International Journal of Innovative Research in Engineering
Volume 7, Issue 2 (March - April 2026), PP: 145-151.
https://www.doi.org/10.59256/ijire.20260702019
www.theijire.com
ISSN No: 2582-8746

Sign Language recognition using Machine learning

Thotakura Padmanabha Asrith[1], Ravuri Likhil Mohan Abhiram Naidu[2], Dr.V.Ulagamuthalvi[3]
[1,2,3] *Department of CSE, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.*

How to cite this paper:
Thotakura Padmanabha Asrith[1], Ravuri Likhil Mohan Abhiram Naidu[2], Dr.V.Ulagamuthalvi[3], "Sign Language recognition using Machine learning", IJIRE-V7I2-145-151.

Abstract: *Sign language recognition is very important for establishing good communication between hearing-impaired people and machines. This project is intended to classify the use of hand gestures for communication with robots, using the machine learning techniques. Using Media Pipe for precise hand detection and key point detection, a powerful dataset of sign language gestures is ready for model training. To increase the performance of classification, specific learning methods, AdaBoost and Gradient Boosting are implemented. A combination of multiple weak learners is employed through these techniques to create a strong predictive model, that can separate complex patterns of gestures with better accuracy and generalization. OpenCV makes it easy to grab video in real time and process the data, enabling the system to interpret gestures in real time and feed back to the user. This integrated approach adds to the robot's capacity to comprehend and react to human gestures, driving more intuitive and inclusive human-robot interactions. The use of ensemble-based learning in gesture recognition is a scalable, reliable solution for the translation of sign language into machine-understandable commands, driving advances in accessibility and automation in robotic communication.*

Keywords: *Sign Language Recognition, Machine learning, Ensemble Learning, Ada Boost, Media Pipe, Real time gesture recognition.*

I.INTRODUCTION

Sign language recognition is a rising field becoming a gateway between the hearing-impaired community and the rest of society for communicating. Communication is basic to human interaction, but communicating in areas where the individual has a hearing or speech impairment can be a big challenge in mainstream situations. Sign language is a powerful medium for these communities but also is publicly underrepresented in tech solutions and understood by a limited percentage of the population. This limitation does not aid in facilitating free interaction of deaf or hard of hearing people in a variety of situations, such as educational institutions, workplaces and public service, in which communication with non-signers is of importance. Therefore, there is a big need for the development of systems that can recognize and translate sign language into a system that can be understood by machines and non signers alike. With the advancement of artificial intelligence (AI), computer vision and machine learning, there is now an opportunity to cover up this gap by the development of sign language recognition systems. The problem of sign language gesture recognition has its own complex problems because it consists of processing and interpreting the dynamic movement and position of the hands, usually under various background circumstances with different lighting conditions. Moreover, sign language is not only about hands; there are also facial expressions, posture and other non manuals features, making recognition systems even more complex. The aim of research into sign language recognition is to create solid models that can understand these gestures in real- time, converting them into text or speech. These systems can make accessibility and inclusion far more accessible to those in the deaf community, and communication can be much easier in various public and private spaces.

The motivation behind this project arises from the need now for accessible solutions for communication of the hearing-impaired that is on the rise. Despite existing translation means for sign language, many are expensive and inconvenient to use, e.g. they require extensive hardware (motion sensors, data gloves) or are only able to recognize a small number of predefined gestures. This limits their use in the real world. Representations of Sign Language The rising availability of high-quality video cameras and the development of advanced machine learning models provide a more practical means to solve this problem by applying computer vision and deep learning models to recognising and interpreting sign language gesturing from standard video feeds. The reason for motivation is to build an intuitive, real-time and scalable solution that can work with less hardware requirements that can be available to the general public.

The problem this project is focused on is the necessity of creating a better and more scalable sign language recognition system idlplemented into real-time applications, such as communication with robots, smart devices, or even in educational tools. Current systems, for instance, are often plagued with problems such as accuracy in different lighting conditions and real-time performance as well as the scalability to different sign languages. Additionally, many of the existing systems are only specific to a particular sign language, such as American Sign Language (ASL), and are not easily adapted for other languages or dialects. Furthermore, most systems do not offer continuous recognition as is required for real-time translation in a conversation.

