Real-Time Marathi Sign Language Recognition Using Deep Learning Techniques

1st Rushali Deshmukh
Computer Department,
JSPM's Rajarshi Shahu College Of Engineering, Pune-411033
Maharashtra, India
radeshmukh_comp@jspmrscoe.edu.in

2nd Tushar Lahamge,
Computer Department ,
JSPM's Rajarshi Shahu College Of Engineering, Pune-411033
Maharashtra, India
lahamgetushar673@gmail.com

3rd Isha Phadatare
Computer Department,
JSPM's Rajarshi Shahu College Of Engineering, Pune-411033
Maharashtra, India
phadatareisha14@gmail.com

4th Dipali Shinde
Computer Department,
JSPM's Rajarshi Shahu College Of Engineering, Pune-411033
Maharashtra, India
drshinde192@gmail.com

5th Raghav Manhas
Computer Department,
JSPM's Rajarshi Shahu College Of Engineering, Pune-411033
Maharashtra, India
raghavmanhas785@gmail.com

***Abstract*— Communication is the most important thing for deaf and dumb people, where sign language plays an essential role in aiding them to communicate with others, however to teach us this form of communication was a challenging task. This work attempts to bridge this communication gap by performing a comprehensive study on real-time Marathi sign language detection. In this work, we have developed a custom dataset devoted to Marathi sign language gestures and investigated various deep learning models such as VGG-16, Convolutional Neural Networks (CNN),MobileNetV2, InceptionV3 for accurately identifying the signs. Rigorous experimentations were carried out and VGG-16 as the final model, boasting a validation accuracy is 95.55% & testing accuracy is 97.38%. In addition, we implemented an easy-to-use GUI which captures the most possible hand gestures in real-time and maps them to Marathi letters correctly. This has the potential to greatly enhance deaf and dumb people's communication exchanges with ordinary society.**

Keywords—- Marathi Sign Language Recognition (SLR), Deep Learning, VGGNET, Marathi Sign Language, Gesture Detection, Real-Time Application

I. INTRODUCTION

As an important communication tool, sign language is an integral part of the life of the deaf and mute people. While there are real-time sign language identification systems for languages like British Sign Language (BSL) and American Sign Language (ASL), regional languages like Marathi are still mainly ignored. The main cause of this gap is the absence of Marathi sign language infrastructure and databases, which creates a significant communication barrier in areas where it is widely used.

However, real-time sign language detection is difficult because of a large variety in gestures as well as variation caused by different lighting conditions and the lack of standard datasets. We address these challenges, drawing inspiration from prior works and focus on the Marathi Sign Language recognition. We generated a custom dataset and experimented with few models such as CNN, VGG-16, MobileNetV2 and InceptionV3. The model was chosen based on training accuracy, and VGG-16 performed well out of the four architectures. This model has a 95.55% validation accuracy and a 97.38% testing accuracy. GUI was created to detect real- time gestures and display Marathi letters helps the deaf community who use sign language to improve in their communication.

II. RELATED WORK

Ankita Wadhawan and Parteek Kumar [1] conducted the first observable comprehensive analysis of sign language recognition systems. This section provides a summary of earlier research on sign language recognition from 2007 to 2017. 396 research papers that are important to SLR systems are systematically reviewed and categorized by the survey. In the course of a comprehensive selection procedure, 117 articles are subjected to a further assessment and classification.

Srivastava et al [2] presented the method for creating an Indian Sign Language (ISL) dataset with a webcam and then using transfer learning to train the TensorFlow model which would result in the creation of real-time sign language recognition systems. The system achieved good accuracy while using a relatively limited dataset. Python and OpenCV were used to collect data from camera images.

Raheja et al [3] identify the challenge of Indian sign language using real-time dynamic hand gesture recognition systems is highlighted in this research. Captured motion pictures are first converted to the HSV color space for preprocessing, and then they are segmented according to skin pixels. To improve accuracy, depth information is also used simultaneously.

