IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

SIGN LANGUAGE RECOGNITION USING TENSORFLOW-KERAS

[1]SOWMIYA S R, [2]SATHIYAVANI E, [3]SOUNDARYA M, [3]KAUSHICA K, [3]BRINDHA P
[1]Faculty of Computer Science & Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur.
[2]Computer Science & Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur.
[3]Computer Science & Engineering, Dhanalakshmi Srinivasan Engineering College, Perambalur.
Corresponding author: Sathiyavani E

ABSTRACT: Communication through signing is a fundamental instrument for individuals who haven't gotten the endowment of talking and tuning in. Ordinary individuals for the most part do not learn gesture-based communication to speak with hearing/discourse-hindered individuals. Likewise, every time these diversely abled individuals can't keep a human interpreter with them. This makes a correspondence hole among typical and hearing/discourse-disabled individuals. So, to overcome this issue, different gesture-based communication interpreters and acknowledgment methods have been created which are talked about in this paper. we got the inspiration to do this project from persons who are facing the difficulties of hearing impairment. So, we decided to derive a solution for this problem. In this paper, we propose a TensorFlow-Keras which is the model of CNN to manage this issue. TensorFlow is the most broadly utilized system because its adaptability additionally gives great comfort to troubleshooting TensorFlow applications. It very well may be considered a decent programming framework where tasks are sent as charts. Here, we used the Convolutional Neural Network for processing the images of the gestures. The Experimental results of our projects bring about our task show of agreeable recognizable proof of motions under different increase strategies. Besides, the proposed approach just requires a tiny number of preparing ages to accomplish 96.29 percent exactness.

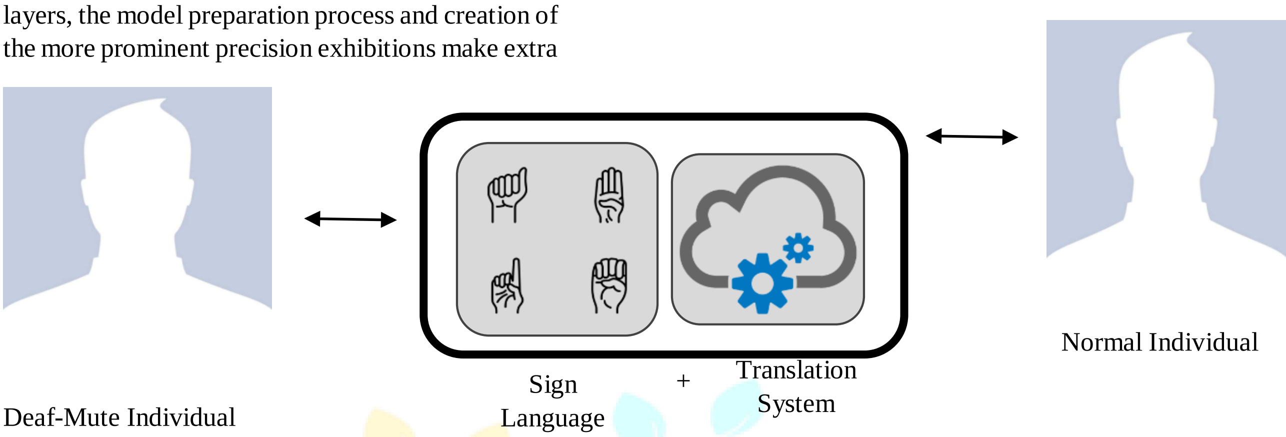
INDEX TERMS: Vision-Based Approach, Communication, TensorFlow, Keras, CNN, Gestures.

