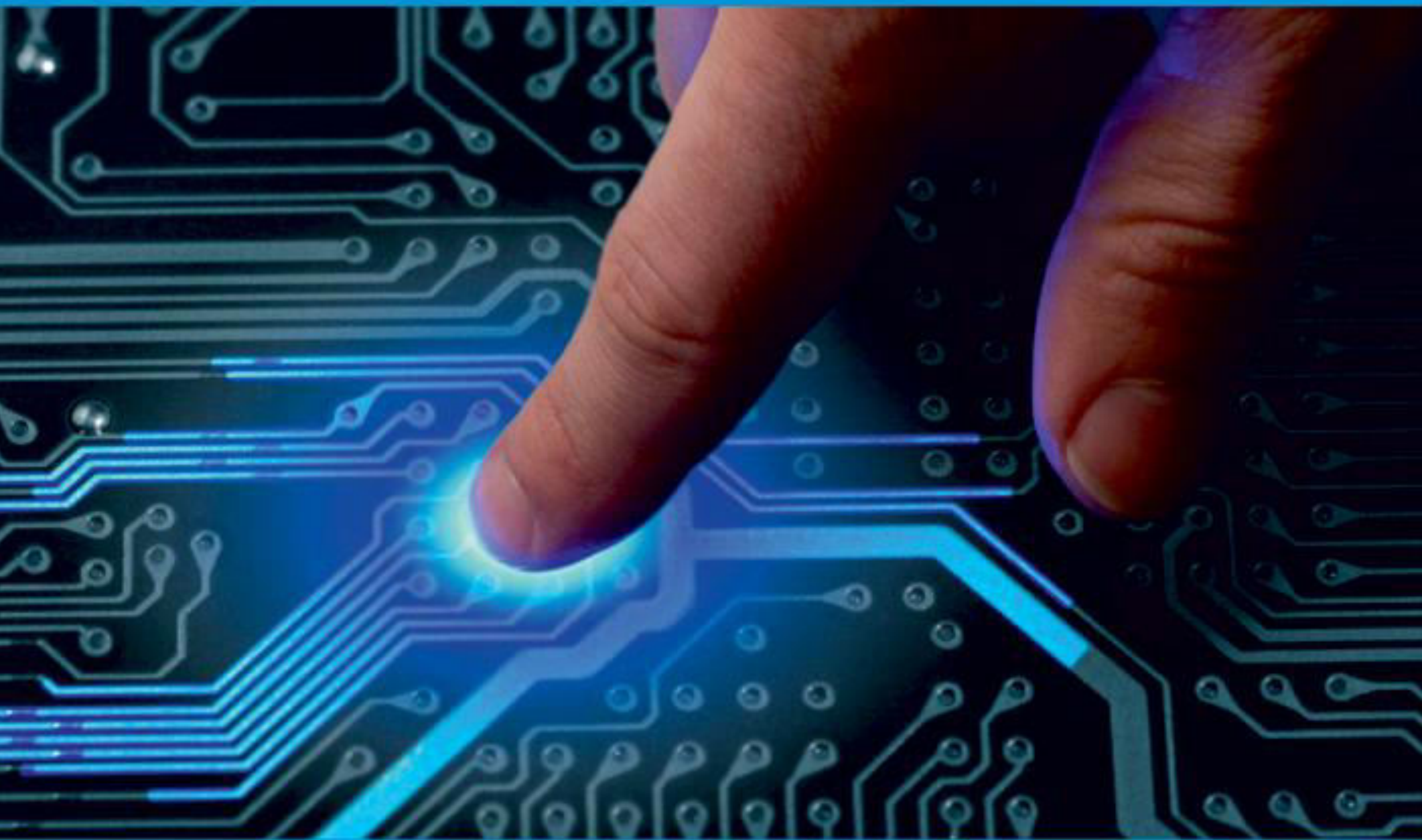
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 10, October 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