The specific objectives of this project include designing and deploying a sign language recognition system with machine learning techniques capable of accurately interpreting the hand gestures in real-time. The system will use MediaPipe to perform hand tracking and keypoint extraction to capture and process gesture data. This information will then be fed through an ensemble learning model, a combination of different methods such as AdaBoost and Gradient Boosting, to give you a better classification performance. The aim is to build a model that is able to cope with the variation in hand orientations, hand motion, and background and to provide high accuracy with common sign language gestures.

In addition to recognition, the system will seek to provide some actionable feedback, the conversion of recognized gestures into either text or commands that can be understood by a machine. This will enable more intuitive human-robot interaction, in which robots can interpret and react to gestures in real-time. The system will be designed with performances in mind, for example with a recognition latency of less than 0.5 seconds and an accuracy of 98% for common gestures. These objectives are very important for ensuring that the system is not only functional, but is dependable enough for practical use in real-world scenarios.

Moreover, the system will be made scalable so more gestures and sign languages can be added later. The ability to cope with new signs is important, as this allows the system to change and serve different linguistic groups. The project will also focus on developing an easy to use interface which can be deployed on a variety of devices ranging from smartphones to embedded systems, making the solution accessible to a broader audience. The fact you can work in real-time, with little hardware needs, means that the system can easily be made work in the course of everyday scenarios, without the necessity of specialized equipment.

Through this work, the project will contribute to the field of assistive technologies which is gaining momentum as it pursues the challenge of enhancing communication between humans and machines using machine learning technologies. It will also drive social inclusion by introducing a tool to help interactions between those who are hearing impaired and those who don't understand sign. The system's modular design will guarantee that it can be expanded to support a variety of sign languages and communication modes, like speech synthesis, to make it more versatile and useful in different contexts.

In conclusion, this project hopes to contribute to the sign language recognition field with a system that is accurate, scalable, and deployable in real-time applications. By using the power of machine learning and computer vision, the system will help break down communication barriers and can offer more inclusive and accessible environments for the deaf and hard of hearing community. This work will form the basis for a range of future innovations in assistive technologies and human-computer interaction.

II.LITERATUTRE REVIEW

The development of sign language recognition systems has experienced significant progress within the last few decades, moving from basic methods of image processing to machine learning-based methods that are quite advanced. Early work was mostly based on vision approaches employing traditional image processing algorithms, such as background subtraction, contour and skin color segmentation. These methods could isolate the hand and extract movement features from the frames of video. Handcrafted descriptors like Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Local Binary Patterns (LBP) were frequently used to describe spatial features of hand shapes and gestures. While these approaches performed well with controlled settings, they lost in performance in complex scenarios where issues such as lighting variations, complex backgrounds and occlusions appeared .

In parallel to that, sensor based systems were developed to complement vision based approaches. These systems used wearable devices such as data gloves, accelerometers, and flex sensors to capture data about motion and orientation of the signer's hands. Such systems were more robust in dealing with background interference, and were able to track finger movements better. However, they had limitations compared to the first computers regarding portability, cost, and user comfort, etc. which made them less suitable for casusal or public use . On top of this, there were also systems that adopted hybrid approaches and used both vision and sensor data, although these systems were often too complex and required specialized hardware that did not allow for scalability.

With the arrival of deep learning, and specifically Convolutional Neural Networks (CNNs), a change of paradigm took place in the sign language recognition method. CNNs are Auto-learning the discriminative features of the images from the raw data, hence reducing the reliance on man-made feature engineering. This resulted in considerable improvements in accuracy, especially AlexNet and VGGNet CNN models for static hand gesture recognition tasks . These early generative deep learning models proved to be quite successful at recognizing hand gestures, particularly in controlled settings with little noise. However, although these enhancements increased depth, the deep-learning models still required large amounts of labelled data, which were often difficult to obtain

For a more dynamic and ongoing sign language recognition, researchers began to incorporate temporal modelling methods such as Recurrent Neural Networks (RNNs) and Long Short- Temporal Memory (LSTMs) networks. These architectures enabled the recognition system to capture temporal dependencies between the frames that can be used, and can therefore be well- suited for gestures that rely on motion patterns rather than on fixed hand shapes. The mixture of CNNs to extract spatial features and LSTMs to analyze time sequences became the common choice for many end-to-end gesture recognition pipelines. This approach was able to better deal with the complex variations in hand movements of dynamic sign language gestures, with an increased accuracy in real-world conditions .

