International Journal of
Algorithms Design and Analysis Review

Volume 1, Issue 2, 2023
July–December
DOI (Journal): 10.37591/IJADAR

https://journals.stmjournals.com/ijadar

Review IJADAR

ASL Mobile Translator with CNN Algorithm

Sasikumar M.[1], Lekha Achuth[2,*]

Abstract

There are around 63 million people in India with speech and hearing disabilities, and the number goes all the way up to 300 million across the world. All of them face issues in their day-to-day life as they can only have conversations with gestures. American Sign Language (ASL) mobile translator with convolutional neural network (CNN) algorithm is an easy-to-use mobile application, which uses complex images and video recognizing models built with an open-source tool called OpenCV. It efficiently recognizes gestures being shown to the camera and translates it to voice message, which can be understood by a normal person. Being a mobile application, it is portable and can be carried anywhere.

Keywords: American Sign Language, convolutional neural networks, image classification, machine learning, object detection

INTRODUCTION

ASL (American Sign Language) mobile translator with CNN (convolutional neural network) algorithm is a cutting-edge smartphone application that has completely changed how the deaf and mute community communicates. This ground-breaking program, which was created with their needs in mind, acts as a handy translator and eliminates the need for a middleman to conduct talks. ASL mobile translator with CNN algorithm has developed into a vital tool for the deaf and mute, enabling them to engage in effective communication thanks to its user-friendly design and cutting-edge gesture recognition technology [1]. ASL mobile translator with CNN algorithm's proficiency at effectively deciphering and interpreting hand gestures is one of its key strengths. The program can record and instantly analyze the user's hand movements by using the built-in camera on a mobile device. ASL mobile translator with CNN algorithm's advanced algorithms swiftly translate these movements into spoken or written discourse. Deaf and mute people can express themselves freely and efficiently without the aid of a middleman because of this quick translation capacity [2].

ASL mobile translator with CNN algorithm does more than only translate gestures. By translating speech communications into visual representations of hand movements, it also eliminates the communication barrier between spoken and sign languages. ASL mobile translator with CNN algorithm automatically transcribes audio messages sent through the app by hearing users and creates a gesture image to go with them. To help the deaf and mute user understand the message and reply accordingly, this graphic is shown to them. By fostering inclusion and greater communication between those who can hear and those who cannot, this ground-breaking function encourages accessibility. ASL mobile translator with CNN algorithm has had a significant positive impact on the deaf and mute community [3]. It gives them more freedom and independence in their daily life, enabling them to participate in interactions and conversations without difficulty. They can convey their thoughts, needs, and emotions without needing other people to serve as mediators. Their prospects for personal and

***Author for Correspondence**
Lekha Achuth
E-mail: lekha@pes.edu

[1]Student, Department of Computer Applications, People's Education Society University, Bengaluru, Karnataka, India
[2]Associate Professor, Department of Computer Applications, People's Education Society University, Bengaluru, Karnataka, India

Received Date: October 27, 2023
Accepted Date: November 02, 2023
Published Date: November 23, 2023

Citation: Sasikumar M., Lekha Achuth. ASL Mobile Translator with CNN Algorithm. International Journal of Algorithms Design and Analysis Review. 2023; 1(2): 31–38p.

professional advancement are increased thanks to their improved social connections, increased confidence, and improved communication skills.

Additionally, ASL mobile translator with CNN algorithm helps advance the larger objective of promoting inclusion and understanding in society. It promotes empathy, an appreciation for diversity, and the understanding of the significant contributions that people with various abilities bring to the table by removing communication obstacles between the deaf and mute community and the hearing community. People of all ages may use the app thanks to its user-friendly interface and intuitive design, which makes it possible for everyone to engage in meaningful conversations and establish deeper connections [4].

