International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 09, September 2022

II. LITERATURE REVIEW.

Saba Joudaki et.al.,8] have presented given a summary of the most important recent research on the vision-based SLR system. The discussion of the current recognition methods focuses on the video-based SLR system and its ability to execute continuous SLR within video sequences. Locating each finger's tip by combining SLR and finger detection is an innovative method. The accuracy rate of this approach, which may be used to identify a class of hand gestures, is close to 96%.

Manisha U.Kakde et.al.,[21] have discussed the basis of sign acquiring methods and sign identifying methods and the many approaches to sign language recognition. ANN establishes a compelling case for the sign acquisition method and the sign identification approach.

Jayshree R.Pansare and Maya Ingale [7] have discussed the American Sign Language Recognition (ASLR) system, which uses static hand gestures, natural lighting, and a complicated background to recognize 26 ASL alphabets. The EOH approach is based on the EOH descriptor model, which calculates the degree of similarity between the running image's EOH and the ASL alphabet images in the training dataset. 88.26% recognition rate, as well.

Sakshi Sharma and Sukhwinder Singh [5] have presented the fundamental idea of a sign language recognition system focused specifically on sign language and used a vision-based approach. To facilitate communication between signers and non-signers, a sign language interpreter is necessary.

M.Ebrahim Al-Ahdal and Nooritawati Md Tahir [20] have explained the discussion is on the Sign Language recognition system, which has been built and is divided into sign capture and recognition procedures. ANN classifiers are suggested as the training process for the required training process as well as the training process needed for extending new vocabulary in a novel way for developing an SLR system based on merging EMG sensors with a data glove [13]. Because it can model words based on collections of predetermined states, the Hidden Markov Model (HMM) classifier also shows intrigue for sign recognition.

Pratibha Pandey and Vinay Jain [22] have proposed Different techniques for hand gesture and sign language recognition. To recognize hand gestures automatically and enable interaction between humans and computers, a hand gesture recognition system is currently being developed. Once complete, this system will be able to control robots and recognize sign language. To capture the gesture and hand posture in sign language, various types of algorithms are used. In the field of hand gesture recognition, vision-based gesture recognition has made impressive strides.

Paulraj M P et.al.,[9] have discussed that Using features derived from head and hand gestures, a basic way of translating sign language into audio signals can be utilized by hearing-impaired individuals to interact with hearing-impaired individuals. For extracting the features from the video sign language, a straightforward features extraction method based on the area of the object in a binary picture and Discrete Cosine Transform (DCT) is suggested. Using the features calculated from the video stream, a straightforward neural network model is created for the recognition of gestures. The categorization rate for the developed system is 92.07%.

Sakshi Sharma and Sukhwinder Singh [17] have proposed the purpose of recognizing gesture-based sign language, and a convolutional neural network (CNN) model based on deep learning has been developed. In this work, VGG-11 and VGG-16 have also been trained and tested to assess the effectiveness of this model. The suggested model achieves the greatest accuracy of 99.96% and 100% for the ISL and ASL datasets, respectively.

Maher Jebali et.al.,[13] have discussed identifying each sign in continuous sign language videos. Therefore, this work presents a computer vision-based system to recognize the signs in continuous sign language video. Signs are classified and recognized using Hidden Markov Model (HMM) and it has been strongly adopted after testing other approaches such as Independent Bayesian Classifier Combination (IBCC). Our system manifests auspicious performance with a recognition accuracy of 95.18% for one gesture and 93.87% for two hand gestures.

P. V. V. Kishore and P. Rajesh Kumar [12] have presented to create a method for explaining a portion of Indian sign language. The task was completed by employing features collected using DWT and Elliptical Fourier descriptors using 10 separate signer features for 80 signs with a recognition rate of 96% to train a fuzzy inference system.

Anuja V. Nair and Bindu V. [23] have explained the recent development of numerous approaches in the fields of image processing and artificial intelligence. Indian researchers have recently begun working on ISL to create systems that automatically recognize Indian sign language. Artificial Neural Networks (ANN), Support Vector Machines (SVM), Hidden Markov Models (HMM), and other key classification techniques are working for recognition.

