Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319 - 2526

Volume 15 Issue 01s, 2026

Gesture-Based Recognition of ISL to Inclusive Communication in Marathi Text

[1]Rohini Vijay Daund, [2]Dr. Monali Y. Khachane

[1]Research Scholar, Dept. of Computer Science, Dr. Annasaheb G.D. Bendale Mahila Mahavidhyalaya, Jalgaon, India

[2]Assistant Professor, Dept. of Computer Science, Dr. Annasaheb G.D. Bendale Mahila Mahavidhyalaya, Jalgaon, India

Email: [1]daund.rohini.32@gmail.com, [2]monalikhachane@gmail.com

Peer Review Information

Submission: 08 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

Hand Gestures, MediaPipe, ANN, TensorFlow, Indian Sign Language.

Abstract

Hand gestures are the media of individuals with total hearing loss but not useful to connect with ordinary person. The hand gesture language recognition becomes most potential area of research. In the present research work the MSL recognition system is developed with help of mediaPipe for real time recognition. It is not possible every time to get the interpreter for translation and understanding of Marathi sign language for common people. In the previous research work many researchers had undertakes the machine learning for classification and extraction of feature in the light of development of said system. They have faced several challenges in these research areas. The present research work will overcome the challenges of these research areas. As a result, it is found that by using direct image data and by using Euclidean distance function the ANN model shows the higher accuracy (approx. 95. %) than logistics regression, decision tree and SGD Model.

Introduction

The hard of hearing and dumb people have communication problem due to that, they unable to communicate effectively with normal people. Marathi Sign language serves as vital tool for eradication of communication problem of hard of hearing person. These individuals can convey message with the other person only who knows the Sign Language and that's very rare. So that there is need of real time SL system. The present research work will translate ISL hand gestures in to the Marathi text. This system employed the mediaPipe framework using python. The Indian sign language recognition system will promote accessibility and regional inclusivity, improving the system's use for native Marathi speakers. In the fields of healthcare, education and public services, it will highly support to the digital inclusion of hearing disabled people. The effectiveness of Indian sign language recognition will also enhance and it will support to the formation of assistive technologies in the context of Indian linguistic diversity. The land marker task which constitutes the hand key point model bundle with detect 3-dimension 21 hand key points coordinates which done by the mediaPipe library. The 30,000+ real captured images are trained in this model and many synthetic hand models are used as a various background. Palmer surface detection model and hand key points detection model are pack in the hand keypoint model bundle. Inside the inputs image the palm detection model locates the hands. Specific hand knuckles are recognized by the hand keypoint detection model.

Objectives

1. To create the dataset for Indian sign language.

2. To develop hand tracking module through Media Pipe.
3. To translate 12 Indian sign language hand gestures into Marathi text.

Related Work

Review of existing literature was conducted and summarized in the Table No 1. Dahibavkar Swaraj et. al. was designed a system to enable deaf people to interact with others by translating Marathi Sign Language into the alphabet. Both right- and left-handed gestures are supported using skin color detection [1]. Shinde Amitkumar, Ramesh Kagalkar et. al. works on static gesture recognition only for Alphabets of Marathi Sign Language (MSL) was reported [2,3]. A Bain, S Birajdar et. al.used KNN classifier for classification [4,5].Chandwani Laveen et. al. worked on static gesture recognition only for Alphabets, Number and some words for American Sign Language (ASL) [6-8]. Kaushal Goyal, Rohit S. S., et. al. used MediaPipe and KNN techniques for American Sign Language (ASL) recognition [8,9]. R Kumar, A Bajpai et. al. only works on numberand alphabet [10-12]. Moustafa, A. M. et. al. proposed system which works on Arabic Sign (ArSL) Language recognition for Arabic 28 Alphabets applying Media pipe and CNN[13]. K. Gomase et. al. Works ASL with Media pipe and KNN classifier for letters [14].It was noted that recognition of reactive computing on Indian sign language (ISL) recognition for Marathi language is still remain to explore .