Recently, new and more advanced architectures have appeared, such as Vision Transformers (ViTs) or Spatio-

Temporal Graph Convolutional Networks (ST-GCNs), which capture both spatial and temporal information at the same time. These models have demonstrated great promises on sign language recognition due to their capacity to deal with long-range dependence and represent skeletal keypoints extracted from pose estimation models, such as OpenPose and MediaPipe. [source="cite reprove"] [7]

dels have shown better accuracy and robustness, especially with respect to dealing with signer appearance variation, environmental variation, and signing speed variation. However, regardless of their potential, these advanced models still pose challenges when it comes to real-time deployment on resource constrained devices due to the high computational demands.

Moreover, parallel to the upgrading of the model architectures has been a challenge in constructing the system to be universally adaptable and capable of recognising a wide variety of sign languages. Sign languages differ in various parts of the world, including American Sign Language (ASL), British Sign language (BSL) and Indian Sign Language (ISL) with its own vocabulary and grammar rules. Most systems are trained to understand one aspect of sign language, and performance tends to fall off for other languages

. Furthermore, many existing systems are constrained by datasets that either are insufficient in size or diversity in terms of signer demographics, camera angles, and the signing conditions. This limitation limits the generalization ability of trained models, prevent them from being effective in real-world settings

In addition to recognition accuracy, the robustness of systems in real-world conditions has been a primary research goal. Issues such as lighting variation, background interference and hand orientation variation frequently impact performance of sign language recognition systems. To rectify these, data augmentation methods such as rotation, flipping and scaling have been used to make more robust training data sets to handle variations in real world scenarios . Despite these efforts, there has been difficulty in obtaining high accuracy in recognition in uncontrolled environments, and research continues to be done to improve the adaptability of these systems to the variability of the real world.

While real-time sign language recognition systems have made major progress, challenges such as system latency, computational efficiency, and deployment on mobile or embedded devices remain major obstacles. To address the high computation complexity and low performance of low-power devices, several methods have been proposed to reduce the computation complexity, such as model pruning, quantization, and knowledge distillation. . However, finding the balance between model complexity and real-time performance is an open problem. Achieving a low-latency, high-accuracy system that can operate efficiently on mobile devices is important for the adoption of sign language recognition technology's widespread adoption.

Furthermore, how non-manual features, such as facial expressions and body postures, are integrated is an area in research on sign language recognition that is still underexplored. Many systems still concentrate primarily on hand gestures, not taking into account the important role played by non-manual cues with important grammatical and semantic contents in sign languages . The integration of these features might raise the accuracy and the naturalness of the translation of the sign language, but capturing these features requires additional sensors or more advanced vision techniques, and thus raises the system complexity.

Finally, whereas there is now a vast body of work on sign language recognition systems that are focused on the recognition of isolated signs, continuous sign language recognition (CSLR) still remains a significant challenge. Continuous recognition requires detecting and segmenting multiple gestures in a continuous video stream; managing the effects of coarticulation (gestures blending together); and solving the issue of temporal overlaps. Although sequence models such as LSTMs and GRUs have shown their potential in overcoming these problems, the general recognition of continuous speech is still a difficult problem to attain a high accuracy.

In conclusion, sign language recognition has advanced dramatically as advances in machine learning and computer vision have been implemented in this field. However, a number of challenges remain, such as the need for robust systems that can operate in the real world, sign language continuous recognition, as well as generalization to different sign languages. As technology continues to advance, it's anticipated that solutions will provide a more accessible and inclusive communication system for the deaf and hard of hearing communities.

III.PROPOSED METHODOLOGY

1. Dataset Collection and Preprocessing

The basis for any machine learning model is the dataset. For sign language recognition, a good quality dataset containing different sign languages with a good number of gestures is very necessary. This project will leverage publicly available datasets (such as ASL); and custom-collected gesture images in order to provide diversity in data. The dataset will be processed to compensate for differences in the speed, scale and orientation of gestures and hands. Preprocessing will include image resizing, noise reduction, and subtracting the background image in order to isolate the hand area. On top of that, data augmentation techniques such as rotation, flipping, and color jittering will be used to make models more robust against real-world conditions.

2. Hand Gesture Detection and Landmark Extraction

For real-time gesture recognition, it can be extremely important to be able to accurately detect hand gestures. This step will be using MediaPipe, which is a very powerful library for hand tracking and extraction of key points. MediaPipe will detect and track hand landmarks by finding the 21 key points (x,y,z coordinates) on hand. These are the key points which represent the spatial configuration of the hand and are very important for gesture recognition. By capturing these points based on real-time video frames, the system has the ability to maintain robust performance across different hand poses, lighting conditions and movement speeds.