J. L. Raheja et.al.,[6] Indian sign identification is based on real-time dynamic hand motion recognition methods. For pre-processing, the recorded video was transformed to HSV color space, and then segmentation based on skin pixels was carried out. From the picture frames, Hu-Moments and motion trajectories were recovered, and a Support Vector Machine was used to classify the gestures. And 97.5% accuracy rate.

Brandon Gracia and Rigoberto Alarcon Viesca [14] have discussed a web application based on a CNN classifier that was developed and trained to translate between English and American Sign Language. For letters a-e, the author can create a robust model, and for letters a-k, a modest one (excluding j). The validation accuracy authors found during training was not immediately repeatable upon testing on the web application due to the lack of diversity in our datasets.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 09, September 2022

Amrutha K and Prabu P [16] have presented the method for sign language recognition (SLR) through various processes. A huge dataset and the best method must be used to train a system that can read and understand a sign. An isolated recognition model is created as the foundation of the SLR system. The approach is based on the detection and recognition of individual hand gestures using vision. The model used KNN for classification and a convex hull to extract features.

Nimisha K P and Agnes Jacob [15] have proposed the recognition of sign language can be done in a variety of ways depending on visual and data glove. VBA's feature extraction working a variety of methods, including YOLO, CNN, PCA, etc. The pre-trained model is the newest and fastest of these methods. SVM, ANN, and CNN classifiers are used in the classification stage. All of these techniques provide very high precision.

III. DATASET

Data collection is part of this work and is an important step to maintain the integrity of the research. Before capturing this dataset, a thorough study of sign language has been carried out and then the dataset has been collected for this research work. The hand gesture recognition for sign language that was used in this paper's available datasets was obtained publicly from datasets. The literature is clear that the authors produced a dataset for Sign Language that included 15000 images and 26 classifications of image size are 256×256.

IV. PROPOSED METHODOLOGY

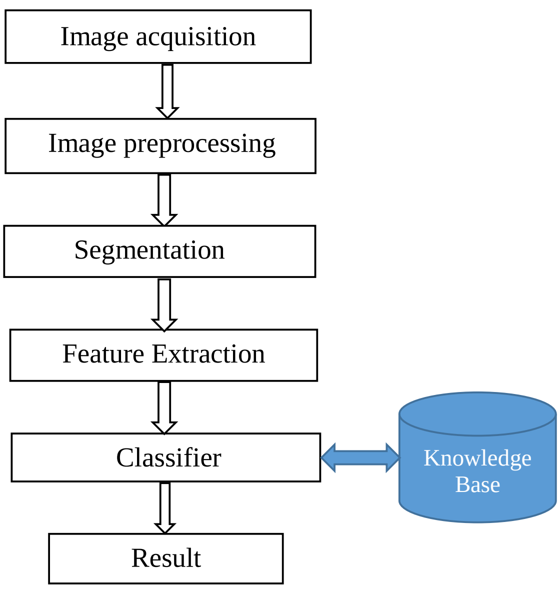


Figure 2: Proposed System Architecture

A. Image acquisition

The suggested system's first stage is data collection to record the hand movements, many research studies have used sensors or cameras. Use the web camera to capture the hand motions for our system the backgrounds are identified and removed from the images through some processing steps using the color extraction method. Every time an instruction is delivered, a webcam is used to take images of the same background to achieve higher consistency. The obtained images are kept in the PNG format it should be noted that there is no quality loss when a PNG image is opened, closed, and then saved once more. PNG is also effective at handling detailed images with great contrast. The images from the webcam will be recorded in RGB color space to the size 256×256. Figure 3 shows an image captured by a web camera.