Finally, ASL mobile translator with CNN algorithm is a revolutionary smartphone software that has revolutionized communication for the deaf and mute community. It enables communication between deaf and mute people and gives them the freedom to express themselves freely and participate in conversations on their own by precisely reading hand gestures and allowing message exchange. ASL mobile translator with CNN algorithm is essential in fostering an inclusive and compassionate society because of its dedication to inclusivity and its support of understanding amongst various populations.

RELATED RESEARCH

Murthy and Jadon [5] presented a system in their study presented at the Advanced Computing Conference in 2010 that could recognize and decipher particular hand movements used to transmit information. The procedure entailed recording a user's hand motions and saving them as video clips. Then, these films underwent processing to become binary pictures and a three-dimensional (3D) space of binary values. The authors used a supervised feed-forward neural network to categorize the hand motions. Lack of looping connections between nodes is a characteristic of this kind of artificial neural network. The backpropagation technique was used by the system to train the neural network. In order to optimize the performance of the network, backpropagation requires changing the weights of the network based on the error rate or loss attained in the previous iteration. Overall, this study showed how feed-forward neural networks and backpropagation may be used to build a system that can recognize and decipher hand movements to efficiently transmit information [6].

Srivastava et al. [7] at the University of Allahabad, presented a sign language recognition system utilizing the TensorFlow Object Detection API in their paper in 2020. The system's goal was to create a real-time sign language detector by using transfer learning to train it on a unique dataset. The authors used a camera to take pictures while utilizing Python and OpenCV to collect the necessary data. The dataset for training the sign language detector was then built using these photos. For its reliability and effectiveness, the TensorFlow Object Detection API, a well-liked framework for object detection tasks, was used. The authors were able to use pre-trained models and modify them for sign language recognition by using transfer learning, which cut down on training time and enhanced system performance. The suggested system was designed to make sign language interpretation and real-time detection possible, helping to close communication gaps between sign language users and non-users.

Johnny and Nirmala [8] revealed a hand gesture recognition system particularly created for persons who are unable to talk or hear, commonly referred to as “dumb” people, in their research published in the *Springer Nature Computer Science* journal in 2022 [8]. The suggested method made use of information obtained from Fifth Dimension Technologies (5DT) gloves, specialized gloves that record hand motions and movements. The goal was to convert these hand gestures into words utilizing the highly accurate and high-quality Australian Sign Language signs collection. The authors investigated many machine learning methods to do this, such as neural networks, decision tree classifiers, and k-nearest neighbors. They tested the efficiency of these methods for understanding and interpreting sign language by putting them into practice in a project named “Talking Hands” and analyzing the outcomes. This study makes a contribution to the creation of tools that help people who use sign language to communicate overcome communication hurdles. The suggested hand gesture detection system has the potential to improve and empower people with speech and hearing disabilities' communication skills.

THEORETICAL FRAMEWORK

Sign Language Detection

The two fundamental kinds of signal-based correspondence affirmation are incessant correspondence by means of motions affirmation and separated motion-based correspondence affirmation. The essential separation between the two social affairs is the regulatory data. While restricted correspondence through motions affirmation is identical to the action affirmation field, steady signal-based correspondence affirmation is more stressed over genuine game plan between the data video segments and the fitting sentence-level imprints than just the affirmation issue itself. Affirmation of steady motion-based correspondence is generally more problematic than affirmation of withdrawn sign. Indeed, gathering separated signal-based correspondence affirmation as a subset of unending correspondence by means of motions acknowledgment is possible [9].

Neural Networks

A type of machine learning algorithm that is modeled after the structure of the human brain is called an artificial neural network, or simply an ANN. Numerous applications, including financial forecasting and analysis, make use of ANNs. Because of their ability to analyze complex data and make predictions based on patterns in the data, neural networks have been widely used in financial markets. These models can be trained to recognize patterns in previous data and then used to forecast future market movements. Some of the common uses of neural networks in finance include algorithmic trading, stock price prediction, credit risk evaluation, and fraud detection. Neural networks can also be used to assess unstructured data, such as news articles or social media postings, to help investors make more informed judgments [10].