Table 1: Related Work

No.	Purpose	Sign Language	Accuracy	Techniques Used	Limitations
1	Translation of Marathi Sign Language into the alphabet (Swar and Vyanan).	Marathi Sign Language (MSL) Dataset size:8600	95%	Canny Edge-Based Segmentation Technique, CNN	Non-skin-colored cloths, dark colored cloths and also non-skin colored background are required.
2	This system offers static Marathi language sign recognition using computer visualization techniques.	Marathi Sign Language (MSL) Dataset size:1000	--	Image Preprocessing, edge detection	System works static data
3	Proposed system works on Marathi Sign Language recognition through offline and Web camera mode.	Marathi Sign Language (MSL) Dataset size:1000	90%	Feature extraction, Pattern matching process, skin filtering	System works on static gesture recognition only for Alphabets.
4	This system detects alphanumeric data in ISL.	Indian Sign languages (ISL) Dataset size:2500	98%	K-means clustering algorithm, KNN, SVM, and Naive Bayes classification	Model does not support to captures temporal inflections in ISL. only work on alphabet and number
5	This system uses a hierarchical centroid feature extractor, a direct pixel value classifier, and a neural network	Regional Language (RSL) Dataset size:5000	97.10%	Neural networks and KNN classifier.	It works only for regional sign language recognition system.

6	A simplified methodology applying Media Pipe library and ML algorithms	American Language (ASL) Dataset size:4000	99%	Media Pipe, SVM	System works on static gesture recognition only for Alphabets, Number and some words
7	Detection of Assamese Sign Language.	Assamese Language (ASL) Dataset size:2094	99%	feed-forward neural network	Limited only for Assamese Sign Language.
8	Sign Language using Media Pipe and recognition through Computer vision alphabets.	American Language (ASL) Dataset size:1800	98%	Media Pipe, KNN	Only Works on 26 alphabets.
9	Detection of American Sign Language applying dynamic approach	American Language (ASL) Dataset size:4500	99.95%	CNN, CV Zone, Media Pipe	Only Works on American Sign Language.
10	Multiple Language Recognition system for alphabets and numbers	American Language (ASL) Dataset size:15600 Indian Language (ISL) Dataset size:4972 Italian Sign Language (Ita.SL) Dataset size:12856 Turky Sign Language (TSL) Dataset size:4124	99%	Media Pipe Naive Bayes, SVM, KNN	Only Works on alphabets and numbers
11	Applying dynamic approach for complex hand gesture detection	American Language (ASL) Dataset size:15600 Indian Language (ISL) Dataset size:4972 Italian Language (Ita.SL) Dataset size:12856	99%	Media Pipe's	Works on Alphabets and Numbers
12	Works on numbers only	American Sign Language (ASL)	95.7%	Media Pipe, KNN	Works on Numbers and only some hand Sign
13	Proposed system works on Arabic Sign (ArSL) Language recognition for Arabic 28 Alphabets	Arabic Sign (ArSL)Language Dataset size:7057 +5600	97.1%	Media Pipe, CNN	Works on only Arabic Alphabets
14	Make a recognizer for A-Z alphabet using MediaPipe	American Sign Language (ASL)	(86-91) %	Media Pipe, KNN	Works on only A-Z Alphabets

Gaps or Limitations of Existing System:

1. Non-Skin colored cloths, Dark background, hand gloves are required
2. Static gesture recognition for Alphabets and Numbers.

Applications of Proposed System:

1. Effective communication leads to the better Educational and Social development.
2. Real time communication for hard of hearing and dumb people.

Methodology

A. Dataset Preparation:

The data set composed from students and teachers of "पडसाद कर्णबधीर विद्यालय, सिडको, नाशिक" and "श्रीमती माई लेले श्रवण व विकास विद्यालय, गंगापूर रोड, नाशिक." The Dataset was prepared by collecting images of hand gestures of Indian Sign language. The 12 hand gestures include कॉल करा, छान, चांगले नाही, चार, इच्छाशक्ती, मजा करा, एक, शांतता, थांबा, ठीक आहे, तीन, विजय. The 120 images are captured for each sign from the age group 14+ of deaf and dumb people. The sample images are given below:

Table 2: Sample Images of Dataset

The proposed framework is shown in Fig. 1.

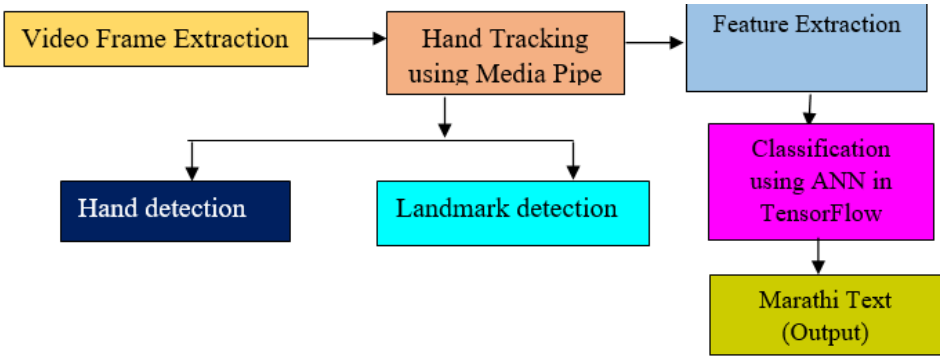
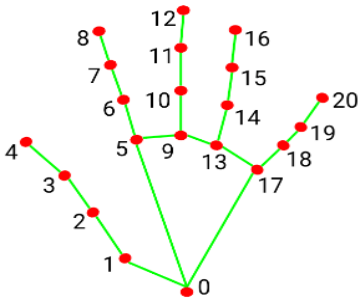