Amrutha, Prabh [4] implemented a vision-based system for recognition and detection of single hand gestures using machine learning techniques such as the feature extraction by convex hull and K-Nearest Neighbors (KNN) for classification. The goal of this method is to develop a simple SLR system that can reasonably accurately identify individual sign language gestures. by establishing the framework for a simple SLR system and demonstrating its capabilities with the 65% accuracy attained.

Athire et al [5] presented a novel vision-based gesture recognition system designed especially for Indian Sign Language (ISL) to help people with speech and hearing impairments. In real-time from live video streams, our system can recognize a variety of gestures, double-handed static movements. Additionally, it uses Support Vector Machine (SVM) classification to achieve outstanding precision in identifying dynamic words (89%) and finger-spelled alphabets (91%).

Katoch et a [6] presented a deep learning network designed specifically for sign language gesture recognition. Using a small number of sign frames to effectively capture both

spatial and temporal information is its distinctive focus. Three networks the dynamic motion network, accumulative motion network (AMN), and sign recognition network (SRN) are included in the framework's hierarchical sign learning module. Promising results were obtained when the system's performance was evaluated using the Arabic sign language KArSL-190 and KArSL-502 datasets. Additionally, using the KArSL-190 dataset, the suggested approach performed much better than competing methods by 15% in the signer-independent mode.

Luqman [7] presented a novel method for real-time recognition of Indian sign language alphabets from A to Z and numbers from live video feeds using the Bag of Visual Words model (BOVW). In addition to predicting and producing the labels as text, the system also translates them into speech.

Golekar et al [8] proposed a method to enable computers to communicate with and understand people in order to create user-friendly Interface that can interact with people who are deaf or have speech impairments. In order to understand real-time movement, the study investigates a number of computer vision techniques, including Support Vector Machines, Neural Networks, and Adaptive Hand Strength. By converting sign language motions into text, the proposed system aims to make it easier for the general public to understand and communicate with people of different abilities.

N. Brindha [9] suggested a real-time system that could interpret the gesture of a live sign. When a hand gesture is detected by the technology as input, the corresponding recognized text is immediately shown on the monitor. The location of the hand can be used to illustrate hand movements through the open CV, and text related to hand movements will show on the screen.

Rajalakshmi et al [10] suggested a method to create a logical system of sign language recognition by combining many components. This requires combining 3D deep neural networks, attention-based Bidirectional Long Short-Term Memory (Bi-LSTM), customized autoencoders, and a hybrid attention module for efficient feature extraction.

Obi et al [11] looked into, the approach comprised a number of consecutive steps. First, they used the Kaggle platform to choose an appropriate dataset. After that, developed a new model with a two-layer Convolutional Neural Network (CNN) architecture. The obtained dataset was then used to train this model. For our program, created and integrated a graphical user interface (GUI). These recognize gestures of hand and smoothly combine them into words that people could recognize. This method has an impressive 96.3 percent accuracy rate. It should be noted that maintaining constant hand gestures for each letter.

Gupta et al [12] suggested had been developed specifically to accomplish real-time sign language identification. It makes use of transfer learning and the TensorFlow Object Detection API. This involves creating a labeled map to link to sign motions, and TF records allow for more efficient data storage throughout the training phase.

Kanchan Dabre and Surekha Dholay [13] suggested a novel approach to ISL recognition that does not depend on markers. The main goal is to convert commonly used sentence-like motions from video to text and then back to audio output. There are several steps involved in the process. To identify and isolate hand shapes, a sequence of image processing techniques is first applied to the continuous stream of frames. Following that, a Haar Cascade Classifier is used to identify the unique signs and their associated meanings.

Zhou et al [14] aims to predict spoken language words and sign gloss patterns from sign motion pictures. The spatial multi-cue (SMC) module processes each frame, generating spatial attributes from several inputs.

Shinde,and Kagalkar [15] proposed method additionally includes a reliable and effective hand segmentation and tracking algorithm. A large collection of samples was collected to identify the 43 different phrases that make up Standard Marathi sign language.

Information regarding the Sign Language Recognition System being utilized by researchers was provided by Suharjitoa and Ricky Anderson [16]. The device that is used to obtain the data, data acquisition, processing mechanism, etc. are all covered in this study. The HMM approach can be implemented in three different ways. Sign language differs from one nation to another due to the lack of comparison between various methods.