Gesture Recognition for Converting Sign Language to Text

Puja Borle, Tushar Kasture, Sakshant Khemnar, Rutwik Bhong, Prof . Anagha Chaudhari

UG Student, Dept. of Computer, PCCOE, Savitribai Phule Pune University, Pune, MH India

UG Student, Dept. of Computer, PCCOE, Savitribai Phule Pune University, Pune, MH India

UG Student, Dept. of Computer, PCCOE, Savitribai Phule Pune University, Pune, MH India

UG Student, Dept. of Computer, PCCOE, Savitribai Phule Pune University, Pune, MH India

Professor, Dept. of Computer, PCCOE, Savitribai Phule Pune University, Pune, MH India

ABSTRACT: The Sign language is a communication method for the deaf and dumb people. Problem arises when a dumb person is trying to communicate with a blind person. This paper proposes a method that provides a basis for the development of Hand Gesture Recognition System Using , KNN, Deep CNN, hand gesture recognition, Supervised Learning where the gesture presented by the dumb person will be converted into an observable format for a blind person.

KEYWORDS: Hand Detection, Hand-Tracking, Feature extraction, K-NN, Deep Learning CNN, Hand Gesture Recognition, Supervised Learning

I. INTRODUCTION

For the speech and hearing impaired community, sign language is one of the key communication languages and a primitive one. Because it uses the hands, face, and eyes rather than the vocal tract, it is a visual language. SLR research has been thriving since the 1990s, and it has progressed to the point where it is now. SLR has seen substantial advancements in performance in recent years. Automatic SLR is the process by which a machine recognises sign motions and translates them into human-readable text or speech. People with normal hearing can also examine the behaviours of the deaf and dumb when the SL is translated into speech. As a result, SL aids in the removal of numerous barriers in the workplace.

As a result, SL aids in breaking down barriers to communication between hearing and non-hearing people. SLR helps non-hearing people integrate into society. Vision-based SLR and sensor-glove-based SLR are the two main types of SLR. For gathering and presenting information, the vision-based approach utilises images, whereas the sensor-based approach uses gloves. The symbol is detected by the sensor's orientation. When compared to vision-based systems, sensor-based systems are more accurate. The maximum recognition accuracy is a challenge for sensor-based SLR.
This review focuses on recent trends in SLR and includes a comparative study of various approaches of SLR, feature mapping techniques and popular classification methods.

Successful recognition can only be achieved by exploring computer vision, machine learning (ML), human computer combination and deaf society . Recently, there is a lot of research in the area of automatic SLR and translation. The recent trend is the increased use of machine learning where a machine is trained with a number of data sets. Usually 80 percentages of the images are for training and 20 percentages, for testing. After training, various image processing steps are carried out on the captured image. This image is compared with the available data set and the sign is classified into its respective class. After classification, the sign is translated into speech using predefined libraries. Different software like ANACONDA, MATLAB etc is used to accomplish these processes. In ANACONDA prompt, python can be used for convenience.

This paper highlights various sign recognition approaches, feature extraction methods, classification methods etc. Further it discusses the challenges, limitations and future scope of automatic gesture recognition. Due to limitation of the conditions handled by the system and the variety of hand shapes, pre-processing keeps developing it to find the most accurate methods. The remaining sections of this paper are organized as follows: Section II describes the different approaches to SLR and Section III looks into the main feature extraction methods and Section

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

IV looks into the experimental results, Section V presents the conclusion of this review.