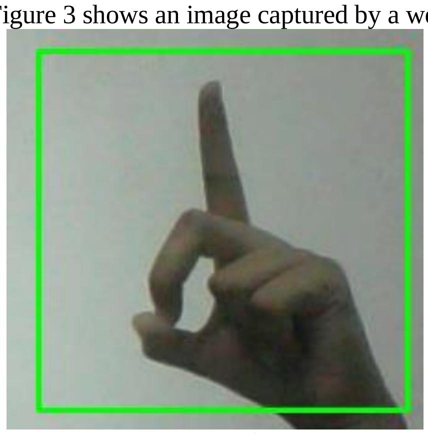


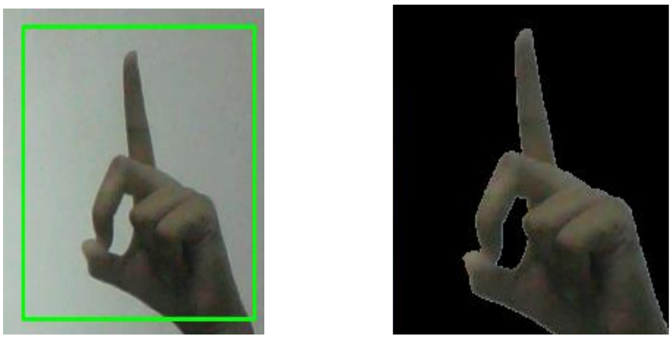
Figure 3: Image capture from web camera

B. Image Preprocessing

Since the collected images are in RGB color space, it is more challenging to separate the hand gestures simply based on skin color. As a result, converting the images to HSV color space is a system that divides an image's colors into three distinct components, namely hue, saturation, and value. . By separating brightness from color space, HSV is a useful tool for enhancing image stability. The background is turned black once a track-bar with H and S values from 0 to 179, 0-255, and 0 to 255 detects the hand gesture. Figure 4 shows the image after the background is set to black using HSV. To extract the relevant information from the current webcam clip, some image preprocessing is required. The background must first be manually divided using a thresholding process. According to the HSV color of the object that was detected, a certain range needs to be specifically stated. The first row of images shows the RGB images that were acquired, while the second row shows the equivalent grayscale images that have been noise- and background-reduced. Gaussian blur is then applied to the image. We can use the Ad Boost Hand Detector to extract the essential image for training by using the Gaussian Blur filter on hands that include skin color The method for hand detection includes background reduction and threshold-based color detection.

The image is blurred and noise is reduced using a linear filter called a Gaussian filter. Edges will be blurred and contrast will be decreased by it simply as the input signals are altered by the Gaussian filter by convolution with a Gaussian function.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 09, September 2022



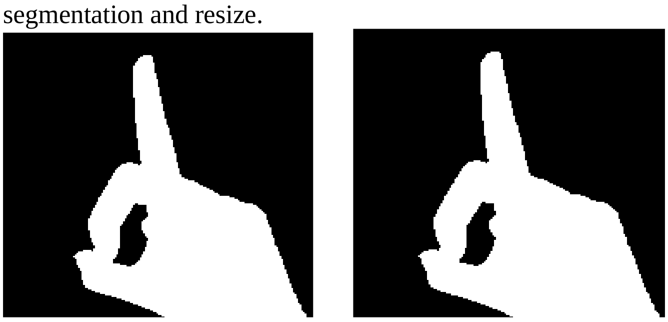
(a) (b)

Figure 4: a) Input image b) Image after the background is set to black using HSV.

C. Segmentation

The initial image is then converted to grayscale. The region of the skin gesture will lose color as a result of this process, but it will also make our system more resistant to variations in illumination. While the other pixels in the converted image remain unmodified and are consequently black, only the non-black pixels are binary. ROI Its main objective is to identify hand motions and extract the most intriguing details and illustrates how the hand region is identified utilizing skin-detection elements from the source image using some predetermined masks and filters the hand gesture is divided into two parts, with the hand gesture serving as our example. First, all related portions of the image are removed, and then only the hand gesture is left. The frame is scaled to 128 by 128 pixels in size.

There is a potential that the threshold-based segmentation will affect the digital image even when done in ideal lighting conditions. The identified objects could be too small or large, blurring the edge of the image. Edge-based segmentation can be utilized to extract the characteristics to prevent this bias Figure 5 shows the a) Image after binarized b) Image after segmentation and resize.



(a) (b)

Figure 5: a)Image after binarization b)Image after segmentation using threshold

D. Feature Extraction

The selection and extraction of key features from an image are one of the most critical steps in image processing. Images typically require a large amount of storage space when they are taken and kept as a dataset since they are made up of so much data by automatically extracting the most crucial aspects from the data, feature extraction assists us in finding a solution to this issue it also helps in preserving the classifier's accuracy and reduces its complexity. In our scenario, the binary pixels of the images have been proven to be essential elements and able to obtain enough characteristics by downsizing the images to 128 pixels to accurately categorize the Sign Language gestures we have a total of 32 to predict hand gestures and extract characteristics from the frames, a CNN model is used.