Kodandaram et al [17] discussed about deep learning for sign language recognition. They had success in creating a useful system that can understand sign language. This system does not support body motions or dynamic movements, but it does cover the 0–9 numerals and the A alphabet.

In various layers,Shiling Huang, and Zhongfu Ye [18] utilized a tensor-train model. Their goal is to investigate the model's performance. They will use these models for sign language in the future.

According to Samiya Kabir Youme and Towsif Alam Chowdhury [19], SLR has been a much-researched area of study. To improve the model's performance, they conduct a variety of experiments.

The proposed Convexity Hull algorithm by Ashish S. Nikam and Aarti G. Ambekar [20] can be used exclusively for number recognition and finger point identification. It has facilities for developing templates for printing symbol recognition. They utilized a variety of feature extraction techniques.

Farman Shah and Muhammad Saqlain Shah [21] evaluated the proposed system using metrics such as F1 score, recall, accuracy, and precision. Multiple kernel learning (MKL) and support vector machines (SVM) are the basis of this system, which uses them to categorize the characteristics that have been extracted. The accuracy of SURF features is only 15%.

Dr. Talwekar [22] is a linguist. One of the most important tools created for the deaf is linguistic communication. The main automatic Arabic language communication recognition system that supports HMM was presented in this paper. The Indian linguistic communication recognition system is implemented in this study.

Mathieu De Coster, Dambre [23] achieved a low accuracy on the unseen test set for a vocabulary of classes by using a Multimodal Transformer Network.

Arunabh Kalita and Ananya Neog [24] utilized Python modu les to recognize speech.Additionally,this model is converted into a mobile application.This program generates pictures in response to the majority of the sounds.

Machine learning and feature extraction were proposed by Hein et al [25] used Myanmar National Sign Language for their dataset. In the future, they plan to use Myanmar Sign Language hand and facial detection.

The paper [26] is a review of current sign language recognition techniques that are targeted for real-time applications in the field of communication. It points out the lack of study around Marathi Sign and presents a dataset dedicated to it, its respective model. The paper gives a survey of popular methods like CNN, indicates their advantages and disadvantages but generally targets for more precise systems to more inclusive communication.

III. METHODOLOGY

Dataset:

The dataset for Marathi Sign Language Recognition project consisting of around 30,000 sign images. Trained and tested our models, injuring different architectures we have managed to obtain different levels of accuracy. A better starting point for us was a simple CNN model that could get 84% accuracy. The accuracy increased significantly with the use of the VGG-16 model, reaching 97.38%, indicating its effectiveness for this task and its improved capacity to catch small characteristics in the images. On the other hand, MobileNetV2 which was designed to be competitive in speed and size has a recognition accuracy of 74.10%, meaning that some tradeoff between model size and performance. As used in this paper, training epochs are reduced by experimentally to determine the adaptation and accuracy of model.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
conv2d_3 (Conv2D)	(None, 12, 12, 128)	147,584
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 128)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 512)	2,359,808
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 43)	22,059

Total params: 2,622,699 (10.00 MB)
Trainable params: 2,622,699 (10.00 MB)
Non-trainable params: 0 (0.00 B)

Figure 1: CNN Layer architecture

Model: "sequential_2"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 4, 4, 512)	14,714,688
flatten_2 (Flatten)	(None, 8192)	0
dense_4 (Dense)	(None, 512)	4,194,816
dropout_2 (Dropout)	(None, 512)	0
dense_5 (Dense)	(None, 43)	22,059

Total params: 18,931,563 (72.22 MB)
Trainable params: 4,216,875 (16.09 MB)
Non-trainable params: 14,714,688 (56.13 MB)

Figure 2: VGG-16-layer Architecture

IV. IMPLEMENTATION & RESULT

We experimented using a handpicked dataset that consisted of images representing the 43 Marathi letters, to ensure both quality and relevance of our findings. The creation of this dataset was made possible by collecting around 30k images, has the intention to capture hand signs in different conditions as generally as possible. The dataset is also complete with PNG images of various sizes and resolutions, from 1380x776 to the full size at 1920x1080 pixels. This carried out a number of important steps in order to achieve high accuracy in our Marathi sign language recognition system. In order to follow the CNN model specifications, the preprocessing step included increasing images to 224x224 pixels, standardizing pixel values between 0 and 1, and utilizing data augmentation methods such as flipping and rotation to improve model generalization.