Fig 1: Research method for ISL into Marathi text

B. Video Frame Extraction*:*
The Recorded video or live feed is captured into program using OpenCV library, which extract video frames into separate images. These separate images act as an input to subsequent models and process.

C. Hand Tracking:



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Fig 2: MediaPipe Hand land Mark

Separate video frame is passed into Media pipe's hand landmark detection model as an input. The model detects hands in the frame and returns output as co-ordinates of hands and hand landmarks in the images. The methodology of detection of hands in each frame therefore in continuous video help us to successfully track the hand in the video. Media Pipe Hands employs ML pipeline containing multiple models work at a time. A palm detection model works on the complete image and gives an oriented hand bounding box. A hand landmark model works on the selected image region cropped by the palm detector and gives high-fidelity 3-Dimension hand key points. Both of these models are as explained below[15].

i) Palmer surface detection model:

In this model, the key points of objects i.e. inner surface of hand and fists are easier than hands detection with articulated fingers. The palm is small object, in case of two hand self-occlusion like hand shakers, the non-maximum suppression algorithm works efficiently. The palms might be demonstrated by utilizing the square bounding boxes; hence it decreases the count of anchors by the factor of 5. So, utilizing this optimization model work fast.

ii) Hand keypoint Model

Performing the palmer surface detection, it detects the land mark accurate key points of hands knuckle coordinates which located the 21 3 D hand key point coordinates by using the regression it detected the hand regions which is the direct coordinated recognition of hand knuckle.

The Media Pipe developed the hand keypoint model which utilized the manually annotated 30,000 real captured images with 3-Dimension 21 coordinates. This model also detects the Z value from the image which is shown below. It will give the best result for all hand gestures.

D. Feature Extraction:

After getting the (X, Y, Z) co-ordinate of each one of 21-hand landmark as an output from hand landmark detection model. In the process of Feature extraction, the Euclidean distance formula is used for the calculation of distance between each landmark, this process denoting distance between all possible combinations of 210 features of each image which dealt as an output of this process. For illustration, If the hand land mark model gives an outputs of WRIST landmark coordinates i.e. (X1, Y1, Z1) and THUMB CMC landmark coordinates are (X2, Y2, Z2), then new feature WRIST THUMB CMC is calculated using following Euclidean distance formula:

WRIST_THUMB_CMC=

$$\sqrt{((X_{(2)}-X_{(1)})^2+(Y_{(2)}-Y_{(1)})^2+(Z_{(2)}-Z_{(1)})^2)} \quad (1)$$

E. Classification Using ANN in TensorFlow:

The google have created the TensorFlow machine learning framework which used to train and build DL models[16].Performing numerical computations, the TensorFlow library is used and then for data flow graphs these computations are done. The nodes show the mathematical operations and edges shows the data in the graph this is a multidimensional array for data also called Tensors which are show the relations between the edges. In a training process TensorFlow allow to form input pipelines for producing input data and also preprocess them. A custom sequential ANN model is built to predict the hand gestures into one of the 12 classes by 210 features extracted from image. The model contains 4 hidden dense layers using as an activation function (Relu). In 4

hidden layer one is dropout layer. This layer drops 25% node information from model to remove the problem of overfitting. In order to build the model, 120 hand images with one of the 12 hand symbols were split into training (80%) and testing (20%) data. The dependent variable (which is nothing but the name of the symbol) is encoded using label encoder. The model is trained with training dataset where softmax activation function is used on output layer, the model's loss function is tracked using sparse categorical cross entropy and is optimized using Adam optimizer.

F. Marathi Text (Output):

The output of the classification model is decoded using label encoder as well as printed on image with hand landmarks using OpenCV and then Marathi text displayed on the screen.