E. Classifier

Using a 3 by 3 filter, scan the images and Apply a 2D CNN model with a tensor flow library, the proposed system's dot product between the frame's pixels. The weights of the filter's convolution layers are computed from the supplied image, this particular stage extracts key traits that are then passed on. After each convolution layer, the pooling layers are then applied the activation map of the preceding layer is decreased by one pooling layer. It combines all of the features that were discovered in the input images of original levels this expands the range of properties the network can represent and reduces the overfitting of the training set. . In our example, the activation function is a Rectified Linear Unit, and the input layer of the convolutional neural network comprises 32 feature maps with a 3 by 3 size. The maximum pool layer measures 2 by 2. The layer is flattened, and the dropout is set to 50%. The network's last layer is a ten-unit, fully connected output layer using the Softmax activation function.

1. *CNN Architecture*

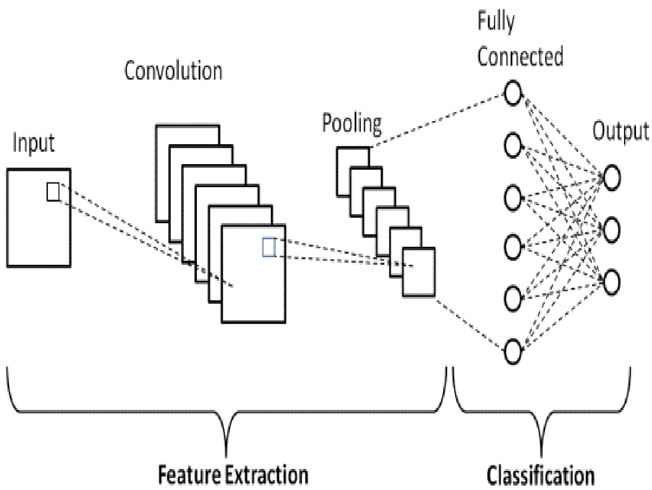


Figure 5: CNN Architecture

a. *Convolution neural network (CNN)*

This study aims to develop a network that can accurately translate a static sign language gesture to its written equivalent. Used Keras and CNN architecture with a variety of layers for data processing and training to get specified results. Each of the 16 filters in the convolutional layer has a 2 × 2 kernel. The spatial dimensions are then reduced to 32 × 32 by a 2 × 2 pooling. Convolutional layer filters are increased from 16 to 32, while Max Pooling filter sizes are increased from 5 × 5. Then, the CNN layers' filter count is raised to 64, but max-pooling remains at 5× 5. With Dropout, each node is randomly disconnected from the current layer and moved to the following layer.The model is now flattened or changed into a vector, and the dense layer is added after that. Together with rectified linear activation, the dense layer specifies the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 09, September 2022

completely linked layer. The SoftMax classifier was used by the author to complete the model and provide the predicted probability.

Convolution neural networks, sometimes known as CNNs, are a class of neural networks used frequently in deep learning for image processing tasks. It was created specifically to process photos. After applying a filter to some input arrays, CNN produces an output array. The filters help in the feature extraction process.

- Starts with an input image.
- Applies many different filters to obtain a feature map.
- Applies a RELU function to increase non-linearity.
- Applies pooling layer to each feature map.
- Flattening the pooled images into one long vector
- Dropout is also used to mitigate overfitting.
- The final fully connected layer provides the voting of the classes.

2. *Proposed System CNN*

In this study, a CNN-based model has been created especially for the recognition of gesture-based sign language. As a result of its clear structure, resource, and energy efficiency, and reduced calculation time, this model is helpful. The proposed model in this study is known as CNN and has 10 layers total, including 2 convolutional layers, 2 pooling layers, 2 dropout layers, 2 fully connected layers, and 2 softmax layers. A CNN small filter size of 3, 2, and 1 that is based on a large filter size is used in the weighted layer.