II. SIGN LANGUAGE RECOGNITION APPROACH

Over 300 SLs exist around the world, according to the "World Federation of the Deaf," and about 90 million individuals utilise them. Signs have phonological characteristics such as hand form, placement, and movement. There are also differences in how SL is executed depending on the locale, age, and hearing status. SLs are divided into American, Indian, and Chinese categories on a global scale. ASL is a natural language having phonemic qualities similar to spoken language and syntax distinct from English. Hand and face movements are used to convey it. ASL is the primary language of the deaf and hard-of-hearing communities in North America, and it is also spoken by a large number of non-deaf people. ISL is well-known in India.ISL is quite popular in India. The ISL of the alphabets A-Z is shown in Figure 1. Unlike ASL and European sign language, ISL is still in its early stages of development and employs two hands to represent the alphabet .The devices utilised to recognise the SL are used to categorise SLR techniques. Hand gesture recognition can be classified into the following categories: (i) Data glove approach and (ii) vision-based approach(instrumented)

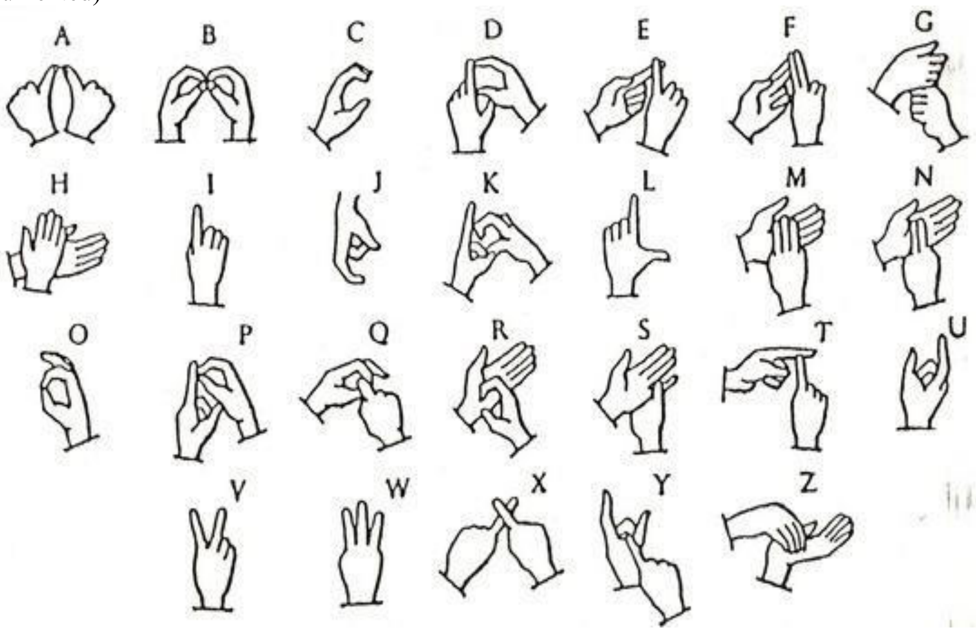


Fig.1:Indian_Sign_Language

III.PROPOSED ALGORITHM

Sensor-based technology necessitates the use of external devices to capture hand movement. However, this technique is limited. Vision-based technology has been introduced to address its limitations. This method only requires an image or video of the gesture. A 3D camera, numerous cameras, a webcam, and other devices are used to capture images. Below is a rudimentary block diagram of a hand gesture recognition system.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

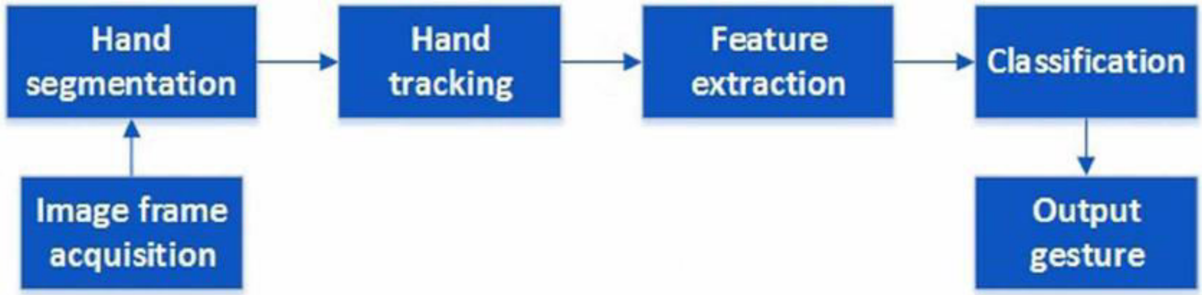


Fig 2 : Block Diagram of Hand Gesture Recognition System

There are several vision-based gesture recognition techniques commonly used for static and dynamic gesture recognition. We have discussed a few of those below:

1) Support Vector Machine:

The Support Vector Machine (SVM) is a supervised machine learning technique that is used to analyse data and recognise patterns in pattern recognition. It's utilised for regression analysis and categorization. It creates a training model and guesses an unknown sample's categorization. The goal of SVM modelling is to create a decision plane with a hyperplane that separates the positive and negative binary classes with the greatest margin. SVM is not confined to two-class classification problems; pair-wise top-down and bottom-up classification algorithms can be utilised to solve multi-class classification problems.

2) K-Nearest Neighbours (K-NN):

The K-nearest neighbour classifier is a machine learning approach that relies on feature vector distance. It finds the most common class among the k-nearest examples in the feature space to classify an unknown data object. The training images' feature vectors and labels are saved by the algorithm. One of the key drawbacks of the K-NN approach is that it compares all characteristics equally, which can lead to classification errors, especially when just a small subset of information is used for classification.

2) Convolutional Neural Network:

The most popular VBA for SLR is the convolutional neural network (ConvNets or CNNs). Object detection, classification, and recognition can all be done with CNN. CNN analyses an image, processes it, and then categorises it based on the likelihood level. To extract the feature, down sample the image to avoid overfitting, and converge the image into a compact feature vector, CNN can have multiple layers. Data is passed through multiple layers of CNN to train the system. Surejya et al proposed a CNN-based feature extraction and classification approach. After extracting features, the Visual Geometry Group (VGG16) is a new CNN for object categorization. There are 16 layers in VGG16.

1. VGG 16 :
VGG 16 is a pre trained model of CNN proposed by Zisserman. Sign language can be also recognised using this model. But there will not any localization (bounding box). For the hand gesture recognition there is no need of localisation. So we can also use VGG16 for the recognition of hand gesture. Figure 3 shows the architecture of VGG16 which consist of stack of layers.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

