DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004 *Int. J. Exp. Res. Rev.*, Special Vol. 34: 26-35 (2023)

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb

Prasanthi Yavanamandha[1], Bodduru Keerthana[2*], Penmetsa Jahnavi[3], Koduganti Venkata Rao[4] and Chatikam Raj Kumar[5]

[1]Department of CSE-AIML & IoT, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telangana, India; [2]Department of Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering (A), Visakhapatnam, Andhra Pradesh, India; [3]Department of Computer Science and Engineering, SRKR Engineering College Bhimavaram, Andhra Pradesh, India; [4]Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Hyderabad, India; [5]Department of Computer Science and Engineering, Gayatri College of Science and Management, Srikakulam, Andhra Pradesh, India

E-mail/Orcid Id:

PY, prasanthi_y@vnrvjiet.in; *BK,* bkeerthana@gvpce.ac.in, https://orcid.org/0009-0000-6820-514X; *PJ,* jahnavi.penmatsa@gmail.com, https://orcid.org/0000-0002-5245-4173; *KVR,* vrkoduganti@gmail.com, https://orcid.org/0000-0002-1633-7236; *CRK,* chatikams@gmail.com

Article History:
Received: 14th Jul., 2023
Accepted: 07th Oct., 2023
Published: 30th Oct., 2023

Keywords:
Sign Language, Computer Vision, Machine Learning, Convolutional Neural Network (CNN)

How to cite this Article:
Prasanthi Yavanamandha, Bodduru Keerthana, ….. and Chatikam Raj Kumar (2023). Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb. *International Journal of Experimental Research and Review*, *34*(Spl.), 26-35.
DOI : https://doi.org/10.52756/ijerr.2023.v34spl.004

Abstract: A deep learning model specifically designed to recognize signs in sign language is the foundation of the Sign Language Recognition system. Sign Language is a visual language used by the deaf and hard of hearing community to communicate with one another and the general public. Sign language is a kind of nonverbal communication based on the use of hand gestures. The ability to communicate socially and emotionally is greatly aided when the speech and hearing challenged have access to sign language. The model developed in this paper captures the images through live web cam and displays the sign language meaning on the screen as text output. The model is trained and built by deep learning framework using Convolution Neural Networks (CNN) in this work. The model is trained with images of hand gestures captured through webcam using Computer Vision and then after successful training, the system performs recognition process through matching parameters for a given input gesture and finally displays the sign language meaning of the gesture as text output on the screen.

Introduction

Sign language recognition involves creating an assistive system that can automatically transform an input sign into the voice or text that corresponds to it (Mittal et al., 2019). Therefore, the sign language recognition system is effective in eliminating the communication gap between communities of hearing and non-hearing individuals, and it opens a new avenue for applications that are based on human-computer interaction (Kanisha et al., 2022; Rakesh et al., 2021; Itkarkar et al., 2021).

Deaf and dumb people specifically use sign language. They communicate words and sentences using different signs (Ali et al., 2022; Rajam and Balakrishnan, 2011). These signs are shown as hand gestures made by particular movements of the hands, making a specific shape out of them. But, communication through sign language becomes hard as it is not something that is common to all. Also, learning such a language is not quite easy (Wadhawan and Kumar, 2019). In order to help such situations where sign language becomes a barrier to communication, a sign language detecting model deals with the task of instantly capturing the signs and recognising them (Goel et al., 2023). The model bridges the linguistic and emotional communication gap between deaf and dumb people and others who do not know sign language. It acts like a translator of sign language.

*Corresponding Author: bkeerthana@gvpce.ac.in

In this work, we proposed a model that is built using machine learning frameworks. Convolution Neural Network (CNN) is used in designing the model. The Computer Vision module gathers data from a webcam and trains a model to recognize signs in real-time. This system, after training recognises and displays the sign meaning as text. The model helps in recognising the signs with a good level of accuracy.

The "Sign Language Recognition system" aims to help those who are hard of hearing or have trouble speaking communicate with those who do not have these difficulties by taking a picture of the sign using a webcam and then displaying the text representation of the meaning.

Using this work, people with either speech or hearing impairments can communicate with regular people without difficulty or confusion. Therefore, this work improves the social lives of those with speech or hearing impairments. Additionally, this work made those people's lives much simpler. The time they can save by using this work is communicating verbally rather than in writing. Not only save time for them but also for the other people who want to communicate with them need not learn the language. This is possible through a vision-based understanding of the sign language. This work is extremely reliable and useful in various circumstances because it may be utilised anywhere.

Existing System

Data gloves approach

Instrumented gloves are another name for data gloves, as shown in Figure 1. Data gloves are required for hand recognition and tracking. These mitts have built-in sensors that track the wearer's hand orientation and movements. To detect hand posture, this technique uses electrical impulses generated by transforming mechanical or optical sensors mounted to a glove. Data gloves make it simple to record the position, orientation, and shape of one's hand, palm, and fingers. The normal amount of engagement with the computer is diminished.

Limitations

Sensor Based Method

Uses gloves embedded with sensors as primary tools. It is expensive because of the hardware.

The user-computer interface will be less natural because of the necessity of wearing the glove and a cumbersome gadget with many cords attached to the computer.