Convolutional neural networks (CNNs), more especially the VGG-16 and MobileNetV2 architectures, were used to extract the features. Convolutional and pooling layers were used in these models to extract important visual characteristics, and the ReLU activation function was then used to add non-linearity and speed up convergence. In order predict one of the 43 Marathi characters, the trained model applied the SoftMax function to the output after passing through these layers in a forward motion. We split the 80% of the dataset for training and 20% of the dataset for testing to develop a reliable model for recognition. To see how the models performed in various scenarios, we trained them using epoch settings of 10, 20, 50, and 100.To ensure the best possible performance for the current classification tasks, these models were adjusted.

A test set, which made up 20% of the dataset, was used for validation, taking evaluation parameters like accuracy into account. The adjusted hyperparameters, used regularization techniques like dropout, and used data augmentation to improve model generalization in order to increase precision. These procedures ensured reliable results and excellent precision when combined with transfer learning from previously trained models and dataset fine-tuning.

Accuracy and Loss Evaluation:

The training process of each model was visualized using graphs depicting both accuracy and loss across different epochs. The accuracy graph demonstrated a steady improvement over time, while the loss graph showed a significant decrease, indicating the model's growing proficiency.

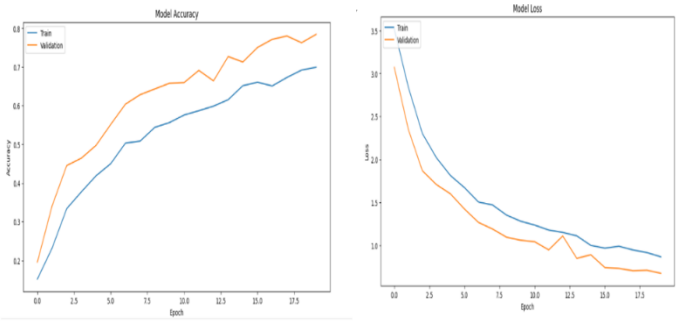

Figure3. Accuracy & Loss Graph of CNN Model

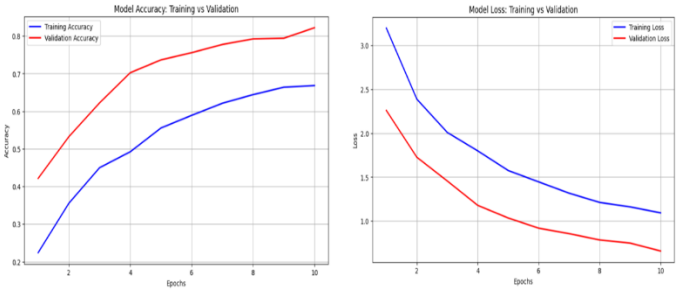

Figure 4. Accuracy & Loss Graph of VGG-16 Model

Classification Metrics and Confusion Matrix:

To further evaluate of our models, utilized confusion matrix to capture the model's classification accuracy across the dataset. The matrix highlights true negative, true positive, false negative, and false positive predictions, providing deeper insights into the model's strengths and weaknesses.

The classification accuracy of model from confusion matrix over dataset to even evaluate our models better. It gives more clarity to model by showing the true negative, true positive, false positive and false negatives predictions from those can determine some advantages of models and its disadvantages.