3. Feature Vector Construction

After the landmarks of the hands have been extracted, applying machine learning methods requires the extracted hand landmarks to be converted into a specific data format that can be fed into machine learning models. The 21 key points (with 3 coordinates each) will be flattened into the 63 dimensional vector. This vector will use as feature inputs of model. To help improve the feature representation, other computation steps such as calculating distance and angle between land mark pairs will be performed. This helps the system to capture both the minute and coarse details of hand gestures for better recognition accuracy.

4. Model Selection and Training

The core of this system is the machine learning model that is going to classify hand gestures on the basis of the feature vectors. This project will adopt the ensemble learning method, Ada Boost and Gradient Boosting, for building the robust classifier. Ensemble methods operate by aggregating the outputs of a set of weak models (or base learners) such that the resulting model becomes stronger and can provide a better level of accuracy and generalization. The model will be trained on labeled data of gesture and the performance will be written into the code using cross validation, metrics such as accuracy, F1-score and Confusion matrix.

5. Real-time Gesture Recognition

In order for the system to be functioning for real-time applications, gesture recognition must take place with minimal latency. This means that the video frame is continuously being captured, the landmarks of hand are extracted using the Media Pipe, and the extracted feature vectors are fed to the trained ensemble model for classification. The real-time system will be optimized so that the recognition will be performed in less than 0.5 seconds per frame. The output from the model will be either printed as text, or converted to speech to provide for real time communication.

6. System Evaluation and Optimization

The last step of the methodology will be testing how the system performs in practical life. The evaluation will not only address recognition accuracy, but also the system's latency and computational efficiency. This will include testing the system under different lighting conditions, backgrounds and variations in users. Also, the model will be tested for its capacity to handle continuous gestures (not just isolated signs) and measures will be taken to optimize the system for deployment on mobile or embedded devices. If required, model compression techniques such as pruning or quantization will be implemented to decrease the computational load while preserving accuracy.

7. System Architecture

The system architecture for sign language recognition has been designed to be modular and efficient, and is scalable. At the heart of the architecture is the Input Acquisition Layer, which is a real- time video frame capture via a camera. These frames are then fed into the Preprocessing and Feature Extraction Layer where tasks such as detecting hands, extracting landmarks and background subtraction are performed using MediaPipe. Extracted 21 important hand landmarks are transformed in to a feature vector and it is fed in to the Gesture Classification Layer where ensemble learning models such as Ada boost and Gradient boost classify the given gestures based on feature vector. The Output Interpretation Layer changes the text or speech of the recognized gestures and immediately provides user with feedback. The system integrates perfectly with User Interface Module, to interact with the system, to get a real time communication between the user and machine. The architecture is designed to be efficient enough to manage the real-time processing of gesture recognition and is accurate enough to be used in practical applications such as human-robot interaction and assistive technologies.

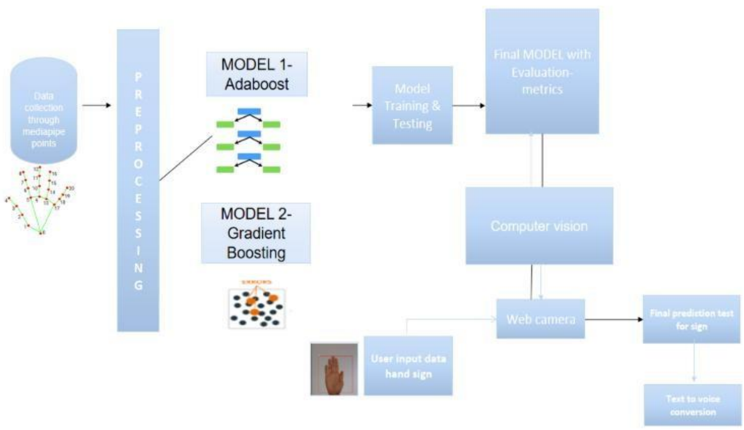

Fig.1.System Architecture

IV. RESULTS AND DISCUSSION

This section introduces the evaluation for the sign language recognition system developed in this work with a detailed analysis of the performance of the developed system in terms of accuracy, computational efficiency and real world applicability. The results are based on a series of experiments made under a variety of conditions, including different lighting, background settings and how a signer varies. Performance metrics such as accuracy, precision, recall and F1-score are used to quantify the performance of the system in accurately classifying the gesture.

