to engage with the system by making hand gestures. The matching recognized text is shown on the display in real-time.

6.Result and Discussion

In this hand sign language detection web application has shown encouraging results in properly identifying and converting hand signs into related text It was instructed utilizing a collection of visual representations of English words. On a convolutional neural network, the model was built architecture, which effectively caught the intricate details using hand motions to allow for trustworthy classification. Through the online application, users can engage in actual time with hand signs that correspond to English words in a smooth and user-friendly interface. The accurate translation of these indications into text on the display shows how well the model understands and interprets a broad range of movements. By allowing as an example of identify a wide variety of signals that are often used in sign language, the dataset which consists of an assortment of hand signs has increased the model's adaptability. Increased precision as well as a wider range of hand sign language communication applications might result from dataset expansion and ongoing model improvement. All things considered, the findings demonstrate how methods for machine learning may become accustomed to create inclusive and accessible software that facilitates communication in sign language users.

7.Conclusion

Conclusively, deep learning more especially, the application of CNNs for hand sign language detection is transforming the way people with hearing loss communicate with one another. This may be accomplished by developing and deploying a CNN model using a multimodal strategy that includes dataset preparation, model architecture design, training, assessment, and deployment. The model shows an impressive ability to independently learn complex spatial information present in hand motions using CNNs. To achieve optimal model performance, the environment setup which includes preprocessing approaches, software frameworks, and hardware infrastructure is crucial. Large annotated datasets are employed in the model's training phase, which is frequently assisted by transfer learning, to help the machine interpret hand signs correctly and broadly. The CNN framework is integrated into interactive systems or user interfaces to guarantee usability and accessibility during deployment. The deployment phase emphasizes responsible and ethical practices and also calls for careful consideration of security and privacy issues. The model must be continuously improved through updates, user feedback, and constant observation as technology advances. Beyond machine learning, the advancement of deep learning in order to hand sign language identification fosters inclusion by enabling successful communication and understanding between the deaf population and non-deaf individuals. Ultimately, this study illustrates how deep learning could enhance accessibility and human empowerment, paving the way for a time when innovative technological solutions will progressively eliminate barriers to communication.

8.Future Scope:

Future applications for an online using hand signs to communication detection application that was trained on a picture-based English word dataset appear promising. Future improvements to this method could identify a greater range among the sign languages, promoting inclusivity furthermore available to those who are hard of hearing and deaf. Instantaneous sign language comprehension and a reduction in communication gaps in a range of contexts could be achieved through integration with wearables and smart devices. In addition, continuous advancements in computer vision as well as machine learning algorithms might improve hand sign recognition's efficacy and accuracy, making this technology useful in real-world situations including communication platforms, hospitals, and educational institutions. Furthermore, collaborative efforts to expand vocabulary coverage and continuous dataset updates may further enhance the system's performance and applicability. By doing this, it would be guaranteed that the system will always be useful in promoting inclusive communication for individuals worldwide.

9.Reference

1. "Sign Language Recognition using Deep Learning," 15th International Conference on Developments in Systems Engineering (DeSE), Baghdad & Anbar, Iraq, 2023, pp. 184-189, doi: 10.1109/DeSE58274.2023.10100055, M. Mahyoub, F. Natalia, S. Sudirman, and J. Mustafina.
2. "Sign Language Recognition System Using Deep Learning," A. Chandrasekhar, L. Singh, S. Tripathi, and N. Malik, 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2023, pp. 67-71, doi: 10.1109/Confluence56041.2023.10048804-8.
3. "Machine Learning-based Hand Sign Recognition," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 356-363, doi: 10.1109/ICAIS50930.2021.9396030, G. Pala, J. B. Jethwani, S. S. Kumbhar, and S. D. Patil.
4. "American Sign Language Recognition using Deep Learning and Computer Vision," by K. Bantupalli and Y. Xie, IEEE International Conference on Big Data (Big Data), 2018 IEEE, Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.
5. "Human Sign Language Recognition Using Deep Learning," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1–5, doi: 10.1109/ICSES55317.2022.9914057, S. Thylashri, D. Femi, V. Jawahar, D. Santhakumar, and Y. V. R. Reddy.
6. Bhadra, R., and Kar, S. (2021, January). Using a deep multi-layered convolution neural network, sign language can be identified from pictures of hand gestures. IEEE Second International Conference on Control, Measurement, and Instrumentation (CMI), 2021 (pp. 196–200). IEEE.
7. Jameel, A. S. M. M., Akter, J., Urmee, P. P., Al Mashud, M. A., & Islam, S. (2019, November). Xception model with augmented dataset for real-time detection of Bangla sign language. In WIECON-ECE 2019, the IEEE International WIE Conference on Electrical and Computer Engineering (pp. 1–5). IEEE.
8. Othadiya, D., Sapariya, K., Bhatt, C., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Deep learning-based sign language identification and recognition. 11, No. 11, Electronics, 1780.

9. In December 2019, Bhagat, N. K., Vishnusai, Y., and Rathna, G. N. Deep learning and image processing are used to recognize Indian sign language gestures. 2019 (pp. 1-8), Digital Image Computing: Techniques and Applications (DICTA). IEEE.
10. Kumar, K., and S. Sharma (2021). ASL-3DCNN: 3-D convolutional neural networks are used in the recognition of American sign language. Tools and Applications for Multimedia, 80(17), 26319-26331.
11. In October of 2018, Ko, S. K., Son, J. G., and Jung, H. Recurrent neural networks are used to recognize sign language using human keypoint detection. In Adaptive and Convergent Systems Research Conference Proceedings, 2018, pp. 326-328.
12. Hu (2021), Zhao (2021), Zhou (2021), Wang (2021), & Li (2021). Pre-training of handmodel-aware representation for recognition of sign language is called SignBERT. 11087–11096) in the Proceedings of the IEEE/CVF International Conference on Computer Vision.
13. Tan, A. W. C., Lee, C. P., Tan, S. C., and Lim, K. M. (2019). Hand energy image and convolutional neural network hand modeling were used to isolate sign language recognition. 78, 19917-19944; Multimedia Tools and Applications.
14. Handra (2021), Bhargav (2021), Sanjanaa (2021), and Gopi (2021). CNN Hand Sign Identification. Performability Engineering International, 17(3), 314.
15. In 2020, Wei, C., Zhao, J., Zhou, W., and Li, H. Continuous sign language recognition using semantic border detection and reinforcement learning. IEEE Transactions on Video Technology: Circuits and Systems, 31(3), 1138–1149.
16. Chung, S. Y., Lee, C. K., Ng, K. K., Chen, C. H., Lau, H. C., & Tsoi, T. (2021). Recurrent neural networks are used in the recognition and training of American sign language. 167, 114403; Expert Systems with Applications.