2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)

Hand Sign Language Detection using Deep Learning And Computer Vision

1st Sujata Joshi
Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
sujata.joshi@nmit.ac.in

2nd Bhaskar Sharma
Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
bhaskar.shashank3333@gmail.com

3rd Dharmender
Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
dharmenderchoudhary21@gmail.com

4th Divyansh Tripathi
Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
divyanshsep9@gmail.com

5th Akshat Shah
Computer Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
akshatshah9816@gmail.com

2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) | 979-8-3503-1646-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIKIIE60097.2023.10390267

***Abstract*—This work presents an innovative system/app designed to assist individuals with visual or hearing impairments in communicating with the outside world using American Sign Language (ASL). By leveraging machine learning and image processing techniques, the proposed system captures and processes hand gestures, enabling the optimal transliteration of sign language alphabets and numbers into machine-readable English text. The system employs deep learning or random forest algorithms to classify the gestures, followed by mapping them to specific alphabets or numbers. The resulting text is presented audibly, bridging the communication gap and facilitating interactions between ASL users and those unfamiliar with sign language. Through the integration of technology, this work aims to enhance inclusivity and accessibility for individuals having disabilities, fostering greater engagement and interaction with the wider world.**

Keywords—Communication Barriers, Visual Impairment, Hearing Impairment, Sign Language, , Deep Learning, Random Forest, American Sign Language, Gesture Recognition, Hand Gestures.

I. INTRODUCTION

Sign Language (SL) is a refined construction of sign language and facial idioms for the purpose of expressing their perspectives about speech and hearing with normal people. Most normal people find it challenging to understand sign language.. So there is a huge gap in communication between the hearing impaired community and the general public. Since it is virtually impossible for human translators to consistently sponsor people with speech impairments in their daily activities, technical support for people with speech disabilities is inevitable[1]. Advances in science and technology may allow us to consider developing approaches that can transcribe gesture signals into human-like or machine- readable text. It facilitates dialogue between the sane and the debilitated.

American Sign Language (ASL) is a natural language that is used by hearing impaired people for communication. A person who uses this language make expresses their thoughts by the movements of the hands and face[2].

Hearing impaired people make use of hand signs to communicate. The normal people may find it challenging to understand the sign language. Hence there is a ere is a need of a system which recognizes the different hand signs and conveys the information to the normal people. In this work, we aim to develop a system which can identify hand gestures of the hearing impaired people and convey the appropriate text along with audio to the normal people.

This paper aims to explore the potential for developing technology that can interpret sign language gestures into human-like or machine-readable text. The paper provides background information on sign language, the lack of understanding between the community of hearing impaired and the society. Ultimately, the paper argues that advances in machine learning offer significant potential for improving the lives of individuals having hearing impairments thus facilitating greater communication between different communities.

II. LITERATURE SURVEY

Interaction, and by accurately recognizing and interpreting hand gestures, a vision-based system can facilitate seamless communication and control between humans and computers. In the context of this work, the main objective is to develop a real-time sign language recognition system using a vision-based approach. Sign language is a visual language used by individuals with speech and hearing disabilities for communication. By recognizing and interpreting sign language gestures in real time, the system aims to enable more inclusive and accessible communication for individuals who rely on sign language.

A literature review on Hand gesture recognition is taken up by authors in [3,4]. In this work, the authors have reviewed the different techniques for recognition of hand gestures and also explored the dataset size and techniques used. A work in [5] is taken up by the authors to study various segmentation techniques for recognizing hand gestures. Orientation histogram method applied have some problems which are; similar gestures might have different orientation histograms. In another work by authors in [6,7], the authors have studied human computer interaction and machine learning Similar kind of sign detection to recognize Arabic sign language is studied by authors in [8].

The authors in [9] have proposed a vision based sign language recognition system using Hidden Markov model. In another work by authors in [10], a system has been proposed to study the American sign language using Genetic Programming. With respect to Indian languages a work is taken up by the authors in [11] to recognize Indian language based on moments. It describes multiple steps collectively contribute to the successful recognition of signs in the Indian Sign Language, facilitating improved communication and

interaction for individuals using this languageAlso a work has been proposed by the authors in[12], in which the authors have worked on the sign language recognition using template matching technique. The authors in [13] have proposed a machine learning model for sign language interpretation using webcam images. In another work proposed by the authors in[14], a method is proposed to convert text to speech for English and Sinhala languages using Intelligent approach. In another work proposed by authors in [15], a work on sign language recognition using modified convolutional neural network model is taken up to explore its use in this domain. One drawback of this sign language recognition system is the limited scope of sign language it can handle.

Recognition for static signs is easy to implement but some signs involve motion in them due to which their recognition becomes more difficult.

III. Methodology

Research involves several steps to build a vision-based system for real-time sign language recognition.