However, using neural networks in the financial markets has several disadvantages. When a model performs well on training data but does not generalize well to new data in these models, overfitting may result. Furthermore, changes in market conditions may have an impact on how well these models estimate future events, making it necessary to retrain the models and regularly update their parameters. Despite these obstacles, the usage of neural networks in finance is growing, and new strategies are being developed to overcome the issues that these models present. Overall, neural networks can deliver useful insights into financial markets as well as improve investment decision-making processes [1].

Convolutional Neural Network

A significant learning network design known as a convolutional mind network propels directly from input, disposing of the essential for human component extraction.

1. *Convolution:* To eliminate features from pictures, convolution is a strategy used in significant learning and picture dealing with. The movement of convolutional channels are applied to the data pictures during the philosophy. Each channel contains an infinitesimal weighted system that is used to convolution the data picture mathematically [11]. Convolution incorporates getting the channel across the picture, expanding its characteristics by the relating picture pixels, and including the results. One more component map is made by reiterating this framework for each pixel in the picture. The outcome is an isolated interpretation of the main picture that causes explicit nuances like edges, surfaces, and models. By stacking various layers of convolutional channels, a significant cerebrum association can progress logically complex components from the data pictures [12].
2. *Rectified linear unit (ReLu):* It is an activation procedure that is consistently used in significant learning. The consequence of an immediate change, like a totally related layer or a convolutional layer, is presented to it part by part. Positive characteristics are left in salvageable shape while negative characteristics are wanted to zero by the ability. Through the introduction of non-linearity as needs be, the association is given the ability to learn seriously baffling data depictions. ReLU commencement is more computationally capable than other establishment capacities like

sigmoid or tanh since it essentially needs a reasonable relationship and errand action. To use incline-based improvement techniques to set up the association, the ability is furthermore differentiable.

3. *Pooling:* Pooling is a procedure for streamlining the consequence of the convolutional layers in convolutional mind associations. It accomplishes this by sampling the outcome, which cuts down its spatial viewpoints. In this manner, the association needs to learn less limits, which fabricates its viability and decreases its affinity for overfitting. There are various approaches to accomplishing pooling, including max pooling (which picks the most raised regard from a get-together of data values) and ordinary pooling (which concludes the common worth of a lot of data values). By cutting down the number of limits that ought to be dominated, pooling similarly upholds reducing the associations computational.

METHODOLOGY OVERVIEW

The designing of the structure is planned to beneficially assemble, handle, and inspect enormous volumes of data. To get the necessary data, the structure first attracts with a cloud programming connection point as shown in Figure 1. This data is cleaned, dealt with, and prepared for adjacent informational collection limit when it shows up at the data processor. As well as going probably as the fundamental data store for the data, the simulated intelligence module can include the close by informational index for planning and testing reasons. The man-made intelligence module uses this data to additionally foster execution and produce exact projections. The module could use the data to perceive models and examples as well as giving the client pieces of information and ideas. The front-end user interface (UI) is planned to be pleasant for data examination and course. Since the data show up through delineations and charts, the client may even more quickly understand the encounters and thoughts made by the artificial intelligence module. Utilizing channels and different data examination techniques, the UI similarly makes it possible to interface with the data. The designing is habitually expected to manage enormous volumes of data while giving the client steady assessment and encounters.

Talking about the implementation of the application. The categorical labels in training labels are converted to numerical labels using the Label Encoder class from Scikit-Learn's preprocessing package. The fit method is used to generate the Label Encoder instance le and fit it to the training_label data. In this stage, the encoder may identify the distinct labels in the data and create a mapping that associates each label with a corresponding numerical value. The training label’s categorical labels are then switched out for their numerical counterparts using the transform method. Using np.array(), the altered labels are turned into a NumPy array. The training_data variable is likewise changed into a NumPy array in a similar manner. With the help of these preparation stages, the data is properly encoded and prepared for training the machine learning model, enabling quick calculations and compatibility with the model architecture.