Figure 1. Data Glove

Proposed Work

Vision-Based Approach

To overcome the existing system drawbacks, the proposed model has been developed. It is vision-based where the webcam comes to play the role of capturing input signs to display predicted meaning as output. This model is built using the CNN algorithm for better classification and recognition of signs. It provides natural user-computer interaction.

Advantages of proposed model

It uses cameras as primary tools.

It removes the need for sensors.

This model reduces the building cost of the system.

The approach is robust for sign language recognition by hand gestures.

Related Work

One such visual-spatial language is Indian Sign Language, named after the country of origin. Indian Sign Language has a unique approach to developing its grammar, phonology, and morphology. Naturalness also applies to the Indian Sign Language. It generates semantic information that conveys words and emotions through the use of arm motions, hand gestures, facial expressions, and body/head movements (Papastratis, 2021).

According to Suneetha et al. (2023), the distinctive feature of the method that has been suggested is that it identifies hand landmarks by making use of Google's Media Pipe, which is both more efficient and more accurate than conventional methods that are based on geometry, form, and edge data. The LSTM model has been shown to be quite successful when it comes to the modeling of sequence data as well as the recognition of gestures.

Indian Sign Language (ISL) movement detection and recognition from grayscale pictures was proposed by Nandy et al. (2010). Their method takes a video source

DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004

with signing gestures and converts it into grey-scaled frames, then uses a directed histogram to extract features from the frames. The signing movements can also be retrieved from the video source. In conclusion, the clustering technique categorises the signals into one of the pre-defined groups according to their characteristics. The study's authors determined that the 36-bin histogram method was superior to the 18-bin histogram method after achieving a 100% sign identification rate.

Using a Convolutional Neural Network, an attention mechanism for recovering spatial data, and bio-inspired deep learning with Long Short-Term Memory (BI-LSTM), Abdul et al. (2021) developed a system for classifying Arabic sign language. Temporal features were extracted using the BI-LSTM. This model was tested under various settings, from varying lighting to costume changes to subject separation from the camera. The resulting model required less time to process than the alternatives because it had fewer deep learning layers and parameters to analyze.

According to the research conducted by Aggarwal et al. (2023), they provided a synopsis of the various pieces of study that have been carried out on the subject of hand sign language recognition. This investigation also contrasted other pieces of previous research and enumerated the benefits and drawbacks associated with doing so.

In order to identify sign language and generate text from the video stream in real-time, Mekala et al. (2011) proposed a neural network architecture. Photos are pre-processed, and feature extraction is performed based on the position and movement of the user's hands, among other things. Each finger, palm, and other hand structure has its own point of interest (POI) (Mekala, 2011). Indicator forecasting using the authors' CNN-layer neural network design was aided by the extraction of 55 features using this method. They claim a 100% identification rate and 48% noise immunity using the full English alphabet as training data.

Using a sensor glove for signing, processing the signs, and presenting the output in a comprehensible sentence, Agarwal et al. attempt to close the gap between persons with speech disabilities and those with normal speech skills. This objective was met by creating a fair playing field for people with and without speech impairments. In the study (Agarwal et al., 2015), subjects acted while wearing the sensor gloves. After the gesture was compared to the database and found to be a match, the data was sent on to be parsed so that a phrase could be constructed using the gesture's components. When first released, the software was only about a third accurate. Version 2 includes a tense-specific keyword that improves accuracy to 100% when using the basic and continuous tenses.

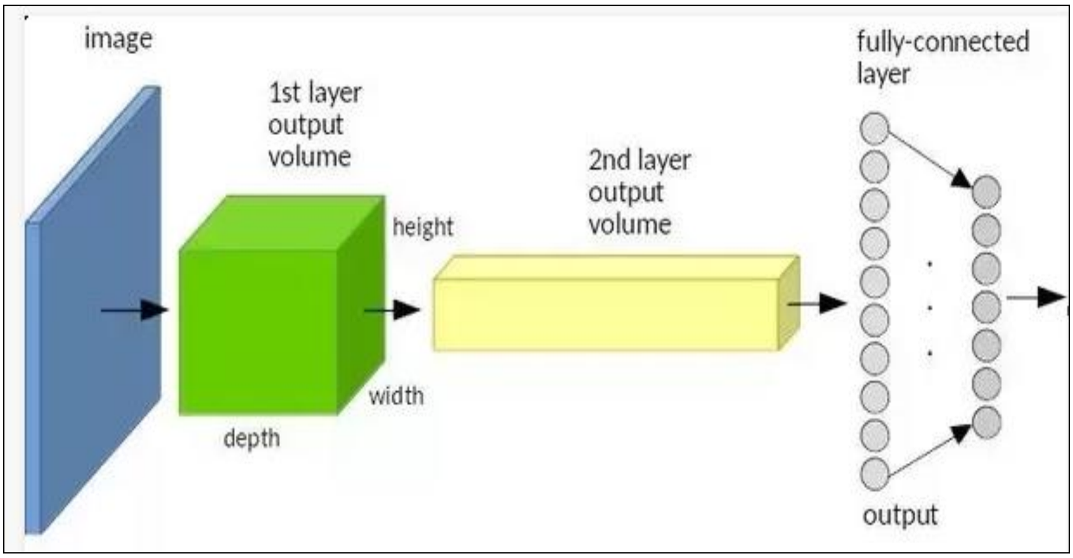

Figure 2. 3-D Arrangement of Neurons in CNN layers

The development of the numerals used in Bhutanese sign language involved the usage of a CNN (Wangchuk et al., 2021). This model used around 20,000 sign images to recognize ten static digits in Bhutanese sign language. Each of these digits was received freely from a separate participant. During the investigation, a comparison was made between several sign languages and the model that CNN suggested. Based on the comparison findings, their proposed model achieved an accuracy of 99.94% during training. The testing accuracy was 97.62%.