Beginning with the convolutional layer, which extracts features by swiping a filter window over an input image, the processing of input gesture images of size [128×128] begins. As features are extracted from an input image, these filters' weights are automatically updated, and learned 32 convolutional filters with the dimensions [3× 3 ×32] have been applied in this layer. As a result, the [128× 128× 32] dimension represents 32 high-level features that are extracted. to become close to the nonlinear decision limits using the max-pooling process, the size of the resulting features map is further reduced by a factor of 2. To create the spatiotemporal representation of the gestures, more sets of convolutional and max-pooling layers are built over this one similarly. There are a total of four convolutional layers used, each with a stride of one and an activation function. As they are arranged in the model, the kernel sizes for each convolutional layer are 3, 1, 3, and 3 with corresponding convolutional depths of 32, 64, 64, and 128. This model makes use of the small kernel size to understand the fine texture of the signs. By working with a filter size of 2 and a stride of 2, the max-pooling operation has been working to reduce the size of feature maps. After that, all previously extracted features for the categorization are linked using a set of the completely connected layer. Two fully connected layers require 512 and 84 hidden units respectively. This model had two dropout layers of 0.40 and 0.40 during training

TABLE I. CONFIGURATION OF PROPOSED CNN

Layer Type	No. of filter	Feature Map Size	Kernel Size	Stride Used
Input Image layer	-	128*128	-	
Convolution 1	32	128*128*32	3*3	1*1
Max-pooling 1	1	64*64*32	2*2	2*2
Convolution 2	32	128*128*32	3*3	1*1
Max-pooling 2	1	64*64*32	2*2	2*2
Dropout 1		0.40		
Dropout 2		0.40		

3. *Training and Testing*

In this study, various datasets of sign language were used to train and test the proposed CNN model and alternative CNN architecture (given in Section 3.1). Before being fed to the feature learning model, the images from each class are split into two sets: 80% for training and 20% for testing. In each training phase, the data is fed into the network in batches of 26 samples, and there have been 5 epochs of training. The Adam optimizer is used to train this hand gesture recognition model, which takes into account its capacity to adjust to learning rates based on a changing frame of gradient updates.

V. *RESULT*

Open-CV was developed with a process on real-time applications to maximize computing efficiency. To facilitate the use of machine perception in commercial products and to provide a common architecture for computer vision applications. As previously said, the dataset images and the input test image are compared to determine whether they are similar to the used 15000 images for this paper, with 600 images for each group or class. This 80% of images are used to train the model. In the test set, 20% of images are included. On the above information, all outcomes are predicated. By providing the model with more images during training, the model's accuracy can be improved. The letter L from the sign is displayed in Figure 7.

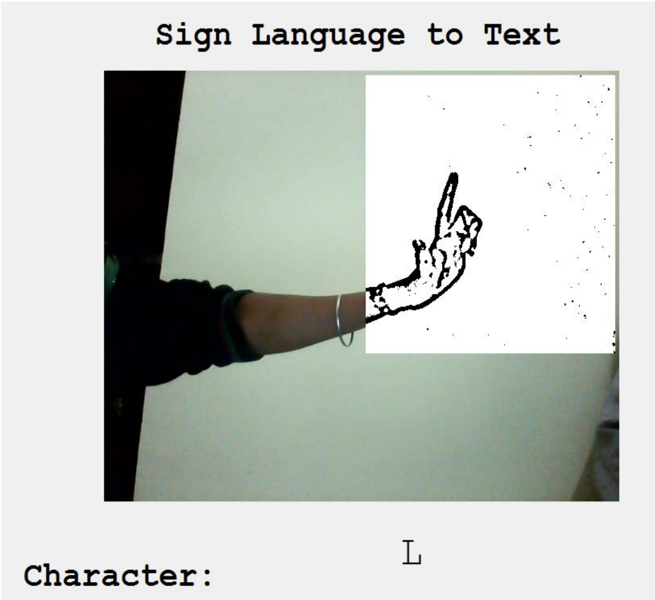


Figure 6: Letter L from the sign

1. Comparison of the classification result

This method's effectiveness has been evaluated in this work in comparison to other methods that have been developed for the same classification problem of hand gesture recognition providing the broad details of this comparison. Since accuracy is the only performance factor that is consistently employed across all state-of-the-art methodologies, comparisons have only been made based on the accuracy achieved. Only, as it is the only widely used performance depends on all the state-of-art approaches. From this table, it has been found that Shruti Chavan et al. [26], Mehreen Hurroo and Mohammad ElhamWalizad [27], and Darden Tasmere [28] have worked with limited numbers of signs, and their achieved accuracy is 87.5%, 90%, and 94% respectively. It is evident from these findings that the CNN model surpasses all the other methods as it achieves the highest accuracy of 98.0%, Sign Language fingerspelling. Table 2 gives the comparison of this proposed work with the existing work.