I INTRODUCTION

Correspondence is fundamental for all human lives to investigate their prerequisites and cooperation with others. Based on late examinations, different specialists tracked down a fascinating and novel style of correspondence in communication through signing across various nations. The communications through signing are visual prompts and coordinate the human manual and non-manual parts emphatically. It significantly upholds the need for a hearing aid and discourse disabled society in getting schooling, occupations, and cultural privileges. The state-run administrations of different countries corrected the various demonstrations to normalize communication through signing to help those in need of a hearing aid and discourse-disabled local area. Since gesture-based communication performs a significant job in need of a hearing aid and discourse-disabled correspondence, the comprehension and answering by the typical individuals requires extra preparation and information. This makes a correspondence hole between customary individuals and the impeded local area. The new progressions in profound learning procedures handle such undertaking productively by including various systems also, numerical methodologies. The advancement of such frameworks brings about tremendous intricacies in different periods of improvement, like misclassification, self-impediment, development epenthesis, equivocalness, clamor, and obscured yield. We explored this multitude of provokes originally to give a superior arrangement and meant to fabricate a strong design to give more noteworthy execution.

The rise of profound learning methods entered all the fields to show their solidarity towards vigorous model development. The profound learning procedures produce great brings about regions like agribusiness, irregularity identification, movement acknowledgment, business examination, crop choice, deformity checking, DNA frameworks, earth examination, extortion recognition, and genomic expectation, these advancements exceptionally persuade us to seek research in the profound learning region. Profound learning models are exceptionally strong and have created understandable accomplishments in a more extensive scope of uses. Notwithstanding, because of the complicated designs and larger number of layers, the model preparation process and creation of the more prominent precision exhibitions make extra difficulties during the model turn of events. These reasons make their appropriateness produce strong models for taking care of perplexing assignments.

As of late, the contribution of vision-based methods has become more famous, of which information is from the camera (web camera). Marcos utilized a variety of coded gloves to make hand discovery simpler. A blend of the two designs is moreover conceivable, which is known as the cross-breed design. While these are more reasonable and less obliging than information gloves, the shortcoming of this approach is lower exactness and high figuring power utilization. The design of these vision-based frameworks is normally separated into two principal parts. The initial segment includes extraction, which removes the ideal highlights from a video by utilizing picture-handling procedures or the PC vision strategy. From the extricated and portrayed highlights, the second part that is the recognizer ought to learn of the design from preparing information and the right acknowledgment of testing information on which machine calculations were utilized. A large portion of the examinations referenced above center around deciphering the signs regularly made by the consultation-disabled individual or the endorser to word(s) that the consultation larger part or non-underwriter can comprehend. Albeit that is what these examinations demonstrated innovation is valuable in such countless ways, their defenders think that these are nosy to a few hearing-weakened people. All things being equal, the defenders proposed a framework that will help those non-endorsers who need to learn essential sign language and not be meddling simultaneously. It is too critical to specify that there are applications carried out on cell phones that help the non-underwriter to learn. sign language through a few recordings introduced on the applications. Be that as it may, most of these applications require a lot of capacity and great web association. The proposed concentrate on plans to foster a framework that will perceive sign motions and convert them into comparing words. A Vision-based approach utilizing a web camera is acquainted with getting the information from the endorser and can be utilized disconnected. The reason for making the framework is that it will act as the learning apparatus for individuals who need to be aware more of the nuts and bolts of communication via gestures like letter sets, numbers, and normal signs. The advocates gave a white foundation and a particular area for picture handling of the hand, subsequently, working on the exactness of the framework and utilizing Convolutional Brain Organization (CNN) as the recognizer of the framework. The extent of the review incorporates essential static signs, numbers, and ASL letters in order (A-Z). One of the primary highlights of this study is the capacity of the framework to make words by fingerspelling without the utilization of sensors and other outer advancements. With the end goal of the review, a portion of the letters in ASL letters in order were changed. It very well may be seen that letters e, m, furthermore, n was overstated contrasted with the first motion, while j and z were changed over completely to motions by getting just their last edge. ASL additionally is severe with regards to the point of the hands while one is hand marking; once more, with the end goal of the review, the points of the hands for letters p, x, and t were changed for their uniqueness, which would incredibly influence the exactness of the framework. Here, we are utilizing convolutional Brain Organization for identifying the tokens of the hands. We as a whole realize that TensorFlow is the top system utilized in profound learning for anticipating the items. In this paper, we proposed to utilize the Keras model through the TensorFlow system to make a superior exactness. Aside from that soundness of the undertaking is generally significant. Thus, we use face detection to add the outlines of the face while acknowledging. It



builds the soundness of the application by staying away from the obstruction of different people into the casing. Subsequently, through the hand following and face acknowledgment, the motion focuses on the finger's edge won't be moved effectively thus the individual can utilize their gesture-based communication with practically no interference.