Using a transformer network, De Coster and colleagues could identify non-manual elements of sign language in film, such as the angle of the mouth and eyebrows (De Coster et al., 2020). A multimodal transformer, video transformer, and posture transformer network were designed to detect signs in several neural networks.

DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004

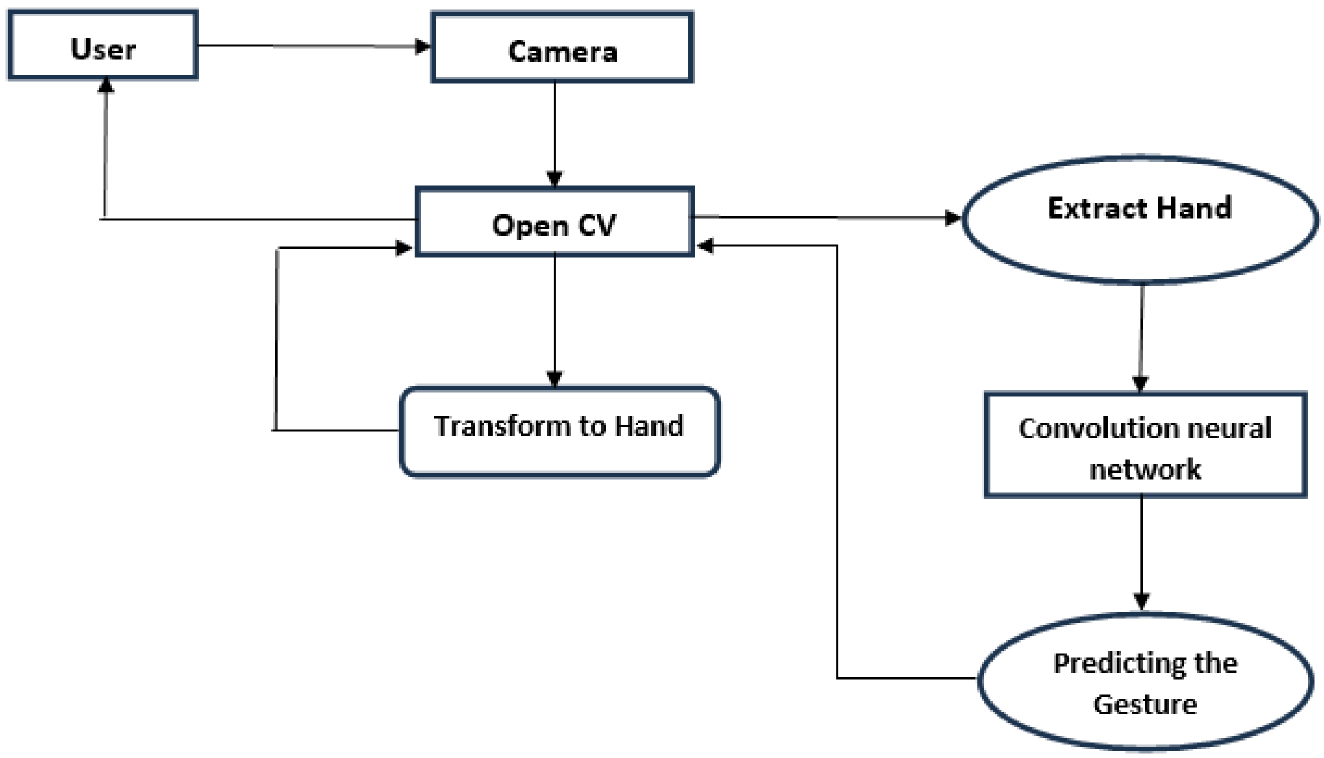

Figure 3. Flow Diagram

In-Depth Learning In recent years, CNN has been increasingly popular in data science studies, such as using vision-based hand gesture identification for sign language interpretation through deep learning (Sharma and Anand, 2021). One experiment utilized a Convolutional Neural Network (CNN) trained with Deep Learning and tailored for sign language detection.

Methodology

Deep learning, often called deep structured learning, uses artificial neural networks and representational learning. Unsupervised and semi-supervised learning is feasible. Deep learning is becoming popular (Saha and Yadav, 2023; Rao et al., 2022). Deep learning models use neural networks. A neural network processes inputs using hidden layer weights changed during training. The model predicts. Changing weights to find patterns improves forecasts. Deep learning lets a computer model classify images, text, and voice directly. Deep learning models can be supervised, semi-supervised, or unsupervised, sometimes outperforming humans. Train models with lots of labeled data. Computer vision, speech recognition, NLP, and audio recognition have employed deep learning architectures like deep neural networks.