TABLE II. COMPARISON OF CNN WITH PUBLISHED WORK FOR SIGN LANGUAGE

Method	Accuracy (%)
CNN(Shruti Chavan et al.[26],2021)	87.5%
CNN (Mehreen Hurroo and Mohammad ElhamWalizad [27], 2020)	90%
CNN(Dardina Tasmere et al.[28],2021)	94%
Proposed	98.0%

VI. CONCLUSION

It has always been difficult to communicate with someone who is deaf-mute and our work aims to lower the obstacle standing between them. Authors have attempted by adding to the topic of understanding Sign Language In this study, the author created a CNN-based system for recognizing human hand gestures. The key aspect of our technique is that do not need to create a model for each action based on the curves and fingertips of the hand a CNN classifier that can identify sign language motions was built. Results from the suggested system for transitive gestures have been good.

The author has also verified our results for the similar-looking gestures that were more likely to be misclassified in this work a functional real-time vision-based sign language recognition system for deaf and dumb people has been developed. It achieved a final accuracy of 98.0% on our dataset and was able to improve our prediction after implementing two layers of algorithms this method can recognize practically all symbols as long as they are displayed correctly, there is no background noise, and the lighting is sufficient as well as examined with the supplemented data, the model is discovered to be rotation- and scaling-invariant. The comparative investigation clearly

REFERENCES

[1] Joudaki, S., Mohamad, D. bin, Saba, T., Rehman, A., Al-Rodhaan, M., & Al-Dhelaan, A. (2014, November 3). "Vision-based sign language classification: A directional review". IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India). Taylor and Francis Ltd. https://doi.org/10.1080/02564602.2014.961576.

[2] Manisha U. Kakde, Mahender G. Nakrani, & Amit M. Rawate. (2016). "A Review Paper on Sign Language Recognition System for Deaf and Dumb People using Image Processing". International Journal of Engineering Research and, V5 (03). https://doi.org/10.17577/ijertv5is031036.

[3] Pansare, J. R., & Ingle, M. (2016). "Vision-based approach for American Sign Language recognition using Edge Orientation Histogram". In 2016 International Conference on Image, Vision, and Computing, ICIVC 2016 (pp. 86–90). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICIVC.2016.7571278.

[4] Sharma, S., & Singh, S. (2020). "Vision-based sign language recognition system: A Comprehensive Review". In Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020 (pp. 140–144). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICICT48043.2020.9112409.

[5] Ebrahim Al-Ahdal, M., & Nooritawati, M. T. (2012). "Review in sign language recognition systems". In 2012 IEEE Symposium on Computers and Informatics, ISCI 2012 (pp. 52–57). https://doi.org/10.1109/ISCI.2012.6222666.

[6] Pandey, P., & Jain, V. (2015). "Hand Gesture Recognition for Sign Language Recognition: A Review". International Journal of Science, Engineering and Technology Research (IJSETR) (Vol. 4).

[7] Paulraj, M. P., Yaacob, S., Desa, H., Hema, C. R., & Wan Ab Majid, W. M. R. (2008). "Extraction of head and hand gesture features for recognition of sign language". In 2008 International Conference on Electronic Design, ICED 2008. https://doi.org/10.1109/ICED.2008.4786633.

[8] Sharma, S., & Singh, S. (2021). "Vision-based hand gesture recognition using deep learning for the interpretation of sign language". Expert Systems with Applications, 182. https://doi.org/10.1016/j.eswa.2021.115657.

[9] Jebali, M., Dakhli, A., & Jemni, M. (2021). "Vision-based continuous sign language recognition using multimodal sensor fusion". Evolving Systems, 12(4), 1031–1044. https://doi.org/10.1007/s12530-020-09365-y.

[10] Kishore, P. V. V., & Rajesh Kumar, P. (2012). "A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic". International Journal of Engineering and Technology, 4(5), 537–542. https://doi.org/10.7763/ijet.2012.v4.427.

[11] V.Nair, A., & V, B. (2013). "A Review on Indian Sign Language Recognition". International Journal of Computer Applications, 73(22), 33–38 https://doi.org/10.5120/13037-0260.