II MOTIVATION AND APPROACH

While the world has advanced in web-based connection through different devices, however, D&M (Deaf and Dumb) individuals face a language obstruction. Thus, they rely upon vision-based correspondence for the association. Assuming there is a typical point of interaction that switches the communication through signing over completely to the message the hand signals can be effectively grasped by the others. Thus, research has been made for a Vision-based interface framework where D&M individuals can partake in the correspondence without truly knowing one another's dialect. The point is to foster an easy-to-use sign-to-discourse transformation programming where the PC grasps human communication via gestures. There are different gesture-based communications everywhere American Gesture-based communication (ASL), French Gesture-based communication, English Gesture-based communication (BSL), Indian Communication through signing, Japanese Communication through signing, and work has been finished on different dialects from one side of the planet to the other.

The problem is to make a PC programming and train a model utilizing CNN which takes a picture of a hand token of American Communication through signing and gives the result of the indication language in text design changes over it into sound configuration and make a totally practical item for individuals who cannot hear, so that, they can get associated with the world without any problem. It had utilized and comprehended advancements like OpenCV, Matplotlib, Keras, Profound Learning, Python, and so forth. The main aim of this project is to make a user-friendly system for the translation of sign language into text using TensorFlow and Keras.

III RELATED WORK

In recent years there has been tremendous research done on gesture recognition. With the help of a literature survey we, realized the basic steps in hand gesture recognition are:-

A Data acquisition

The different approaches to acquiring data about hand gestures can be done in the following ways: – Use of sensory devices It uses electronic devices to provide exact hand configuration, and position. Different glove-based approaches can be used to extract information Bu t it is expensive and not user-friendly – Vision-based approach In vision-based methods computer camera is the input device for observing the information of hands or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing 7 artificial vision systems that are implemented in software and/or hardware. The main challenge of vision-based hand detection is to cope with the large variability of the human hand's appearance due to a huge number of hand movements, different skin-color possibilities as well as to variations in view scales, and speed of the camera capturing the scene.

B Data pre-processing and Feature extraction

In [1] the approach for hand detection combines threshold-based color detection with background subtraction. We can use an XGBoost face detector to differentiate between faces and hands as both involve similar skin color – We can also extract the necessary image The filter can be easily applied using open computer vision also known as OpenCV and is described in [3]. – For extracting the necessary image which is to be trained we can use instrumented gloves as mentioned in [4]. This helps reduce computation time for pre-processing electronics and can give us more concise and accurate data compared to applying filters on data received from video extraction. – We tried doing the hand segmentation of an image using color segmentation techniques but as mentioned in the research paper skin color and tone are highly dependent on the lighting conditions due to which output we got for the segmentation we tried to do were not so great. Moreover, we have a huge number of symbols to be trained for our project many of which look similar to each other like the gesture for the symbol 'V' and digit '2', hence we decided that to produce better accuracies for our large number of symbols, rather than segmenting the hand out of a random background we keep the background of hand a stable single color so that we don't need to segment it based on skin color. This would help us to get better results. Gesture classification – In [1] Hidden Markov Models (HMM) are used for the classification of the gestures. This model deals with dynamic aspects of gestures. Gestures are extracted from a sequence of video images by tracking the skin-color blobs corresponding to the hand into a body–face space centered on the face of the user. The goal is to recognize two classes of gestures: deictic and symbolic. The image is filtered using a fast-lookup indexing table. After filtering, skin color pixels are gathered into blobs. Blobs are statistical objects based on the location (x, y) and the colorimetry (Y, U, V) of the skin color pixels to determine homogeneous areas [2] Naive Bayes Classifier is used which is an effective and fast method for static hand gesture recognition. unlike many other recognition methods, this method is not dependent on skin color. The gestures are extracted from each frame of the video, with a static background. The first step is to segment and label the objects of interest and extract geometric invariants from them. The next step is the classification of gestures by using a K nearest neighbor algorithm aided with a distance weighting algorithm (KNNDW) to provide suitable data for a locally weighted Naive Bayes classifier. – According to a paper on "Human Hand Gesture Recognition Using a Convolution Neural Network" by Hsien-I Lin, Ming-Hsiang Hsu, and WeiKai Chen graduates of the Institute of Automation Technology National Taipei University of Technology Taipei, Taiwan, they construct a skin model to extract the hand out of an image and then 9 apply the binary threshold to the whole image. After obtaining the threshold image they calibrate it about the principal axis to center the image about it. They input this image into a convolutional neural network model to train and predict the outputs. They have trained their model over 7 hand gestures and using their model they produce an accuracy of around 95% for those 7 gestures.