Results and Discussion

In a result, many machine learning models are used for hand gesture detection. The models are assessed on the based-on Train accuracy, Test accuracy, F1 Score, precision and recall. In the Direct image data model type, the ANN shows 86.64% for Train accuracy, 87.74% for Test accuracy, 93.46% for F1 Score, 93.97% for precision and 92.55% for recall. The ANN Model shows the higher accuracy in every parameter than decision tree model, logistic regression model and SGD Model. By applying Euclidean distance function, the ANN higher accuracy as compared to using direct image data. The effectiveness of ANN shows 95.05% for Train accuracy, 95.02% for Test accuracy, 95.08% for F1 Score, 95.52% for precision and 95.02% for recall. The table shows the accuracy of ANN, Decision tree, logistic regression and SGD Model.

Table 3: Result and Accuracy

Model type	Model name	Train Accuracy	Test Accuracy	F1 Score	Precision	Recall
Using direct image data	ANN	86.64%	87.74%	93.46%	93.97%	92.55%
	Decision tree	100.00%	70.01%	69.78%	69.39%	67.80%
	Logistic Regression	79.73%	78.17%	80.16%	80.81%	79.47%
	SGD	75.95%	76.03%	74.47%	79.95%	76.46%
Using distance function	ANN	95.05%	95.02%	95.08%	95.52%	95.02%
	Decision tree	100.00%	83.75%	81.55%	83.15%	81.97%
	Logistic Regression	86.28%	86.55%	86.76%	89.00%	86.63%
	SGD	85.34%	86.01%	86.62%	88.94%	87.24%

The confusion matrix by using ANN algorithm for 12 gestures.

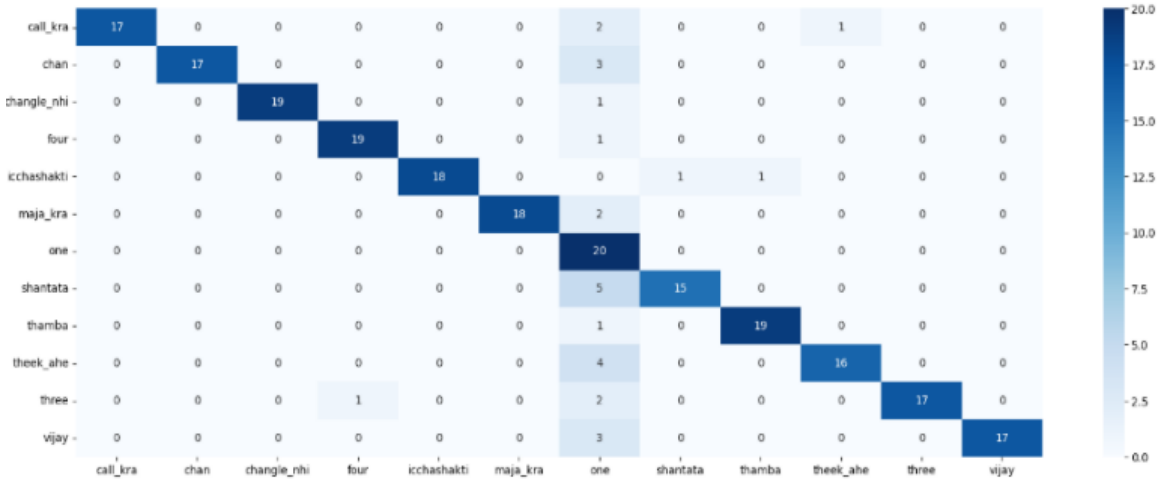
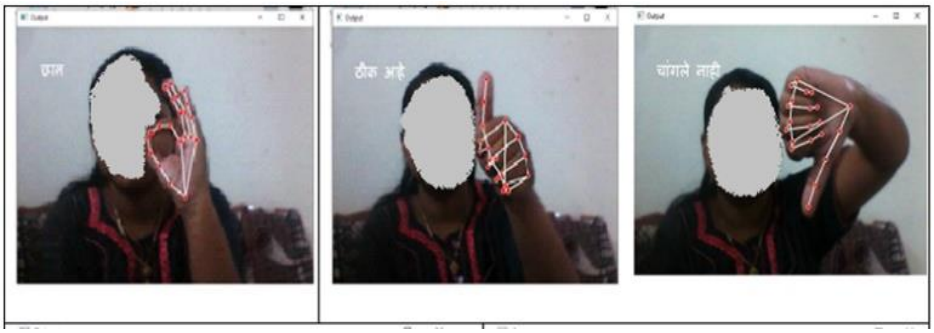


Fig 3: Confusion Matrix by using ANN Algorithm

In the above confusion matrix, the diagonal elements show the better result of by using ANN algorithm for 12 gestures. The ANN was performed on the dataset in TensorFlow. It can be observed from the following output the Marathi signs like छान, ठीक आहे, चांगले नाही, आनंद घ्या, मला कॉल करा are detected which are presented in the following table.

