International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

Real-Time Marathi Sign Language Translation into Text and Speech Using Hand Gesture Recognition

Saili Bhinganiy[1], Rutuja Gunjal[2], Saburi Game[3], Shreya Gore[4], Swapnali Gawali[5]
[1,2,3,4]Student, Dept. of CSE, Sanjivani College of Engineering, Kopargaon, Maharashtra, India.
[5]Professor, Dept. of CSE, Sanjivani College of Engineering, Kopargaon, Maharashtra, India.
Email ID: *bhinganiyasaili@gmail.com[1], rutujagunjal16@gmail.com[2], gamesaburi@gmail.com[3], shreyagore1@gmail.com[4] ,gawaliswapnaliit@sanjivani.org.in[5]*

Abstract
This project aims to create a system that can translate Marathi Sign Language (MSL) into text and speech in real time using hand gesture recognition. Sign language is an essential way of communication for people who are deaf or hard of hearing, but it is not widely understood by everyone. This project helps solve this problem by using technology to recognize and translate hand gestures. The system works by using a camera to capture hand gestures. These gestures are then analyzed by a deep learning model to identify their meaning. The recognized gestures are converted into Marathi text, and this text is further converted into speech using text-to-speech technology. The system is designed to be fast, easy to use, and capable of understanding a wide range of gestures, including those with complex movements. It can also learn and improve over time by adapting to individual user styles. This project helps to make communication easier between sign language users and others, promoting better understanding and inclusivity for the hearing-impaired community in Marathi-speaking areas.

Keywords: *Marathi Sign Language (MSL), Deep Learning, Long-Short-Term Memory (LSTM), Sign Language Recognition, MSL Dataset.*

1. Introduction
In traditional communication systems, individuals who use Marathi Sign Language (MSL) face significant barriers when interacting with those who are unfamiliar with sign language. These challenges often result in limited conversations and reliance on interpreters, which can be inefficient and impractical in everyday scenarios. To address these communication barriers, we propose a real-time system that translates Marathi Sign Language hand gestures into both words and speech using advanced hand gesture recognition techniques. This system will accurately capture and interpret hand and finger movements, converting them into readable text or synthesized speech. The goal of the solution is to enable smooth interaction between non-hearing and hearing individuals, eliminating the dependence on human interpreters. Furthermore, the model will maintain a history of recognized gestures, enabling users to review and verify the translations by automating the gesture recognition and translation process, this approach enhances communication speed accuracy and accessibility for Marathi sign language users in real-world interactions. This paper is structured as follows. Section II provides a summary of current research and contributions from scholars in the relevant areas. Section III outlines the materials and methods utilized in the study, focusing on the experimental and implementation phases. Section IV discusses the analysis of the results, providing an assessment of the study's findings. Finally, Section V the paper wraps up by outlining the essential conclusions and implications of the research. [1]

2. Background and Literature Overview
The most crucial stage in the process of developing a software product is literary research. It is essential to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