IV METHODOLOGY:

The framework will be carried out through a work area with a 1080P Full-HD web camera. The camera will catch the pictures of the hands that will be taken care of in the framework. [20] Note that the underwriter will conform to the size of the casing so that the framework will want to catch the direction of the underwriter's hand. At the point when the camera has caught the motion from the client, the framework orders the test and thinks about it in the put-away motions in a word reference, and the related yield is shown on the screen for the client.

A) DATASET GENERATION:

For the venture, we attempted to find as of now made datasets, yet we were unable to find the dataset as crude pictures that matched our prerequisites. Consequently, we chose to make our informational index. The means we followed to make our informational index are as per the following.

[9]We have chosen to utilize the Open PC vision(OpenCV) library to produce our dataset. First and foremost, we will catch around 800 pictures of each of the images in ASL for the end goal of preparing and around 200 pictures for every image for testing. To start with, we catch each edge shown by the webcam of our machine. In each casing, we characterize a district of interest (return on initial capital investment) which is indicated by a blue-limited square. From this entire picture, we extricate our return for capital invested which is RGB, and convert it into a dim-scale Picture.

B) GESTURE CLASSIFICATION

Algorithm Layer (CNN):

(a)1st Convolution Layer: The input picture has a resolution of 128x128 pixels. It is first processed in the first convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 126X126 pixel image, one for each Filter-weights.

(b) 1st Pooling Layer: The pictures are downsampled using max pooling of 2x2 i.e we keep the highest value in the 2x2 square of the array. Therefore, our picture is downsampled to 63x63 pixels

(c) 2nd Convolution Layer: Now, these 63 x 63 from the output of the first pooling layer is served as an input to the second convolutional layer. It is processed in the second convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 60 x 60-pixel image.

(d) 2nd Pooling Layer: The resulting images are downsampled again using a max pool of 2x2 and are reduced to 30 x 30 resolution of images.

(e) 1st Densely Connected Layer: Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of 30x30x32 = 28800 values. The input to this layer is an array of 28800 values. The output of this layer is fed to the 2nd Densely Connected Layer. We are using a dropout layer of value 0.5 to avoid overfitting.

(f) 2nd Densely Connected Layer: Now the output from the 1st Densely Connected Layer is used as an input to a fully connected layer with 96 neurons.

(g) Final layer: The output of the 2nd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

C) ACTIVATION FUNCTION:

We have chosen to utilize ReLU (Rectified Straight Unit) as our enactment capability in each layer. It is a basic capability expressed as

f(x) = max (0, x) for each info pixel. Using this actuation function we would have the option to diminish the required calculation power, Gaussian Plummet losing issues, and so on.