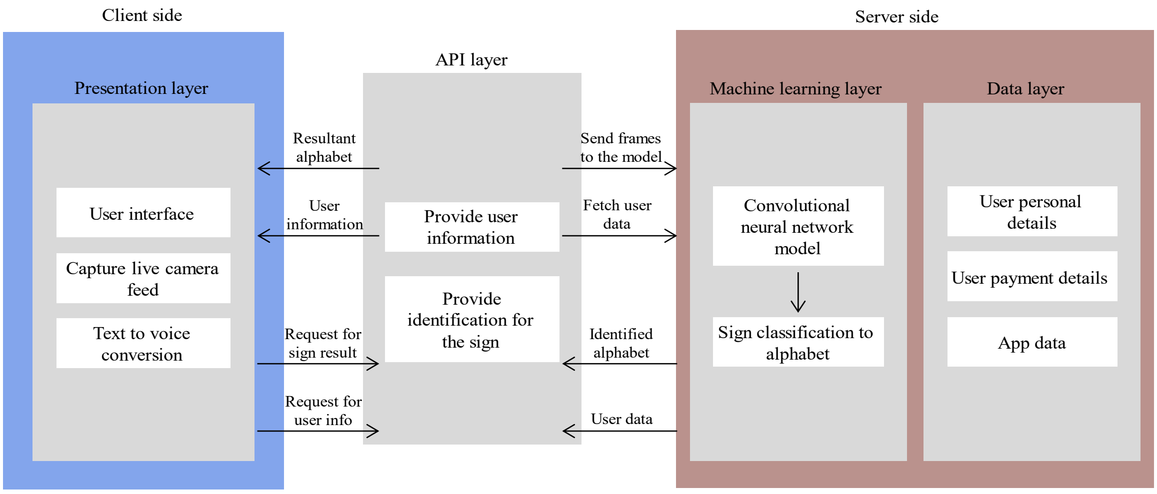

Figure 1. Architecture diagram for American Sign Language (ASL) mobile translator.

Layers are added sequentially, one after the other, to a model. Conv2D layer with 32 filters and kernel size of (7, 7) are the model's initial building blocks. It accepts input images with the shape (200, 200, 3), which stands for images with a 200-pixel width and height and three RGB color channels. The convolutional procedure travels one pixel at a time thanks to the stride of (1, 1). The feature maps are then down sampled using a MaxPooling2D layer that conducts max pooling with a pool size of (2, 2) after that. To add non-linearity, the ReLU activation function is used after the pooling layer. In order to prevent overfitting, a Dropout layer is subsequently introduced, which randomly disables 30% of the neurons during training. Another Conv2D layer, a second MaxPooling2D layer, and a second ReLU activation come next in the model.

The feature maps are then loaded into the fully connected layers after being flattened using a Flatten layer. With ReLU activation functions, three completely connected Dense layers are introduced, each with 128, 64, and 32 units. The output layer, which is appropriate for multi-class classification issues, consists of 25 units with a softmax activation function. The sparse_categorical_crossentropy loss function, accuracy as the evaluation measure, and the adam optimizer are all used in the model's compilation. This full implementation prepares the CNN model for training and testing on the preprocessed data. The proposed process flow diagram explains the complete flow of the application as shown in Figure 2.

The hand sign to voice message change system, explicitly, is figured out in the process stream diagram for the application. The client enters a picture of a hand sign to start the method. To set up the picture for input into the model, the image is then preprocessed, which incorporates exercises like picture scaling and normalization. The picture is preprocessed before being sent into the model, which uses computer-based intelligence strategies to calculate the hand movement depicted in the data picture. The American sign language motion based correspondence type of the foreordained letter set is then returned by the model. The program can make an understanding of hand signals into texts on account of this instrument. The resulting use case, in which the program recognizes a voice message from the client as data, is comparatively shown in the process stream graph. Talk-to-message development is then used to change over the sound correspondence into a text. The hand sign delineations for every