Convolution Layer

Connecting neurons to all neurons in the preceding volume is impractical for high-dimensional inputs like pictures (Rao et al., 2023a). We'll connect each neuron to an input volume section. Convolutional layers form CNNs (Rao et al., 2023b), the layer's parameters are a set of learnable kernels with a small receptive field but full input volume (Liu et al., 2023; Dhulipala et al., 2022). In the forward pass, each filter convolved across the width and height of the input volume computes the dot product between its entries and the input and creates a 2-dimensional activation map shown in Figure 2. Thus, the network learns filters that activate when it detects a specific feature at a specific spatial position in the input (Krishnan et al., 2023; Reddy et al., 2023). This network topology ignores the spatial structure of high-dimensional inputs like pictures, making connecting neurons to all neurons at the previous level hard. Convolutional networks use spatially local correlation by connecting each neuron to a small section of the input volume (Kothadiya et al., 2022). Neurons' receptive fields determine this connection (Adaloglou et al., 2021; Likhar et al., 2020). Local in width and height, the connections always extend along the input volume's depth. This architecture trains filters to respond best to spatially local input patterns.

DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004

Control Flow Diagram

Control flow diagrams explain processes shown in Figure 3. It shows us where control begins and ends and may branch off in specific instances. You're writing machine-starting software. What if the engine floods or a spark plug breaks? Control then redirects software flow. Diagram these branches. The flow diagram helps stakeholders and systems professionals understand it (Breland et al., 2021). Laypeople can understand the notion even if they don't understand particular symbols.

Figure 4. Reference Dataset

Experiments and Results

Dataset

To address a wide range of Artificial Intelligence problems, such as picture or video categorization, datasets often consist of photos, texts, audio, videos, numerical data points, etc.

The foundation of any practical AI application is a high-quality dataset. Therefore, it's important to know where to look. While ideal datasets would be simple, clean, and well-organized, real-world datasets are much more complicated, messy, and poorly organised. Quantity, quality, and relevance of the dataset are all crucial to the the success of any Machine Learning or Deep Learning model. Striking a middle ground is a challenging endeavor.

In this paper, we have created our own data set using a computer vision module in Python. We have accessed the webcam and collected images of specific pixels (300) for various signs. The data is collected with reference to the alphabet of American Sign Language and a few other general signs.

DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004

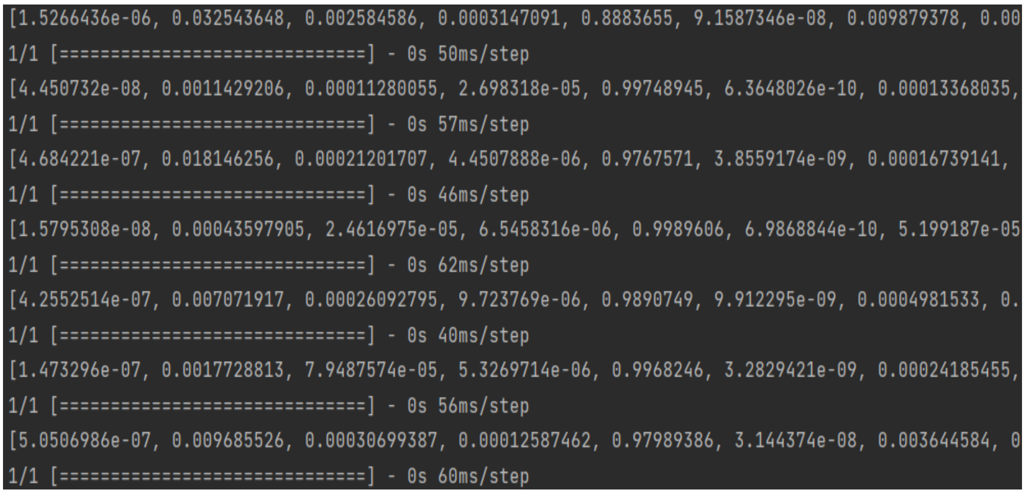

Figure 5. Accuracy Scores

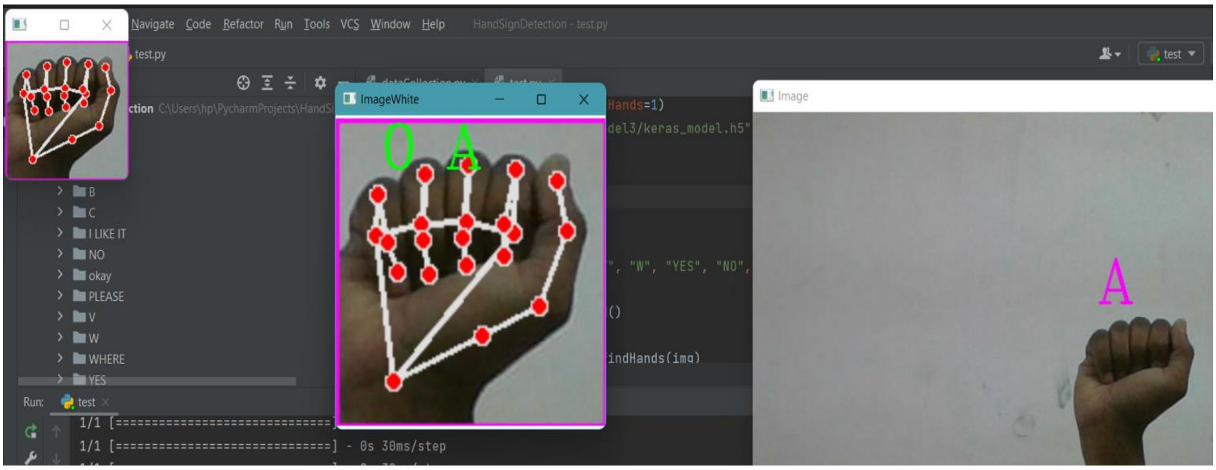

Figure 6. 'A' Sign

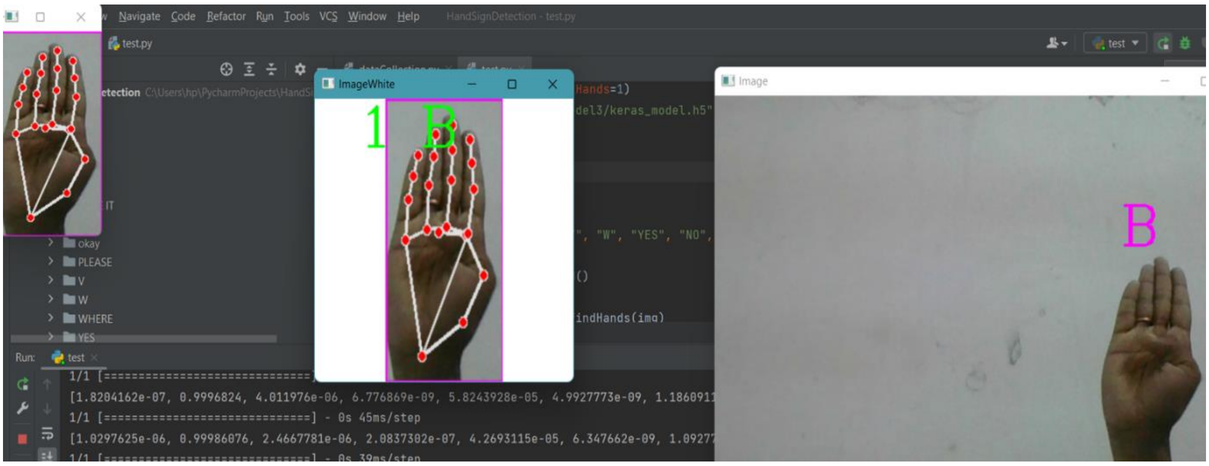

Figure 7. 'B' Sign

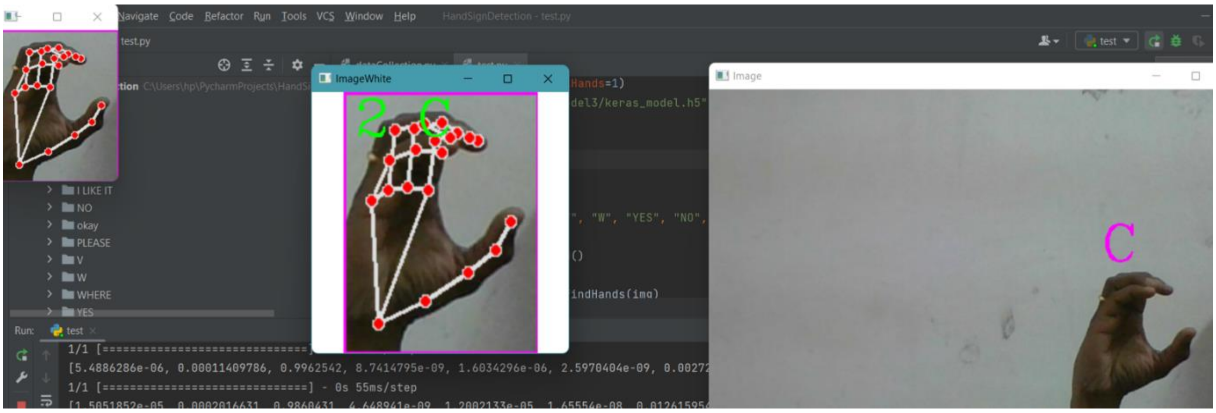

Figure 8. 'C' Sign

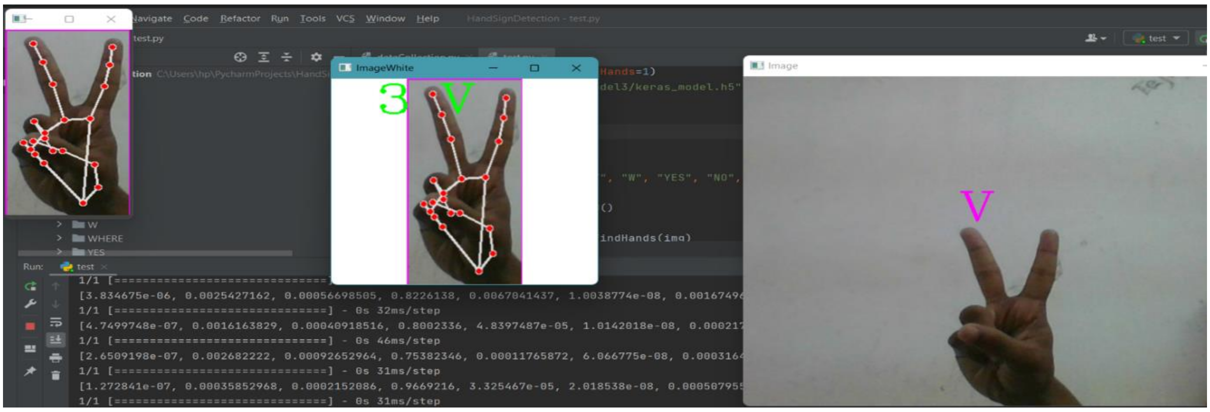

Figure 9. 'V' Sign

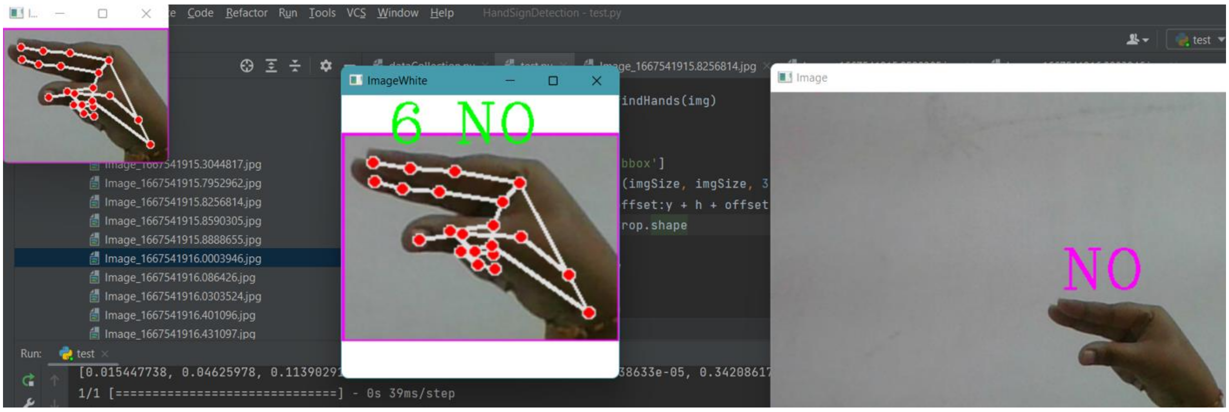

Figure 10. 'No' Sign

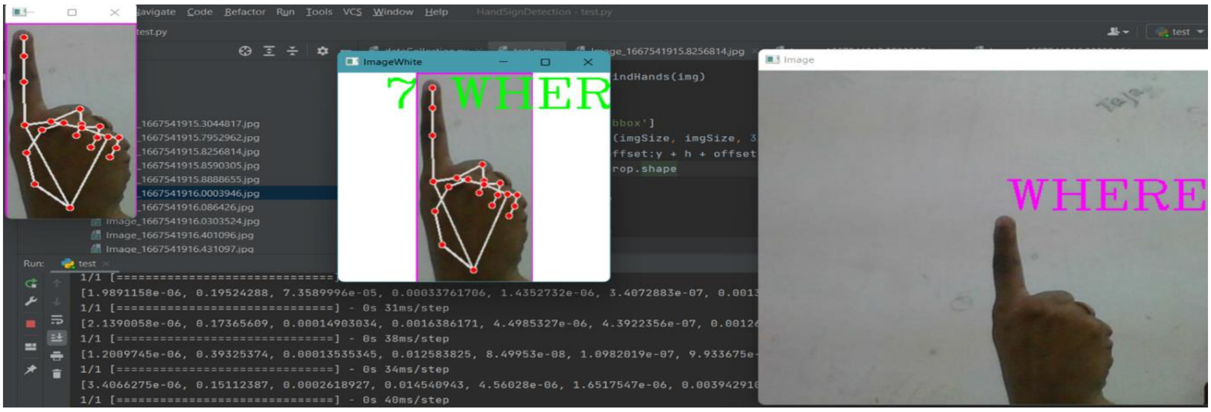