D) POOLING LAYER:

Pooling is a component regularly soaked up into Convolutional Brain Organization (CNN) models. The fundamental thought behind a pooling layer is to "collect" highlights from maps created by convolving a channel over a picture. Officially, it can logically diminish the spatial size of the portrayal to decrease how many boundaries and calculations are in the organization. The most widely recognized type of pooling is max pooling. Max pooling is finished to a limited extent to help over-fitting by giving a preoccupied type of portrayal. Also, it diminishes the computational expense by lessening the number of boundaries to learn and gives essential interpretation invariance to the inside portrayal. Max pooling is finished by applying a maximum channel to (normally) non-covering subregions of the underlying portrayal. Different types of pooling are: normal, and general[17].

We apply Max pooling to the info picture with a pool size of (2, 2) with a ReLu actuation function. This decreases how much boundaries subsequently reducing the calculation cost and lessening overfitting [8].

E) Optimizer:

We will utilize the Adam analyzer for refreshing the model in light of the result of the misfortune capability. Adam consolidates the benefits of two expansions of two stochastic inclination drop calculations to be a specific versatile slope algorithm (ADA Graduate) and root mean square propagation (RMSProp)

F) Training the system:

The preparation for character and SSL acknowledgment was finished independently; each dataset was partitioned into two: preparing and testing. This was finished to see the exhibition of the calculation utilized. The organization was carried out and prepared through Keras and TensorFlow as its backend utilizing a Designs Handling Unit GT-1030 GPU. The organization utilizes a Stochastic inclination plunge enhancer as its enhancer to prepare the organization to have a learning pace of 1 × 10−2. The all-out number of ages used to prepare the organization is 50 ages with a clump size of 500. The pictures were resized to (50, 50, and 1) for preparation and testing. [13] We utilize a stochastic inclination plummet enhancer, too known as the gradual slope drop, to limit the cluster size of huge datasets. The cluster slope plunge performs excess calculations for enormous datasets as slopes are recalculated before every boundary update for comparative models. By performing each update in turn, SGD kills this overt repetitiveness. It is normally a lot quicker and can likewise be utilized for web-based learning [6].

V TESTING

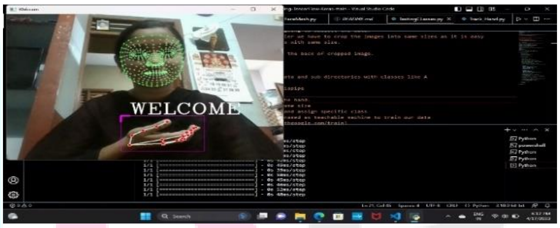
The venture was tried by 30 people: 6 were sign language mediators, and 24 were understudies with and without information in utilizing communication through signing. Thirty examples were liked to have

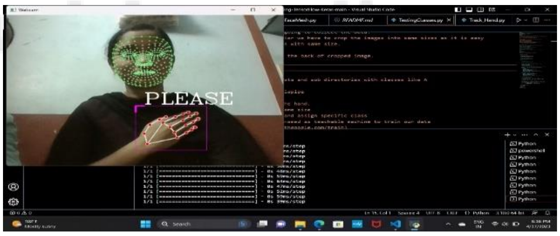
$$\text{Accuracy rate} = \frac{\text{Total nos. of correct recognized letters from user}}{(\text{Total no. of users})(\text{No. of Trails})}$$

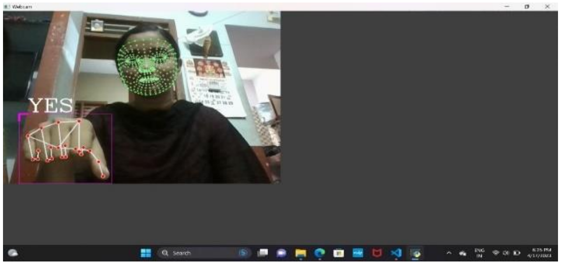
the option to utilize the Understudy's t-test or the importance approval of the review and to demonstrate the dependability of the framework in perceiving static hand signals from the hands of individuals who were not in the dataset Three preliminaries were directed in each letter/number/word signal acknowledgment. Every preliminary has a span of 15 seconds per letter/number/word signal acknowledgment. On the off chance that the framework did not print the same text of the signs inside the assigned time, the result was viewed as inaccurate.