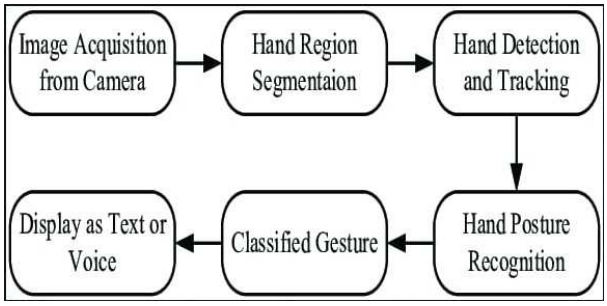

Fig. 1. Architecture of the system.

The steps are as shown in Fig. 1. The images are acquired using the camera. To begin, we collected a comprehensive dataset of hand gesture images, representing different signs used by individuals with hearing impairments. The dataset encompassed a wide range of hand orientations and variations in lighting conditions to ensure robustness and generalizability of the trained model.

The collected data will be pre-processed to extract features that are relevant for gesture recognition. This step involves removing noise, background subtraction, and segmenting the hand region from the image. The hand region area is then segmented using CNN model where an image is partitioned into smaller groups of pixels..

For the CNN model, we utilized a deep learning architecture specifically designed for image classification tasks. The dataset is partitioned into training, validation, and testing sets. The acquired dataset is trained using CNN model with multiple convolutional layers, pooling layers, and fully connected layers. The model is learned to extract meaningful features from the hand gesture images and make predictions based on these learned features.

Next, relevant features will be extracted from the pre-processed data. This step involves analysing the hand position and motion to extract spatial and temporal features.

The developed model is then tested using the test data. A vision-based system is developed to recognize sign language gestures in real-time. This system will use the trained model to classify the gestures in real-time video streams.

The performance of the system will be evaluated using a separate test dataset. The accuracy and efficiency of the system will be measured using standard evaluation metrics. Fig. 2 depicts the detection of left hand.

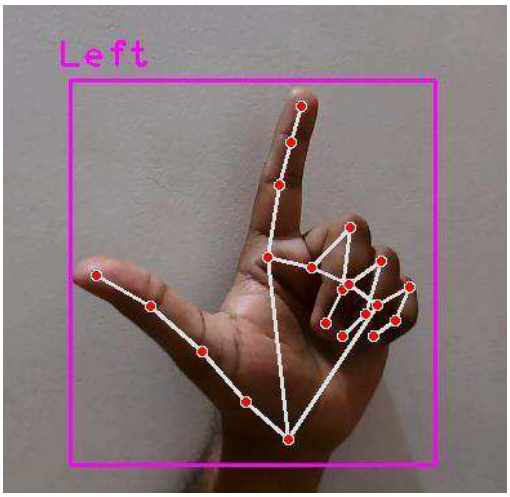

Fig. 2. Left hand detection using computer vision.

The pseudocode for prediction is presented here.

- Load the trained model from the pickle file
- Open a video capture object for accessing the webcam
- Initialize the hand detection components
- Create a dictionary mapping for label indexes to characters
- Repeat until images are read
  - Read a frame from the video capture using cap.read() and assign it to 'ret' and 'frame'.
  - Convert the frame to RGB color using cv2.cvtColor().
  - Process the frame with the hand detection model:
  - Use hands.process() to detect hand landmarks in the frame.
  - If hand landmarks are detected:
    - Draw the hand landmarks and connections on the frame using mp_drawing.draw_landmarks().
    - Extract hand features
    - Iterate over each detected hand landmark
    - Normalize the coordinates and append them to the 'data_aux' list
    - Calculate the bounding box coordinates for the hand region.
    - Predict the hand sign using the trained model
    - Initialize the text-to-speech engine
    - Convert the predicted label to speech

IV. Results

The developed interface utilizes a web camera to capture hand signs, and the system employs the trained model which is obtained in previous steps, to predict the corresponding hand gestures. Fig. 3 demonstrates an instance where the model successfully predicts the hand gesture as the alphabet 'I' and 'U'.

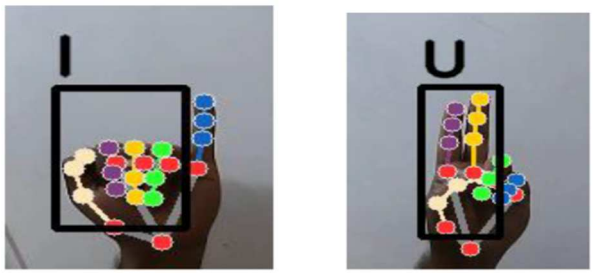

Fig. 3. Live Prediction of hand signs and prediction of hand signs letter 'I' and 'U'.