Figure 11. 'Where' Sign

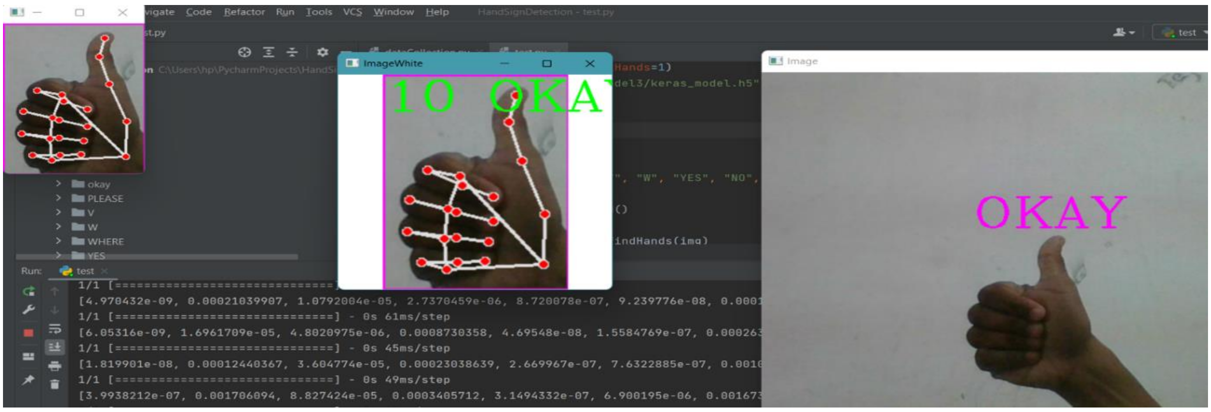

Figure 12. 'Okay' Sign

DOI: https://doi.org/10.52756/ijerr.2023.v34spl.004

Results

During this study, the authors considered different hand gestures for recognition. Different signs as 'A', 'B', 'C' etc of all alphabet and also signs like 'ok', 'where', 'are' etc. for recognition for communication with the Deaf and Dumb. Figure 5 shows an accuracy of 97.98% when compared with existing works, our proposed model shows improvement.