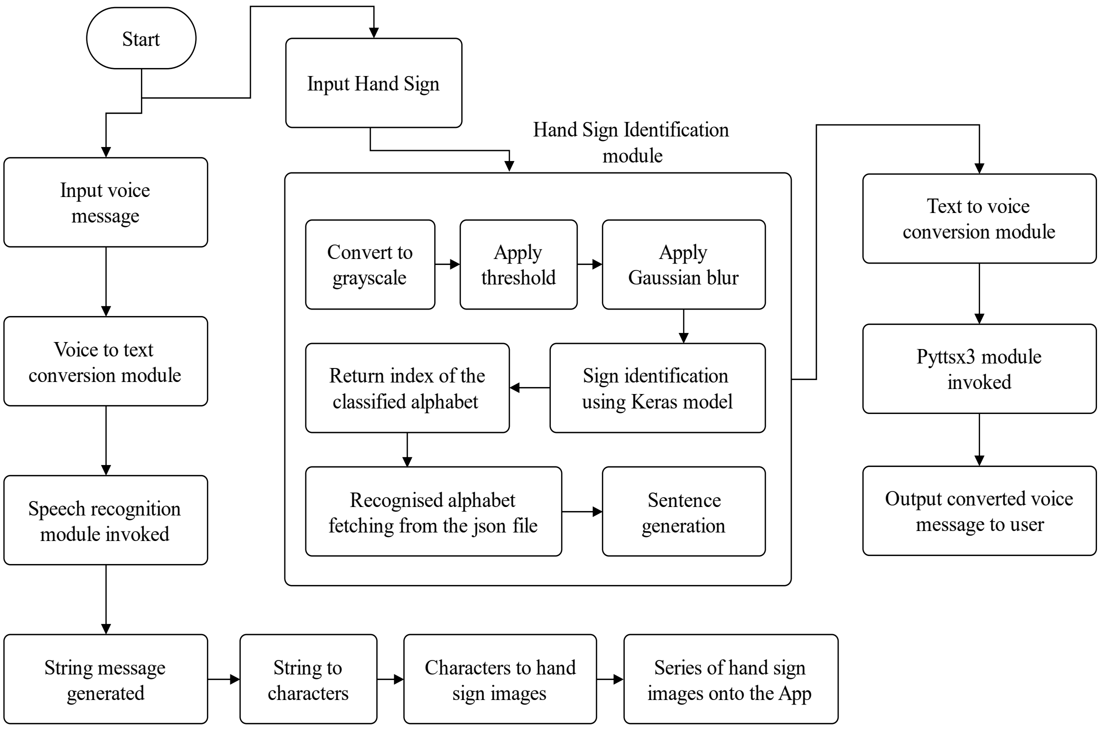

Figure 2. Process flow diagram for American Sign Language (ASL) mobile translator.

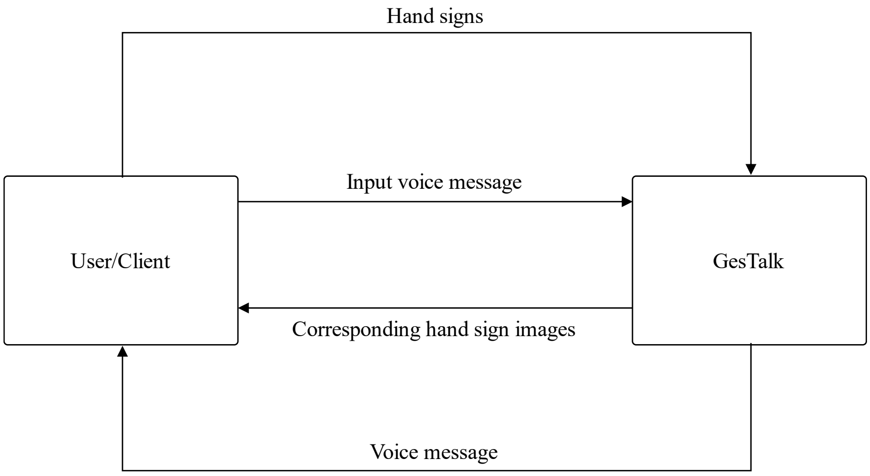

Figure 3. Data flow diagram for American Sign Language (ASL) mobile translator.