Table 4: Translation of Hand Landmarks in to Marathi Text



Conclusion

The proposed system will helpful in educational and social development of deaf and dumb people as well as will be supportive to the new researchers. As a result, it is found that by using direct image data and by applying Euclidean distance function the ANN model shows the higher accuracy (approx. 95%) than logistics regression, decision tree and SGD Model. In the preset research work after the data analysis the results gives approx. 95% accuracy. The model utilized mediaPipe for feature extraction and ANN for classification, providing a robust and accurate solution for recognizing.

In proposed system it can be concluded that by using mediaPipe and TensorFlow framework, Real time 12 Indian sign gestures are detected. The proposed system evaluated on 12 Indian sign gestures. So, in future a greater number of Indian sign gestures are considered.

Future Scope

1. Addition of more Indian Sign gestures.
2. Development of Mobile application.

Acknowledgment

I am incredibly grateful to my research guide, for their priceless advice, inspiration, steadfast support and guidance during my research journey. I would like to thanks and deepest appreciation to the principal, teachers and students of "पडसाद कर्णबधीर विद्यालय, सिडको, नाशिक" and "श्रीमती माई लेले श्रवण व विकास विद्यालय, गंगापूर रोड, नाशिक, whose suggestions and knowledge significantly improved the calibre of this study. I would especially like to thank the members of the hard of hearing community who shared their stories and made it stress-free for me to see how useful sign language recognition is in real-world situations

References

[1] S. Dahibavkar, J. Dhopte, M. Patole, and S. Madachane, "Marathi sign language recognition," 2020.

[2] A. Shinde, "Sign language to text and vice versa recognition using computer vision in Marathi," International Journal of Computer Applications, vol. 118, no. 13, pp. 1–7, 2014.

[3] A. Shinde and R. Kagalkar, "Advanced Marathi sign language recognition using computer vision," International Journal of Computer Applications, vol. 118, no. 13, pp. 1–7, 2015.

[4] A. Bain, S. Birajdar, and M. S. Manonmani, "Sign language recognition system using machine learning," International Journal of Engineering Research and Technology, vol. 8, no. 8, 2021.

[5] M. Sharma, R. Pal, and A. K. Sahoo, "Indian sign language recognition using neural networks and KNN classifiers," International Journal of Computer Applications, vol. 9, no. 8, 2014.

[6] L. Chandwani, J. Khilari, K. Gurjar, P. Maragale, A. Sonare, S. Kakade, A. Bhatt, and R. Kulkarni, "Gesture based sign language recognition system using Mediapipe," Preprint, 2023, doi: 10.21203/rs.3.rs-3106646/v1.

[7] Sujan Hiregundagal Gopal Rao. (2023). A Review of Cybersecurity Threats in Automotive Semiconductor Control Units. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(1), 927–932. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7998.

[8] K. Goyal and V. G. Velmathi, "Indian sign language recognition using Mediapipe Holistic," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2020.

[9] R. Kumar and A. Bajpai, "Mediapipe and CNNs for real-time ASL gesture recognition," in Proc. Int. Conf. Smart City and Emerging

Technology, 2021, doi: 10.13140/RG.2.2.32364.03203.

[10] A. Haldera and A. Tayade, "Real-time vernacular sign language recognition using MediaPipe and machine learning," International Journal of Research Publication and Reviews, 2021.

[11] K. M. Kavana, "Recognization of hand gestures using Mediapipe Hands," International Research Journal of Modernization in Engineering Technology and Science, vol. 4, no. 5, 2022.

[12] R. S. Rohit, R. Kandoi, and S. Kumar, "Hand gesture recognition for human computer interaction through KNN algorithm and Mediapipe," International Journal of Computer Sciences and Engineering, vol. 11, pp. 14–18, 2023.

[13] A. M. J. Al Moustafa, M. S. Mohd Rahim, B. Bouallegue, M. M. Khattab, A. M. Soliman, G. Tharwat, and A. M. Ahmed, "Integrated Mediapipe with a CNN model for Arabic sign language recognition," Journal of Electrical and Computer Engineering, vol. 2023, pp. 1–15, 2023, doi: 10.1155/2023/8870750.

[14] K. Gomase, A. Dhanawade, P. Gurav, S. Lokare, and J. Dange, "Hand gesture identification using Mediapipe," International Research Journal of Engineering and Technology, 2022.

[15] Google, "MediaPipe," 2023. https:// github.com /google/mediapipe

[16] TensorFlow, "TensorFlow," 2023. https:// www. Tensorflow.org