A. Testing Procedure

Before the genuine acknowledgment of the signs, the client must adjust the light to guarantee that the skin covering of the hand is recognized and has less commotion; the alignment should be possible by moving the lampshade sideways. It is suggested that







the light isn't straightforwardly raising a ruckus around town. The framework is delicate to light; consequently, deciding the appropriate spot of the light ought to be thought of. On the off chance that the edges of the hand in the concealing are recognized obviously, the client might start to utilize the interpreter. For the signs to be perceived, the hand should be before the camera. The discovery must be finished assuming the hand is inside the case that should be visible on the screen of a PC's screen. Since the size of the hand of every individual is unique, a client might move his/her hand this way to fit inside the virtual box. The client ought to then hang tight for the framework to produce what could be compared to the signs in a text-based structure. It additionally suggested that the client's hand does not make any development until the framework creates the yield. To know the pace of learning, the scientist gauges the season of creating the deciphered static signs utilizing a stopwatch and rehashes this multiple times.

B. Testing of accuracy formula

To check the precision of the letter/number/word motions acknowledgment, the quantity of the accurately perceived letters/words/numbers that showed up on the screen was added also, separated by the result of the complete number of clients increased by the number of preliminaries. The right acknowledgment is gained when the signs made by the client are deciphered and their separate counterparts are created in a text-based structure within the term of 15 seconds. If the framework creates the same word/letter/number past 15 seconds, it is excluded from the complete number of right perceived letters/words/numbers.

VI RESULTS:

In this report, a utilitarian constant vision-based Indian communication through signing acknowledgment for not-too-sharp individuals has been produced for ASL alphabets. We can accomplish the last precision of 98.0 percent on our informational collection. We would have the option to get to the next level of our expectation after carrying out two layers of calculations in which we check and anticipate images that are more like one another. This way we can identify practically every one of the images given that they are shown appropriately, there is no commotion in the foundation and the lighting is sufficient. we had utilized TensorFlow and Keras which make the model for anticipating the gestures. It offers reliable and basic APIs, limits the number of client activities expected for normal use cases, and gives clear and noteworthy criticism of client mistakes.

VII CONCLUSION:

Gesture-based communication acknowledgment frameworks have heaps of potential applications in the field of human-PC connection. A Vision-based static sign motion acknowledgment framework is fundamental to diminish the correspondence hole between ordinary individuals and outwardly weakened individuals. The proposed technique presents transformer-based communication via gesture acknowledgment for static signs. A multi-head consideration-based encoding structure can accomplish great exactness with a tiny number of preparing layers and ages.

In this report, a utilitarian constant vision-based American communication through signing acknowledgment for not-too-sharp individuals has been produced for ASL alphabets. We can accomplish the last exactness of 98.74 percent on our informational collection. We would have the option to work on our forecast after executing two layers of calculations in which we confirm and predict images that are more like one another. This way we can distinguish practically every one of the images given that they are shown appropriately, there is no clamor in the foundation and the lighting is satisfactory.

VIII REFERENCES

[1] B. Saunders, N. C. Camgoz, and R. Bowden, ‘‘Progressive transformers for end-to-end sign language production,’’ in Proc. Eur. Conf. Computer. Vis. (ECCV), 2020, pp. 687–705.

[2] B. Saunders, N. C. Camgoz, and R. Bowden, ‘‘Continuous 3D multi-channel sign language production via progressive transformers and mixture density networks,’’ Int. J. Computer. Vis., vol. 129, no. 7, pp. 2113–2135, Jul. 2021.