individual in the translated text are then effectively shown on the flexible application. This framework enables the program to make a translation of talk trades into hand signals, which might be important while talking with need of a hearing aid or hard of hearing individuals. The cooperation stream chart summarizes the whole technique by which the program changes two wellsprings of data — a hand sign and a communicated message — into the looking at yields — American Motion based correspondence and photographs of hand signals, independently. The program takes advantage of computer-based intelligence and typical language dealing with to achieve this collaboration.

The flow of data in the application is explained with the data flow diagram shown in Figure 3. The framework by which the program gets two information sources — a hand signal and a voice message and yields the connected voice message and hand sign pictures is shown in this data stream diagram. The client starts the framework by entering a voice message and a hand sign. To start with, a hand sign affirmation module gets the data and assessments and converts it into a modernized depiction. The information about these mechanized hand signals is then saved in an informational index for extra use. Using a talk affirmation module, the sound message input is changed into message. The message is subsequently transported off a module for ordinary language dealing with, where the structure assessments and handles it. Ensuing to taking care of, the text is saved in an informational index for future reference. The database is then searched for the saved hand sign data and voice message, which is in this way sent through an image creating module. The material hand sign picture and voice message picture are made by this module using the data that has been saved. Finally, the client comes by the outcome, which integrates the made pictures of hand signs and spoken messages. To do this, the program unites picture affirmation, standard language taking care of, and picture making propels. This methodology has applications in different endeavors, including correspondence, clinical benefits, and tutoring. The overall data stream frame shows how the program recognizes commitments of a hand signal and a voice message, examinations those information sources, saves them, and a short time later sends the fitting hand sign and voice message pictures back to the client.

Text to Hand Sign Translation

The final process of breaking down the translated text message into series of corresponding hand sign accuracy of its predictions.

RESULTS

Model Performance

The presented findings demonstrate how well a model performed during training and validation across 200 iterations. The model shows steady progress in terms of training accuracy, rising from 0.791 in the first epoch to 0.846 by the 200th. This shows that the model is successfully identifying patterns

Table 1. Model performed during training.

Epoch	Training accuracy	Training loss	Validation accuracy	Validation loss
100	0.791	0.13	0.783	0.087
200	0.846	0.06	0.871	0.044

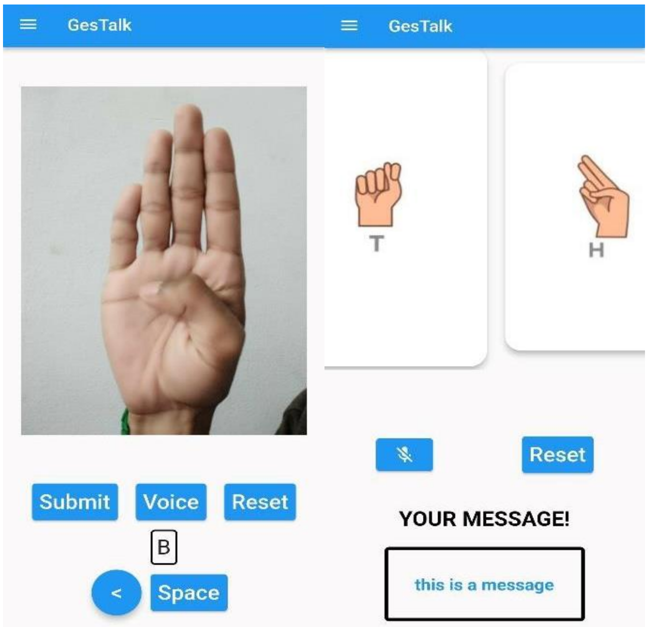

Figure 4. Test results of validation sets.

in the training data and learning from them. Similar to the training loss, which starts at 0.13 and drops to 0.06 throughout the course of the epochs as shown in Table 1. A smaller training loss shows that the model is reducing mistakes and improving accuracy of its predictions. The validation accuracy, which started at 0.783 and increased to 0.871 by the 200th epoch, is likewise outstanding. This shows that the model performs properly on the validation set and generalizes adequately to new data.