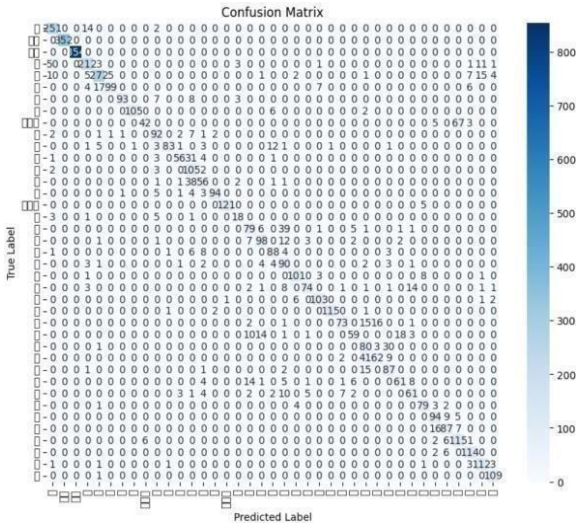

Figure 5: Confusion matrix

In addition to the AUC-ROC score, evaluated performance using recall, precision and F1-Score. These metrics compares each class and was computed for every one of them to make sure that how good a model is at predicting the different categories available within out dataset.

To evaluate the models' capacity to generalize over a range of image kinds and resolutions, additional datasets were used for validation. Even though the initial results are positive, more research is required to calculate the model's accuracy and adaptability to various datasets and imaging circumstances. Definitely, the model will learn real life examples and not only about training dataset after this.

Output and Results:

The architecture of our Marathi sign language recognition system follows a clear flow from capturing the hand sign to displaying the predicted output. When a user performs a hand sign corresponding to a particular Marathi letter, the system processes the input in real-time, leading to the display of the recognition results.

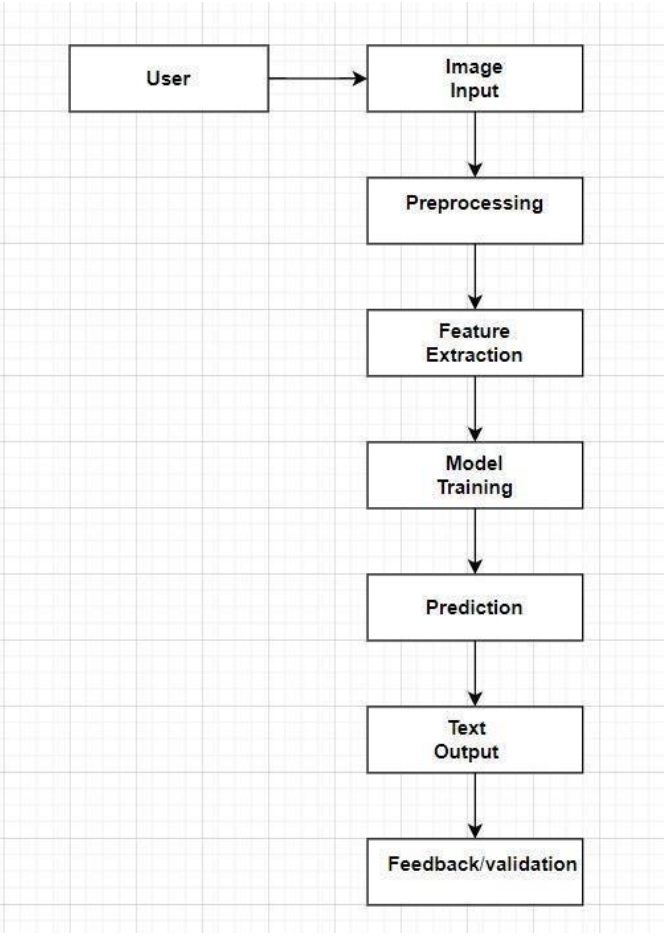

Figure 6: Proposed system

The comparison of models for Marathi Sign Language recognition, VGG-16 performed the best with an accuracy of 97.38%, owing to its deep architecture with 16 layers, including 13 convolutional layers. CNN, with fewer layers, achieved a moderate accuracy of 83.32%, while MobileNetV2, known for its lightweight structure, had the lowest accuracy at 74.10%. InceptionV3 demonstrated an accuracy of 81.68%. The results highlight those deeper networks, like VGG16, generally offer better accuracy for complex tasks like sign language recognition.