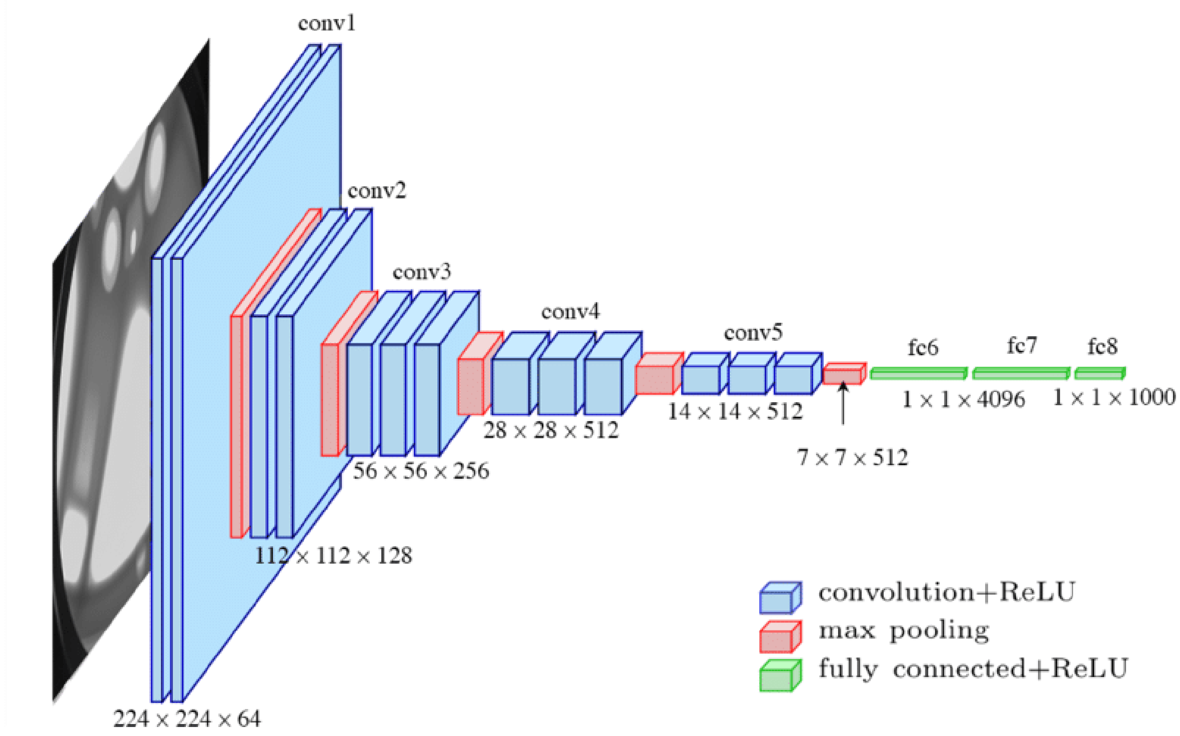


Fig 3: Architecture of VGG16

Input to the first Convolution Layer (CL) is an M x M image. Then the image is passed through consecutive CL which have many filters. It may be 3 x 3 or 1 x 1 sized. Pooling is done by the help of Max Pooling (MP) layer. VGG16 uses five MP layers which give spacial pooling. MP layer performs sampling over 2 x 2 pixel window. There also exist three Fully Connected (FC) layers in the end of the CL stack. There are three FC layers in VGG16. Using VGG 16 thousand classes can be classified. The last layer of this model is the Softmax Layer (SML) which is an output layer. Vgg16 do not give a localisation or bounding box to an object. Another method called YOLO which is more efficient method compared to VGG16.