A similar pattern can be seen in the validation loss, which falls from 0.087 to 0.044. Lower validation loss indicates that the model is not overfitting and is capable of producing accurate forecasts for fresh data. The model appears to be learning successfully overall, as evidenced by its high accuracy and low loss on both the training and validation sets as shown in Figure 4.

CONCLUSION

As a powerful translator, ASL mobile translator with CNN algorithm is a mobile application that makes use of a variety of cutting-edge techniques and technology. Python, Flask, TensorFlow, Flutter, Android Studio, OpenCV, and Teachable Machine are just a few of the technologies that ASL mobile translator with CNN algorithm uses to provide a smooth and effective translating experience. Real-time language translation is made possible by the combination of these techniques, giving users an easy way to communicate in several languages. The application's outstanding performance and capacity for contextual adaptation are demonstrated by the training and validation accuracy, loss, and improvement over epochs metrics. ASL mobile translator with CNN algorithm is a living example of how technology may help people communicate more effectively by removing language barriers.

REFERENCES

1. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, June 27–30, 2016. pp. 770–778.

2. Jenkins J, Rashad S. An innovative method for automatic American Sign Language interpretation using machine learning and Leap motion controller. In: IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, December 1–4, 2021. pp. 633–638. doi: 10.1109/UEMCON53757.2021.9666640.
3. Boulares M, Jemni M. Mobile sign language translation system for deaf community. In: W4A 2012 – International Cross-Disciplinary Conference on Web Accessibility, Lyon, France, April 16–17, 2012. Article 37, pp. 1–4. doi: 10.1145/2207016.2207049.
4. Bird JJ, Ekárt A, Faria DR. British Sign Language recognition via late fusion of computer vision and Leap motion with transfer learning to American Sign Language. Sensors (Basel, Switzerland). 2020; 20: 5151. doi: 10.3390/s20185151.
5. Murthy GR, Jadon RS. A review of vision based hand gestures recognition. Int J Inform Technol Knowledge Manage. 2009; 2 (2): 405–410.
6. Tawalare S, Karale N, Pande S, Khamparia A. Identification of characters (digits) through customized convolutional neural network. Lecture Notes Data Eng Commun Technol. 2022; 122: 38–47. doi: 10.1007/978-981-16-6285-0_38.
7. Srivastava S, Gangwar A, Mishra R, Singh S. Sign language recognition system using TensorFlow Object Detection API. In: Woungang I, Dhurandher SK, Pattanaik KK, Verma A, Verma P, editors. Advanced Network Technologies and Intelligent Computing. Volume 1. Berlin, Germany: Springer; 2022. pp. Volume 1, pp. 634–646. doi: 10.1007/978-3-030-96040-7_48.
8. Johnny S, Nirmala SJ. Sign language translator using machine learning. SN Computer Sci. 2022; 3: 1–6.
9. Al-Saffar AAM, Tao H, Talab MA. Review of deep convolution neural network in image classification. In: International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications, (ICRAMET), Jakarta, Indonesia, October 23–24, 2017. pp. 1–6. doi: 10.1109/ICRAMET.2017.8253139.
10. Szegedy S, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 7–12, 2015. pp. 1–9. doi: 10.1109/CVPR.2015.7298594.
11. Goyal S, Sharma I, Sharma S. Sign language recognition system for deaf and dumb people. Int J Eng Res Technol. 2013; 2 (4): 382–387.
12. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017; 60 (6): 84–90. doi: 10.1145/3065386.