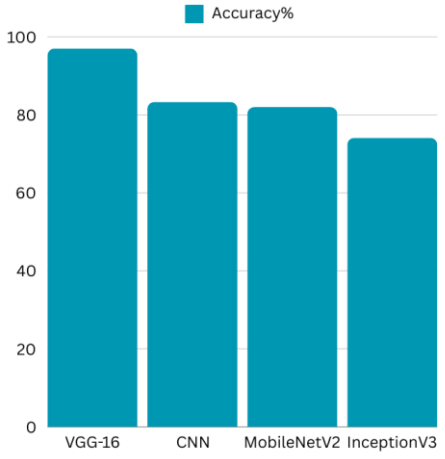

Figure 7: Accuracy Vs Models Comparison Graph

For instance, when the user shows the hand sign for the Marathi letter 'अ', the system captures the gesture, processes it through the architecture, and predicts the letter. The output frame then updated to reflect the recognition result.

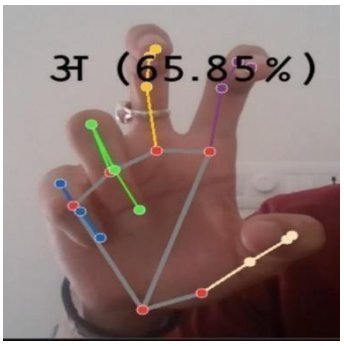

Figure 8: Letter Prediction

By presenting the detected letter and its corresponding accuracy, the system enhances user interaction, ensuring that users can trust the predictions made by the model. The intuitive interface allows users to see how well the model performs in real-time, making it an engaging and informative experience.

V. CONCLUSION

This study uses deep learning models to provide an innovative approach for real-time Marathi SLR. After a custom dataset was created and many models, including CNN, VGG-16, MobileNetV2, and InceptionV3, were tested, it was found that VGG-16 performed the best, reaching testing accuracy of 97.38%. Furthermore, a user- friendly graphical user interface (GUI) was developed that can accurately translate hand motions into Marathi characters in real-time, offering a useful tool to bridge the gap in communication and greatly enhancing the deaf and dumb community's involvement with the public. Future work can concentrate on the usage of advanced deep learning architectures and optimization techniques to increase model accuracy and performance. This model can be very flexible when we expand our dataset by adding a few other sign languages, for instance, ASL or ISL. It can also provide a foundation for future research to build more sophisticated models that uses real-time translation through an NLP pipeline and transfer learning techniques, enabling it to rapidly deploy the model. Also, by designing a powerful whole-sentence reading model with advanced sequence prediction algorithm or Transformer network to identify sentences will automate sign language recognition on-the- fly and lead this system closer toward seamless real-time communication.

REFERENCES

[1] Ankita Wadhawan and Parteek Kumar “Sign Language Recognition Systems: A Decade Systematic Literature Review.” Arch Computat Methods Eng 28, 785–813 (2021).https://doi.org/10.1007/s11831-019-09384-2

[2] Srivastava, S., Gangwar, A., Mishra, R., Singh, S. (2022). “Sign Language Recognition System Using Ten-TensorFlow Object Detection API”.In: Woungang, I., Dhu-randher, S.K., Pattanaik, K.K., Verma, A., Verma, P. (eds) Advanced Network Technologies and Intelligent Computing. ANTIC 2021. Communications in Computer and Information Science, vol 1534. Springer, Cham.https://doi.org/10.1007/978-3-030-96040-7-48

[3] Raheja, J.L., Mishra, A. and Chaudhary, “A. Indian sign language recognition using SVM. Pattern Recognit. Image” Anal. 26, 434–441(2016). https://doi.org/10.1134/S1054661816020164

[4] K. Amrutha and P. Prabu, “ML Based Sign Language Recognition System,” 2021 International Conference on Innovative Trends in Information Technology (IC- ITIIT), Kottayam, India, 2021, pp. 1-6, doi: 10.1109/ICI-TIIT51526.2021.9399594.

[5] P.K. Athira, C.J. Sruthi, A. Lijiya “A Signer Independent Sign Language Recognition with Co- articulation Elimination from Live Videos: An In- dian Scenario Author links open overlay panel

[6] Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, “Indian sign Language recognition system using SURF with SVM and CNN” www.sciencedirect.com/journal/array

[7] H. Luqman, ”An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion,” in IEEE Access, vol. 10, pp. 93785-93798, 2022, doi: 10.1109/ACCESS.2022.3204110.