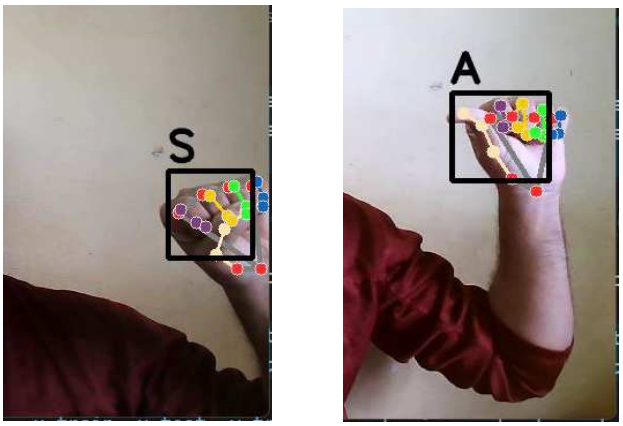

Fig. 4. Prediction of letter 'S' and letter 'A'.

The proposed system incorporates advanced techniques to differentiate between a human face and hand, allowing it to accurately read and interpret hand signs. Fig. 4 demonstrates an instance where the interface successfully predicts the hand sign as the alphabet 'S' and alphabet 'A'. The ability of the proposed interface to distinguish between a human face and hand is a crucial aspect of its functionality. This capability ensures that the system focuses on and analyses the hand gestures specifically, enabling precise and reliable interpretation of sign language.

	A	S	I	L	U
A	**839**	154	7	0	0
S	243	**749**	8	0	0
I	3	13	**935**	49	0
L	0	0	32	**917**	51
U	0	0	53	61	**886**

Fig. 5. Confusion Matrix.

A confusion matrix for five classes: A, S, I, L, and U is presented in Fig. 5. Based on the given table, we can interpret it as follows: For the class A, there were 839 instances correctly predicted as A, 154 instances incorrectly predicted as S, 7 instances incorrectly predicted as I, and no instances incorrectly predicted as L or U. For the class S, there were 243 instances incorrectly predicted as A, 749 instances correctly predicted as S, 8 instances incorrectly predicted as I, and no instances incorrectly predicted as L or U. For the class I, there were 3 instances incorrectly predicted as A, 13 instances incorrectly predicted as S, 935 instances correctly predicted as I, 49 instances incorrectly predicted as L, and no instances incorrectly predicted as U. For the class L, there were no instances incorrectly predicted as A, S, or I. Instead, there were 32 instances incorrectly predicted as I, 917 instances correctly predicted as L, and 51 instances incorrectly predicted as U. For the class U, there were no instances incorrectly predicted as A, S, or I. Instead, there were 53 instances incorrectly predicted as I, 61 instances incorrectly predicted as L, and 886 instances correctly predicted as U.

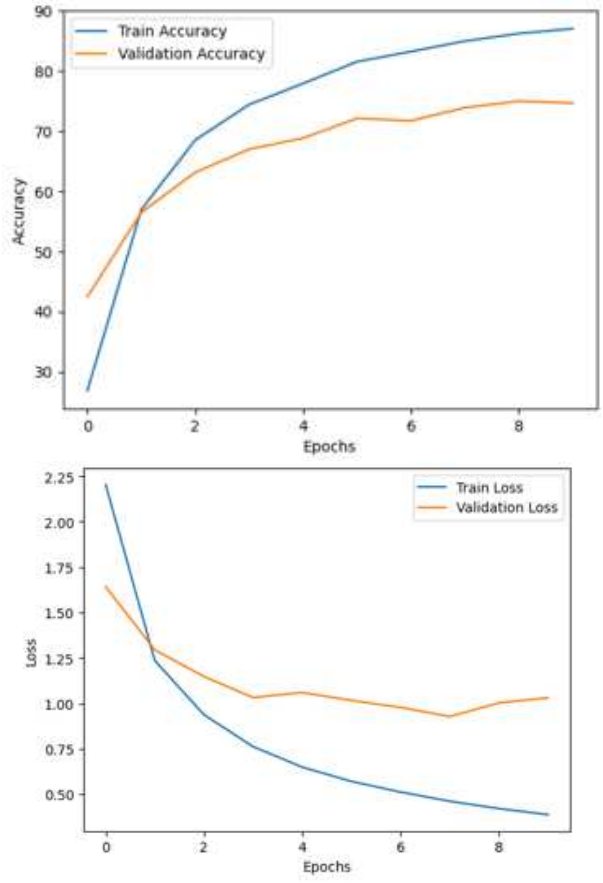

Fig. 6. Accuracy vs Epoch Graph and Loss vs Epoch Graph.