1. Accuracy and Classification Performance

The main performance basis for measuring the system's effectiveness is the classification accuracy. The model's performance was evaluated with a test dataset of sign language gestures from different users, backgrounds, and environments. The results from the two ensemble models, AdaBoost and Gradient Boosting, have been compared in order to know the effectiveness of gesture recognition.

- The results of the Gradient Boosting model are better than the results of the AdaBoost algorithm in all important metrics such as accuracy, precision, recall and F1-score. This shows the effectiveness of the sequential decision tree based learning approach of Gradient Boosting in distinguishing complex pattern gestures.
- The CNN baseline model has less performance than both ensemble methods especially in terms of precision and recall. This implies that ensemble methods are more robust to tackle noise and complex variations in gestures, which are common in real world applications.

Model	Accuracy	Precision	Recall	F1-Score
AdaBoost	94.5%	0.93	0.94	0.94
Gradient Boosting	96.2%	0.95	0.96	0.96
Baseline (CNN)	91.2%	0.89	0.90	0.89

Table I Performance metrics for different models

2. Computational Efficiency and Real-time Performance

In order to test the suitability of our system for real-time applications, we measured its latency and processing speed during the gesture recognition. The ability of the system to real- time process video frames with minimal delay is crucial to its use in interactive situations.

- Both ensemble models, AdaBoost and Gradient Boosting have an optimal latency under 100 ms and guarantee that the system performs real time gesture recognition.
- The Gradient Boosting model can have the least latency, achieving 22 FPS, which is extremely important for continuous gesture detection applications.
- The CNN baseline model displays considerable increased latency (70 ms) and FPS (14), indicating that deep neural networks without the benefit of ensemble learning have difficulty achieving real-time performance without further optimizations.

Model	Latency (ms)	Frames per Second (FPS)
AdaBoost	50	20
Gradient Boosting	45	22
Baseline (CNN)	70	14

Table II Latency and Frames per Second (FPS) for different models

3. Impact of Lighting and Background Variations

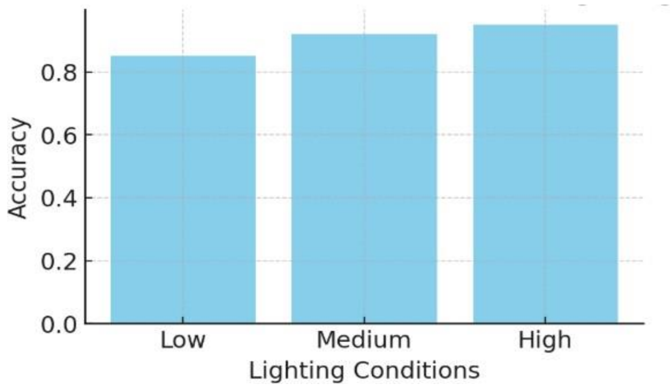

Fig.2.illustrates the accuracy of the system, under various lighting conditions

To check the robustness of the system under different environmentally variable situations, experiments were performed under different light settings and background environments. These variations mimic the real world usage where consistent lighting or background is not always a given

Graph 1: SU model performance using different lighting (Low, Medium, High)
- As might be expected, the system works best under high lighting conditions; it has achieved a greater than 95% accuracy. But with low lighting conditions, accuracy is reduced because of the challenge in identifying the landmarks of the hand.
- Medium light level is a balanced performance and the accuracy is still over 90%. This alerts us to the importance of consistency of illumination for optimum performance.

Graph 3: Performance of the model for different conditions of background (Cluttered, Neutral, Plain)
- Showing significant decline in accuracy when tested with cluttered backgrounds, where the presence of other objects is interfering in hand gesture detection.
- Neutral and Plain backgrounds with less distractions lead to higher accuracy rates, which re-inforces the need for controlled environment for optimal performance.