[8] Dipalee Golekar, Ravindra Bula,Rutuja Hole, Sidheshwar Katare, Prof. Sonali Parab, “Sign language recognition using Python and OpenCV” www.irjmets.com

[9] N. Brindha, Hand Gesture Detection, http://restpublisher.com/book- series/daai

[10] E. Rajalakshmi et al., “Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture,” in IEEE Access, vol. 11, pp. 2226-2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[11] Yulius Obi, Kent Samuel Claudio, Vetri Marvel Budiman, Said Achmad, Aditya Kurniawan, “SLR for communicating to people with disabilities” www.elsevier.com/locate/procedia

[12] U. Gupta, S. Sharma, U. Jyani, A. Bhardwaj and M. Sharma, “Sign Language Detection for Deaf and Dumb students using Deep Learning: Dore Idioma,” 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-5, doi: 10.1109/CISCT55310.2022.10046657.

[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), Mumbai, India, 2014, pp. 317-321, doi: 10.1109/CSCITA.2014.6839279.

[14] H. Zhou, W. Zhou, Y. Zhou, and H. Li, “Spatial-Temporal Multi-Cue Network for Sign Language Recognition and Translation,” in IEEE Transactions on Multimedia, vol. 24, pp. 768-779, 2022, doi: 10.1109/TMM.2021.3059098.

[15] Amitkumar Shinde and Ramesh Kagalkar. Article: “Advanced Marathi Sign Language Recognition using Computer Vision.” International Journal of Computer Applications 118(13):1-7, May

[16] Suharjitoa and Ricky Anderson “Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output”,2017, Elsevier, doi: https://doi.org/10.1016/j.procs.2017.10.028

[17]. Satwik Ram Kodandaram and 2N Pavan Kumar “Sign Language Recognition”,2021, doi: https://www.researchgate.net/publication/354066737

[18] Biao Xu, shilling Huang, and Zhongfu Ye “Application of Tensor Train Decomposition in S2VT Model for Sign Language Recognition”2021, IEEE, doi: 10.1109/ACCESS.2021.3059660

[19] Samiya Kabir Youme and towsif alam Chowdhury, “Generalization of Bangla Sign Language Recognition Using Angular Loss Functions”,2021,IEEE, doi: 10.1109/ACCESS.2021.3134903

[20] Ashish S. Nikam and Aarti G. Ambekar, “Sign Language Recognition Using Image-Based Hand Gesture RecognitionTechniques”,2016, IEEE, doi: 10.1109/GET.2016.7916786

[21] F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud, and D. S. Abdelminaam, "Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language," in IEEE Access, vol. 9, pp. 6754867558, 2021, doi: 10.1109/ACCESS.2021.3077386.

[22] Patil, Sandeep Baburao, and Rajesh H. Talwekar. "Implementation of Indian sign language recognition system using scale-invariant feature transform (sift)." International Journal of Computer Science and Information Security 15.2 (2017): 493.

[23] De Coster, Mathieu, Mieke Van Herreweghe, and Joni Dambre. "Sign language recognition with transformer networks." 12th international conference on language resources and evaluation. European Language Resources Association (ELRA), 2020.

[24] Neog, Anannya Priyadarshini, Arunabh Kalita, and Ms Nithyakani Pandiyarajan. "Speech/text to indian sign language using natural language processing." (2023), doi: 10.21817/indjcse/2023/v14i3/231403030

[25] Z. Hein, T. P. Htoo, B. Aye, S. M. Htet and K. Z. Ye, "Leap Motion based Myanmar Sign Language Recognition using Machine Learning," 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering(ElConRus), St. Petersburg, Moscow, Russia, 2021, pp. 23042310, doi: 10.1109/ElConRus51938.2021.9396496

[26] R. Deshmukh, T. Lahamge, I. Phadatare, D. Shinde and R. Manhas, "A Comprehensive Review on Real-Time Sign Language Detection for Deaf and Dumb People," 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2024, pp. 169-176, doi: 10.1109/ICSADL61749.2024.00034.