[12] Patil, R., Patil, V., Bahuguna, A., & Datkhile, G. (2021). "Indian Sign Language Recognition using Convolutional Neural Network". ITM Web of Conferences, 40, 03004. https://doi.org/10.1051/itmconf/20214003004.

[13] Garcia, B., & Viesca, S. A. (2017). "Real-time American Sign Language Recognition with Convolutional Neural Networks". 2017 IEEE International Autumn Meeting on Power, Electronics and Computing, ROPEC 2017, 2018-Janua, 1–6.

[14] Amrutha, K., & Prabu, P. (2021). "ML-based sign language recognition system". In 2021 International Conference on Innovative Trends in Information Technology, ICITIIT 2021. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICITIIT51526.2021.9399594.

[15] Nimisha, K. P., & Jacob, A. (2020). "A Brief Review of the Recent Trends in Sign Language Recognition". In Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing, ICCSP 2020 (pp. 186–190). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCSP48568.2020.9182351.

[16] Cheok, M. J., Omar, Z., & Jaward, M. H. (2019). "A review of hand gesture and sign language recognition techniques". International Journal of Machine Learning and Cybernetics, 10(1), 131–153. https://doi.org/10.1007/s13042-017-0705-5.

[17] Tolentino, L. K. S., Serfa Juan, R. O., Thio-ac, A. C., Pamahoy, M. A. B., Forteza, J. R. R., & Garcia, X. J. O. (2019). "Static sign language recognition using deep learning". International Journal of

Machine Learning and Computing, 9(6), 821–827. https://doi.org/10.18178/ijmlc.2019.9.6.879.

[18] Sudeep, K. S., & Pal, K. K. (2017). “Preprocessing for image classification by convolutional neural networks”. In 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings (pp. 1778–1781). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/RTEICT.2016.7808140.

[19] Rao, G. A., & Kishore, P. V. V. (2016). “Sign language recognition system simulated for video captured with smartphone front camera”. International Journal of Electrical and Computer Engineering, 6(5), 2176–2187. https://doi.org/10.11591/ijece.v6i5.11384.

[20] Nandy, A., Prasad, J. S., Mondal, S., Chakraborty, P., & Nandi, G. C. (2010). “Recognition of isolated Indian sign language gesture in real time”. In Communications in Computer and Information Science (Vol. 70, pp. 102–107). Springer Verlag. https://doi.org/10.1007/978-3-642-12214-9_18.

[21] Kakkoth, S. S., & Gharge, S. (2018). “Visual Descriptors Based Real Time Hand Gesture Recognition”. In 2018 International Conference on Advances in Communication and Computing Technology, ICACCT 2018 (pp. 361–367). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICACCT.2018.8529663.

[22] Aloysius, N., & Geetha, M. (2020). “Understanding vision-based continuous sign language recognition”. Multimedia Tools and Applications, 79(31–32), 22177–22209. https://doi.org/10.1007/s11042-020-08961-z.

[23] Starner, T., & Pentland, A. (1995). “Real-time American Sign Language recognition from video using Hidden Markov models”. In Proceedings of the IEEE International Conference on Computer Vision (pp. 265–270). IEEE. https://doi.org/10.1109/iscv.1995.477012.

[24] Wadhawan, A., & Kumar, P. (2021). “Sign Language Recognition Systems: A Decade Systematic Literature Review”. Archives of Computational Methods in Engineering, 28(3), 785–813. https://doi.org/10.1007/s11831-019-09384-2.

[25] Cheok, M. J., Omar, Z., & Jaward, M. H. (2019). “A review of hand gesture and sign language recognition techniques". International Journal of Machine Learning and Cybernetics, 10(1), 131–153. https://doi.org/10.1007/s13042-017-0705-5.

[26] Chavan, S., Yu, X., & Saniie, J. (2021). Convolutional Neural Network Hand Gesture Recognition for American Sign Language. In IEEE International Conference on Electro Information Technology (Vol. 2021-May, pp. 188–192). IEEE Computer Society. https://doi.org/10.1109/EIT51626.2021.9491897.

[27] Tasmere, D., Ahmed, B., & Das, S. R. (2021). Real-Time Hand Gesture Recognition in Depth Image using CNN. International Journal of Computer Applications, 174(16), 28–32. https://doi.org/10.5120/ijca2021921040.