The graph in Fig. 6. shows the graph in terms of the performance metrics namely accuracy and Loss with respect to the number of epochs used in the training. A high training accuracy is generally considered a positive sign in machine learning. It indicates that the model can effectively learn and memorize the patterns present in the training data. The loss function measures the error or discrepancy between the predicted and true values in machine learning. Decreasing the loss function states that model has improved accuracy, generalization, optimization convergence, and facilitates

model evaluation, ultimately leading to better performance and more reliable predictions.

V. Conclusion

This work aimed to address the challenges faced by individuals with hearing impairments by developing a system for hand gesture recognition. The objectives are successfully achieved, which included collecting a comprehensive dataset of hand orientations for different signs used in sign language. The dataset thus curated is utilized and acts as a source for training and evaluating our machine learning models. By extracting relevant features from the collected images, we were able to capture the distinctive characteristics of each hand gesture. These features were essential for the accurate detection and classification of the gestures. We employed various machine learning techniques, including CNN algorithm to develop robust model that could effectively recognize and interpret the hand gestures. Our system holds great potential for enabling effective communication between individuals who use sign language and those who do not. By accurately interpreting hand gestures, we have created a tool that enhances inclusivity and promotes greater accessibility for individuals with hearing impairments.

While we have achieved significant milestones in our project, further research and innovation are necessary to improve the system's performance, expand its vocabulary, and refine its user interface. Collaboration with the hearing-impaired community and continuous feedback will be invaluable in making ongoing improvements and ensuring the system's relevance and effectiveness.

In conclusion, our work has made strides in developing a machine learning-based hand gesture recognition system that can enhance communication and empower individuals with hearing impairments to interact with the world more effectively.

References

[1] W. Li, H. Pu and R. Wang, "Sign Language Recognition Based on Computer Vision," 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2021, pp. 919-922, doi: 10.1109/ICAICA52286.2021.9498024

[2] Ss, Shivashankara & S, Dr.Srinath. (2018). American Sign Language Recognition System: An Optimal Approach. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.

[3] Khan, Rafiqul Zaman & Ibraheem, Noor. (2012). Hand Gesture Recognition: A Literature Review. International Journal of Artificial Intelligence & Applications (IJAIA). 3. 161-174.10.5121/ijaia.2012.3412.

[4] A. Kumar, K. Thankachan and M. M. Dominic, "Sign language recognition," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 2016, pp. 422-428, doi: 10.1109/RAIT.2016.7507939

[5] Kiran P, Parameshachari BD. Resource optimized selective image encryption of medical images using multiple chaotic systems. Microprocessors and Microsystems. 2022 Jun 1;91:104546.

[6] Kumar, Amit & Tewari, Dr. Naveen & Kumar, Dr Rajeev. (2019). Study towards the Analytic Approach for Human Computer Interaction using Machine Learning. The International journal of analytical and experimental modal analysis. 11. 1456-1466.

[7] Bansal, H., Khan, R., : A Review Paper on Human Computer Interaction, International Journals of Advanced Research in Computer Science and Software Engineering, ISSN: 2277-128X Volume-8, Issue-4, pp. 53-56, 2018

[8] Sidig, Ala & Luqman, Hamzah & Mahmoud, Sabri. (2018). Arabic Sign Language Recognition Using Optical Flow-Based Features and HMM. 10.1007/978-3-319-59427-9_32.

[9] Jayaprakash S, Nagarajan MD, Prado RP, Subramanian S, Divakarachari PB. A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. Energies. 2021 Aug 27;14(17):5322.

[10] Fahad Ullah, “American Sign Language Recognition System for Hearing Impaired People Using Cartesian Genetic Programming”, 5th International Conference on Automation, Robotics and Applications, pp.96-99, 2011.

[11] U. Patel and A. G. Ambekar, "Moment Based Sign Language Recognition for Indian Languages," 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017, pp. 1-6, doi: 10.1109/ICCUBEA.2017.8463901.

[12] S. Shrenika and M. Madhu Bala, "Sign Language Recognition Using Template Matching Technique," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020, pp. 1-5, doi: 10.1109/ICCSEA49143.2020.9132899.

[13]] K. Dabre and S. Dholay, "Machine learning model for sign language interpretation using webcam images," 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014, pp. 317-321, doi: 10.1109/CSCITA.2014.6839279

[14] P. Jayawardhana, A. Aponso, N. Krishnarajah and A. Rathnayake, "An Intelligent Approach of Text- To-Speech Synthesizers for English and Sinhala Languages," 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 2019, pp. 229-234, doi: 10.1109/INFOCT.2019.8711051

[15] Suharjito, H. Gunawan, N. Thiracitta and A. Nugroho, "Sign Language Recognition Using Modified Convolutional Neural Network Model," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), Jakarta, Indonesia, 2018, pp. 1-5, doi: 10.1109/ INAPR.2018.8627014.