[3] B. Saunders, N. C. Camgoz, and R. Bowden, ‘‘Mixed Signals: Sign language production via a mixture of motion primitives,’’ in Proc. IEEE/CVF Int. Conf. Computer. Vis. (ICCV), Oct. 2021, pp. 1919–1929

[4] G. A. Gates and J. H. Mills, ‘‘Presbycusis,’’ Lancet, vol. 366, no. 9491, pp. 1111–1120, Sep. 2005.

[5] K. Emmorey, Language, Cognition, and the Brain: Insights from Sign Language Research. New York, NY, USA: Psychology Press, 2001.

[6] L. Goyal and V. Goyal, ''Text to sign language translation system: A review of literature,'' Int. J. Synth. Emotions (IJSE), vol. 7, no. 2, pp. 62–77, 2016.

[7] M. J. Cheok, Z. Omar, and M. H. Jaward, ''A review of hand gesture and sign language recognition techniques,'' Int. J. Mach. Learn. Cybern., vol. 10, no. 1, pp. 131–153, Jan. 2017.

[8] G. A. Haidar, R. Achkar, D. Salhab, A. Sayah, and F. Jobran, ''Sign language translator using the backpropagation algorithm of an MLP,'' in Proc. 7th Int. Conf. Future Internet Things Cloud Workshops (FiCloudW), 2019, pp. 31–35.

[9] M. Hassan, K. Assaleh, and T. Shanableh, ''Multiple proposals for continuous Arabic sign language recognition,'' Sens. Imag., vol. 20, no. 1, p. 23, Jan. 2019.

[10] Z. Ren, H. Li, C. Yang, and Q. Sun, ''Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning,'' Knowl.-Based Syst., vol. 188, Jan. 2020, Art. no. 105040.

[11] I. Lauriola, C. Gallicchio, and F. Aiolli, ''Enhancing deep neural networks via multiple kernel learning,'' Pattern Recognit., vol. 101, May 2020, Art. no. 107194.

[12] Q. Xiao, Y. Zhao, and W. Huan, ''Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network,'' Multimedia Tools Appl., vol. 78, no. 11, pp. 15335–15352, Jun. 2019.

[13] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, and D. A. Kumar, ''3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2–stream CNN,'' Neurocomputing, vol. 372, pp. 40–54, Jan. 2020.

[4] M. De Coster, M. Van Herreweghe, and J. Dambre, ''Isolated sign recognition from RGB video using pose flow and self-attention,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2021, pp. 3441–3450.

[15] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, and Q. Xin, ''Hypertuned deep convolutional neural network for sign language recognition,'' Comput. Intell. Neurosci., vol. 2022, pp. 1–10, Apr. 2022.

[16] M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. Guo, and M. M. Elahi, ''Sign language recognition for Arabic alphabets using transfer learning technique,'' Comput. Intell. Neurosci., vol. 2022, pp. 1–15, Apr. 2022.

[17] M. M. Kamruzzaman, ''Arabic sign language recognition and generating Arabic speech using convolutional neural network,'' Wireless Commun. Mobile Comput., vol. 2020, pp. 1–9, May 2020.

[18]Ministry of Health, Labour and Welfare, ''The survey of difficulties of daily life (survey of nationwide home children with disabilities, and others, 2011),'' (in Japanese), Dept. Health Welfare Persons Disabilities, Tokyo, Japan, Tech. Rep., 2013.

[19] S. Tanaka, M. Takuma, and Y. Tsukada, ''Recognition of finger alphabet remote image sensor,'' (in Japanese), in Proc. 76th Nat. Conv. (IPSJ), 2014, vol. 76. no. 2, pp. 2275–2276.

[20] S. He, ''Research of a sign language translation system based on deep learning,'' in Proc. Int. Conf. Artif. Intell. Adv. Manuf. (AIAM), Oct. 2019, pp. 392–396.

[21] E. K. Kumar, P. V. V. Kishore, M. T. Kiran Kumar, and D. A. Kumar, ''3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2 stream CNN,'' Neurocomputing, vol. 372, pp. 40–54, Jan. 2020.