Conclusion

In this work, we train a model to recognize various signs and then use that information to predict new sets of signs. We conclude that our algorithm can classify diverse hand gestures with sufficient accuracy after running the model under various test conditions. The solution is meant to help those in need and maintain its societal significance. The system's simplicity and ease of use ensure that it will be widely adopted. The software reduces or eliminates the need for costly hardware or software. Therefore, the model can be easily expanded to a massive size by increasing the amount of the dataset. Some of the constraints on the model reduce the detection accuracy, such as low light intensity and an uncontrolled backdrop.

The proposed sign language recognition system used to recognize sign language letters can be further extended and trained to enhance model's ability to recognize gestures and facial expressions. The scope of different sign languages can be increased. More training data can be added to find the best with more accuracy. Additionally, training the neural network model to well organized identify symbols well requires two hands. This work can be expanded to convert symbols into speech.

Conflict of interest

None

References

Abdul W., Alsulaiman, M., Amin, S.U., Faisal, M., Muhammad, G., Albogamy, F.R., & Ghaleb, H. (2021). Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. *Comput. Electr. Eng.*, *95*, 107395. https://doi: 10.1016/j.compeleceng.2021.107395.

Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G.T., Zacharopoulou, V., Xydopoulos, G.J., Atzakas, K., Papazachariou, D., & Daras, P. (2021). A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition. *IEEE Transactions on Multimedia*, *24*, 1750-1762. https://doi.org /10.1109/TMM.2021.3070438

Agarwal, S.R., Agrawal, S.B., & Latif, A.M. (2015). Article: Sentence Formation in NLP Engine on the Basis of Indian Sign Language using Hand Gestures. *Int. J. Comput. Appl.*, *116*, 18–22. https://doi.org /10.5120/20428-2757.

Ali, M., Abbas, A.H., & Shahadi, H.I. (2022). Real-time sign language recognition system. *International Journal of Health Sciences*, *6*(S4), 10384-10407. https://doi.org/10.46501/IJMTST0801006.

Breland, D.S., Dayal, A., Jha, A., Yalavarthy, P.K., Pandey, O.J., & Cenkeramaddi, L.R. (2021). Robust Hand Gestures Recognition Using a Deep CNN and Thermal Images. *IEEE Sens. J.*, 21, 26602–26614. https://doi: 10.1109/JSEN.2021.3119977.

Aggarwal, D., Ahirwar, S., Srivastava, S., Verma, S., & Goel, Y. (2023). Sign Language Prediction using Machine Learning Techniques: A Review. *Second International Conference on Electronics and Renewable Systems (ICEARS)*, Tuticorin, India. pp. 1296-1300, https://doi.org/10.1109/ICEARS56392.2023.10084924.

De Coster, M., Herreweghe, M.V., & Dambre, J. (2020). Sign Language Recognition with Transformer Networks. *In Proceedings of the Conference on Language Resources and Evaluation* (LREC 2020), Marseille, France, pp. 6018–6024. https://aclanthology.org/2020.lrec-1.737

Dhulipala, S., Adedoyin, F.F., & Bruno, A. (2022). Sign and Human Action Detection Using Deep Learning. *J. Imaging.*, *8*(7), 192. https://doi: 10.3390/jimaging8070192.

Goel, A., Wasim, J., & Srivastava, P. (2023). A Noise reduction in the medical images using hybrid combination of filters with nature-inspired Black Widow Optimization Algorithm. *International Journal of Experimental Research and Review*, *30*, 433-441. https://doi.org/10.52756/ijerr.2023.v30.040.

Itkarkar R.R., Nandi, A.K.V., Mungurwadi, V.B. (2021). Indian Sign Language Recognition Using Combined Feature Extraction. In: Mukherjee, M., Mandal, J., Bhattacharyya, S., Huck, C., Biswas, S. (eds) Advances in Medical Physics and Healthcare Engineering. Lecture Notes in Bioengineering. Springer, Singapore. pp. 1-7. https://doi.org/10.1007/978-981-33-6915-3_1

Kanisha, B., Mahalakshmi, V., Baskar, M., Vijaya, K., Kalyanasundaram, P. (2022) Smart communication using tri-spectral sign recognition for hearing-

Impaired people. *J. Supercomput.*, *78*, 2651–2664. https://doi.org/10.1007/s11227-021-03968-1.

Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A.B., & Corchado, J.M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. Electronics, *11*, 1780. https://doi.org/10.3390/electronics11111780

Krishnan, G., Liya, B. S., Lakshmi, D.S.V., Sathyamoorthy, K., & Ganesan, S. (2023). Monkeypox Detection Using Hyper-Parameter Tuned Based Transferable CNN Model. *International Journal of Experimental Research and Review*, *33*, 18-29. https://doi.org/10.52756/ijerr.2023.v33spl.003.

Likhar, P., Bhagat, N.K., & Rathna, G.N. (2020). Deep Learning Methods for Indian Sign Language Recognition. 2020 *IEEE, 10th International Conference on Consumer Electronics (ICCE-Berlin)*, pp. 1-6. https://doi.org/10.1109/ICCE-Berlin50680.2020.9352194.

Liu, Y., Nand, P., & Hossain, M.A. (2023). Sign language recognition from digital videos using feature pyramid network with detection transformer. *Multimed Tools Appl.*, *82*, 21673–21685. https://doi.org/10.1007/s11042-023-14646-0.

Mittal, A., Kumar, P., Roy, P.P., Balasubramanian, R., & Chaudhuri, B.B. (2019). A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion. IEEE Sensors Journal, 19, 7056-7063. https://doi.org/10.1109/JSEN.2019.2909837.

Mekala, P., Gao, Y., Fan, J., & Davari, A. (2011). Real-time sign language recognition based on Neural network architecture. In Proceedings of the IEEE 43rd Southeastern Symposium on System Theory, Auburn, AL, USA, 14–16 March. https://doi.org/10.1109/SSST.2011.5753805.

Rao, M.S., Uma Maheswaran, S.K., Sattaru, N.C., Abdullah, K.H., Pandey, U.K., & Biban, L. (2022). A Critical Understanding of Integrated Artificial Intelligence Techniques for the Healthcare Prediction System. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 728-731. https://doi: 10.1109/ICACITE53722.2022.9823678.

Nandy, A., Prasad, J., Mondal, S., Chakraborty, P., & Nandi, G. (2010). Recognition of Isolated Indian Sign Language Gesture in Real Time. *Commun. Comput. Inf. Sci.*, *70*, 102–107. https://doi.org /10.1007/978-3-642-12214-9_18.

Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2021). Artificial Intelligence Technologies for Sign Language. *Sensors*, *21*, 5843. https://doi.org/10.3390/s21175843.

Rakesh, S., Bharadhwaj, A. & Sree, H.E. (2021). Sign language recognition using convolutional neural network. *In Innovative Data Communication Technologies and Application*, 59, 7-7-719. https://doi.org/10.1007/978-981-15-9651-3_58.

Rajam, P.S., & Balakrishnan, G. (2011). Real time Indian Sign Language Recognition System to aid deaf-dumb people. 2011 IEEE, *13th International Conference on Communication Technology*, pp. 737-742. https://doi: 10.1109/ICCT.2011.6157974.

Rao, M., Kumar, S., & Rao, K. (2023a). Effective medical leaf identification using hybridization of GMM-CNN. *International Journal of Experimental Research and Review*, *32*, 115-123. https://doi.org/10.52756/ijerr.2023.v32.009.

Reddy, N. S., & Khanaa, V. (2023). Diagnosing and categorizing of pulmonary diseases using Deep learning conventional Neural network. *International Journal of Experimental Research and Review*, *31*(Spl Volume), 12-22. https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.002.

Rao, M.S., Kumar, S.P., & Rao, K.S. (2023b). A Methodology for Identification of Ayurvedic Plant Based on Machine Learning Algorithms. *International Journal of Computing and Digital Systems*, *2023*, 10233-10241. http://dx.doi.org/10.12785/ijcds/140196.

Saha, A., & Yadav, R. (2023). Study on segmentation and prediction of lung cancer based on machine learning approaches. *International Journal of Experimental Research and Review*, *30*, 1-14. https://doi.org/10.52756/ijerr.2023.v30.001.

Sharma, P., & Anand, R.S. (2021). A comprehensive evaluation of deep models and optimizers for Indian sign language recognition. Graph. *Vis. Comput.*, *5*, 200032. https://doi: 10.1016/j.gvc.2021.200032.

Suneetha, M., Shivananda, B., Neha, B., Kiran, S.B. (2023). Hand gesture recognition and voice conversion for deaf and Dumb. *4th International Conference on Design and Manufacturing Aspects for Sustainable Energy* (ICMED-ICMPC 2023). https://doi.org/10.1051/e3sconf/202339101060.

Wadhawan, A., & Kumar, P. (2019). Sign Language Recognition Systems: A Decade Systematic Literature Review. *Archives of Computational Methods in Engineering*, *28*, 785 - 813. https://doi.org/10.1007/s11831-019-09384-2.

Wangchuk, K., Riyamongkol, P., & Waranusast, R. (2021). Real-time Bhutanese Sign Language digits recognition system using Convolutional Neural Network. *Science Direct. ICT Express*, *7*, 215–220. https://doi: 10.1016/j.icte.2020.08.002.

How to cite this Article:

Prasanthi Yavanamandha, Bodduru Keerthana, Penmetsa Jahnavi, Koduganti Venkata Rao and Chatikam Raj Kumar (2023). Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb. *International Journal of Experimental Research and Review*, *34*(Spl.), 26-35.

DOI : https://doi.org/10.52756/ijerr.2023.v34spl.004