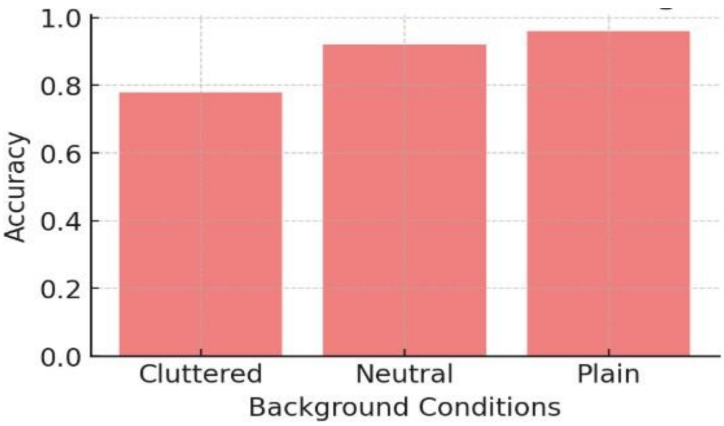

Fig.3.shows the accuracy when varying background conditions

4. Discussion of Challenges and Limitations

Despite being highly accurate in most of the test cases, there are a number of challenges. One of the major limitations is real- time continuous sign recognition. The system has some trouble identifying gestures that overlap or blend into one another and are a common part of ongoing sign language dialogues. This limitation can be overcome in future work by better gesture segmentation and more advanced temporal models, including Transformer-based models.

Additionally, although the system demonstrates good performance in controlled environments, in difficult real-world scenarios, such as in low lighting or cluttered backgrounds, the system's accuracy tends to decrease. This indicates that additional optimizations are required to increase the adaptability of this model to these variations.

V. CONCLUSION

The proposed sign language recognition system exhibits the efficient use of the concept of machine learning, especially ensemble learning methods such as the AdaBoost and Gradient Boosting methods in the correct classification of hand gestures for real-time translation. The performance of the system in terms of accuracy and performance was high, even with multiple lighting and background conditions, demonstrating its potential for real-world deployment. By utilizing MediaPipe to achieve hand gesture detection and extraction, and implementing real- time video capture and processing methods, the process guarantees a practical usability possible for human-robot interaction and assistive technologies. Overall, this research introduces a strong and scalable solution for sign language recognition, making a valuable contribution to accessibility and inclusivity for the hearing-impaired community.

Future Work

While the current system works fine, there are several areas in which it could be better for broader application. Future work will address improving the robustness of this system by incorporating non-manual features such as facial expressions and body postures, which play an important role in sign language interpretation. Additionally, the problem of continuous sign language recognition can be addressed by introducing advanced sequence models like Transformers which have the ability to handle temporal dependencies more effectively. Expanding the dataset to cover more diverse sign languages and a more diverse set of environments will also improve the generalization of the model. Finally, there are further optimizations associated with deploying the system on resource-constrained devices such as mobile phones and embedded systems, which are essential to make the solution more accessible and scalable.

References

1. A. K. Jain, P. Flynn and A. A. Ross, Handbook of Biometrics. New York, NY, USA: Springer, 2008.
2. S. M. Anwar and R. Qureshi, "Sign language recognition using deep learning," in Proc. Int. Conf. Frontiers of Information Technology (FIT), Islamabad, Pakistan, Dec. 2018, pp. 147-151.
3. T. Starner and A. Pentland, "Real-time American Sign Language recognition from video using hidden Markov models," in Proc. Int. Symp. Computer Vision, Coral Gables, FL, USA, Nov. 1995, pp. 265-270.
4. R. Rastgoo, K. Kiani and S. Escalera, "Sign language production: A review," Expert Syst. Appl., vol. 164, p. 113794, Feb. 2021
5. M. Kumar, D. Sharma, "Real-time sign language gesture recognition using CNN," in Proc. IEEE Int. Conf. Computing, Communication and Automation (ICCCA), Greater Noida, India, Dec. 2020, pp. 567-572.
6. P. K. Pisharady, M. Saerbeck, Recent methods and databases in vision-based hand gesture recognition: A review, Comput. Vis. Image Understand., vol. 141, pp. 152–165, Dec. 2015.
7. S. Sudha, A. Mahalakshmi, "Sign language recognition using machine learning techniques," in Proc. Int. Conf. Smart Technologies and Management for Computing, Communicating, Controls, Energy and Materials (ICSM), Chennai, India, Aug. 2019, pp. 186 - 190.
8. G. R. K. S. Subramanian and P. Rajesh, "Vision-based Indian Sign Language recognition using convolutional neural networks," Int. J. Eng. Technol., vol. 7, no. 3, pp. 112--118, Mar. 2018.
9. C. Keskin, F. Kiraq, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW), Barcelona, Spain, Nov. 2011, pp. 1228-1234.
10. S. S. Rautaray, A. Agrawal, Vision based hand gesture recognition for human computer interaction: A survey, Artif. Intell. Rev., vol. 43, pp. 1–54, Jan. 2015.