IV. RESULTS

As input, a CNN takes tensors of shape (image _height, image _width, color _channels), ignoring the batch size. If you are new to these dimensions, color _channels refers to (R,G,B) and grayscale or black and white images channel is one. In this implementation we will configure our CNN to process inputs of shape (64, 64, 1). You can do this by passing the argument input _shape to your first layer. We can see that the output in fig 4 of every Conv2D and MaxPooling2D layer is a 3D tensor of shape (height, width, channels). The width and height dimensions tend to shrink as you go deeper in the network. The number of output channels for each Conv2D layer is controlled by the first argument (e.g., 32 or 64). Typically, as the width and height shrink, you can afford (computationally) to add more output channels in each Conv2D layer.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_2 (Conv2D)            (None, 62, 62, 32)        320
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 29, 29, 32)        9248
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 6272)              0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               802944
_________________________________________________________________
dense_3 (Dense)              (None, 6)                 774
=================================================================
Total params: 813,286
Trainable params: 813,286
Non-trainable params: 0
_________________________________________________________________
```

Fig 4: Output of CNN Model

To complete the model, we will feed the last output tensor from the convolutional base (of shape (29, 29, 32)) into one or more Dense layers to perform classification. Dense layers take vectors as input (which are 1D), while the current output is a 3D tensor. First, we will flatten (or unroll) the 3D output to 1D, then add one or more Dense layers on top. We use a final Dense
layer with 6 outputs. The network summary shows that (29,29, 32) outputs were flattened into vectors of shape (1024) before going through two Dense layers.

V. Conclusion

Different methods based on visual and data gloves can be used to recognise sign language. The majority of systems do not contain facial features. Feature extraction in VBA is a useful tool. YOLO, CNN, PCA, and other algorithms are used. Among The pre-trained model is the most recent and fastest of these techniques. It's also the best because it employs a large amount of data contribute to great precision, which is the primary goal of SLR stands for Single Lens Reflex. SVM, ANN, and CNN are used in the classification stage classifiers. All of these strategies are extremely accurate. Despite all of the proposed methodologies and trials that have been examined, a near-perfect system for sign translation remains a long way off. This review paper aims to give a layperson an overview of the field and to inform the reader about common SLR approaches.

References

1. M. Haenlein and A. Kaplan, 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence', Calif. Manag. Rev., vol. 61, no. 4, pp. 5–14, 2019.
2. M. Huang, R. Rust, and V. Maksimovic, 'The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI', Calif. Manag. Rev., vol. 61, no. 4, pp. 43–65, 2019.
3. S. Chundru, "Leveraging AI for Data Provenance: Enhancing Tracking and Verification of Data Lineage in FATE Assessment," International Journal of Inventions in Engineering & Science Technology, vol. 7, no.1, pp. 87-104, 2021.
4. F. Mod Ma'asum, S. Sulaiman, and A. Saparon, 'An Overview of Hand Gestures Recognition System Techniques', IOP Conf. Ser. Mater. Sci. Eng., vol. 99, p. 012012, 2015.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910017 |

5. F. Zhan, ‘Hand Gesture Recognition with Convolution Neural Networks’, in 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI) Information Reuse and Integration for Data Science (IRI), 2019, pp. 295–298
6. B. Xu, Z. Zhou, X. Chen, Y. Yang, and Z. Yang, ‘Arm removal for static hand gesture recognition’, J. Intell. Fuzzy Syst., vol. 35, no. 6, pp. 6489–6500, 2018.
7. N. Shah and J. Patel, ‘Gesture Recognition Technique:A Review’, Int. J. Recent Trends Eng. Res., vol. 3, no. 4, pp. 550–558, 2017.
8. Vimal Raja, Gopinathan (2017). Predicting Default Rates in Credit Scoring Models using Advanced Mining Algorithms. International Journal of Innovative Research in Science, Engineering and Technology 6 (12):23188-23193.
9. P. Premaratne, Human Computer Interaction Using Hand Gestures. Singapore: Springer, 2014.
10. K. Kennedy, ‘Sign Language’, Salem Press Encycl., 2020, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=ers&AN=93788
11. &site=eds-live&scope=site.
12. National Institute of Deafness and Other Communication Disordes, ‘American Sign Language’, NIDCD, Aug. 18, 2015.
13. Mohit, Mittal (2018). Exploring Generative Adversarial Networks (GANs) For Realistic Image Synthesis. International Journal of Innovative Research in Computer and Communication Engineering 6 (2):1720-1730.
14. https://www.nidcd.nih.gov/health/american-sign-language (accessed May 25, 2020).
15. Iurii Krak, Iurii Kryvonos, Waldemar Wojcik, ”Interactive Systems For Sign Language Learning”, IEEE Sixth International Conference on Application of Information and Communication Technologies (AICT), pp. 1-3, 2012.
16. Sujay S Kumar, Tenzin Wangyal, Varun Saboo, Dr. Ramamoorthy Srinath, ”Time Series Neural Networks For Real Time Sign Language Translation”, 17th IEEE International Conference on Machine Learning and Applications, pp. 81-88, 2018.
17. Shivashankara S, Srinath S, ”A Review on Vision Based American Sign Language Recognition, its Techniques, and Outcomes”, IEEE 7th International Conference on Communication Systems and Network Technologies, pp.293-299, 2017.
18. Pavan Reddy, Vaka (2014). Home Depot Breach or The Rise of Ransomware. Neuroquantology 12 (4):485-491.
19. W C Stokoe, D Casterlines and C Croneberg, A Dictionary of American Sign Language on Linguistic Principles, Linstok Press, 1965.
20. Muthu Mariappan, Dr. Gomathi V, ”Real-Time Recognition of Indian Sign Language”, Second International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6 IEEE.
21. Shinichi Tamura and Shingo Kawasaki, ”Recognition of Sign Language Motion Images” Pattern Recognition 21, 4 (1988), 343–353.
22. DOI: http://dx.doi.org/10.1016/00313203(88)90048-9