establish time factors, business, and company strength before creating a tool. Once all of this has been accomplished, the subsequent 10 steps must identify which operating system and language can be utilized for development. Once a programmer develops a tool, they require substantial external assistance. This assistance can be obtained from respected programmers, publications, or online resources. Before constructing the system, the considerations mentioned previously concerning the development of the proposed system are taken into account. The tasks verified here are as follows: AmitKumar Shinde et al. [1] discussed various sign language recognition methods, focusing on regional languages using vision-based techniques, machine learning, deep learning, and technologies such as Convolutional Neural Networks (CNNs). The challenges faced during this study were variability in hand shapes, and movements and the lack of standardized datasets persist. The proposed system recognizes 43 Marathi sign language (MSL) gestures, translating them into text and vice versa. It operates in offline mode using dataset images and in real time via webcam with preprocessing and pattern recognition to ensure precision [1]. To recognize MSL, the system uses two modes, offline recognition and web camera-based recognition. In offline mode, users can learn MSL through features like pre-processing, feature extraction, and pattern recognition using datasets. The web camera mode will capture hand gestures, process them and process them through resizing, color-based detection, noise reduction, and center of gravity calculation to identify signs. This dual mode approach ensures effective learning and recognition of MSL, bridging communication gaps effectively. Future work involves improving the accuracy of the proposed system and extending dynamic hand gesture recognition along with facial expressions and aims to make the system more comprehensive and allow deaf individuals and sign language users to communicate independently without relying on interpreters [1]. Rajalakshmi et al. [2] Detections (SLRs) inspected through sign language have been significantly developed and are intended to end communication in the language community. In early research, we used the Stokoe paradigm to create basic theories and defined drawing components in the form. Research primarily focused on sensor-based approaches using infrared and depth sensors, which, while effective, often required cumbersome devices. Recent progress has shifted to visual systems where learning techniques such as folding networks Convolutional Neural Networks (CNNS) and long- term memory (LSTM) are used to extract spatial and temporal functions without technical technology [2]. Despite these advancements, challenges remain, including variability in signing styles, occlusion, and background complexity. Researchers have explored hybrid models combining manual and non-manual elements to improve accuracy. The author have demonstrated the effectiveness of attention mechanisms and multi-modal data integration in enhancing SLR systems. Overall, while progress has been made, the need for comprehensive datasets and methodologies that address diverse signing contexts persists, paving the way for more inclusive and effective SLR technologies [2]. Sunusi Bala Abdullahi et al. [3] The proposed paperwork's discusses IDF-Sign (Inconsistent Depth Features) model which is use to improve sign language recognition by addressing inconsistent depth features in hand and body motions. Sunusi Bala Abdullahi used Temporal Depth Feature Model to capture hand motions and Spatial Depth Feature Model which used to extract spatial features from video frames [3]. To identify and eliminate inconsistent features PairCFR (Pair Consistency Feature Ranking) is used along with threshold value selection for optimization. For classification Three- based models are used such as Random Forest, Rotation Forest and Optimization Forest which enhances recognition accuracy through ensemble techniques. Existing methods used 3D depth data from sensors like Microsoft Kinect and Leap Motion often struggle with dynamic signs [3]. Inconsistent depth features in sign language models lead to issues like misclassification and lower recognition accuracy. Jestin Joy et al. [4] Signal language research serves as an important tool for communication between pigeons and hard working communities. However, learning can be a challenge due to the lack of accessible resources and effective

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

training methods. Traditional learning often relies on peer groups and printed materials, which are insufficient for conveying the nuances of sign language (SL). Current tools frequently depend on expensive external sensors, limiting accessibility for many learners. Finger spelling, a key component of sign language, is used to represent letters and fill gaps where specific signs do not exist. However, its effective teaching remains under explored. Research shows that technology-enhanced learning can significantly improve sign language acquisition [4]. Yet, many existing applications do not provide real-time feedback, making independent learning difficult. Research shows that deep learning techniques (CNNS) such as folding networks can effectively classify a large number of neural networks, but they are often used for recognition primarily for educational purposes. The integration of automatic sign language recognition (ASLR) in educational frameworks illustrates a promising way to improve long-term skills, especially for learners, without access to specialized educators [4]. Sign Quiz addresses these gaps by providing an inexpensive web-based solution that allows you to learn finger writing signs in Indian Sign Language (ISL) without external support [4]. Deep R. Kothadiya et al. [5] Discussed the method of recognizing Indian Sign Languages using a Vision Trans former model. Processes photos with patches embedded. SIGN language detection includes computer vision for deep learning and gesture identification. Sign language can be divided into static and dynamic types, including manual body parts such as hands and facial expressions. Speech recognition [5]. The Vision Transformer (ViT) represents a significant advancement in image recognition, competing robustly with Convolutional Neural Networks (CNN) and surpassing them in computational efficiency and accuracy. In this framework, images are divided into fixed-size patches, which are embedded and augmented with positional encoding to enhance classification tasks. In the Vision Transformer (ViT), a model based on multi-head self-attention (MHSA) encoding features capture relationships between image patches and Multi-Layer Perceptron (MLP) to refine learned representations. In conjunction with the normalization layer, this combination improves the model, trains complex patterns, and improves the classification accuracy [5]. Jaya Prakash Sahoo et al. [6] explains the detection of human-robot interactions. HGR allows non-verbal communication between humans and robots or machines. The system is based on the compact folding model (CNN) as a DR CAM focused on recognizing hand gestures, particularly finger movements. The proposed model was tested with custom data records and well-known sign language data sets (ASL-FS) [6]. The proposed method showed superior accuracy compared to the existing methods. This model is integrated into software for the control of mobile robots in real-time [6]. Although the system is optimized for indoor use due to hardware limitations, with a depth camera operating at a distance of 1–1.5 meters. Deep R. Kothadya et al. [7] discussed the development of AI-based systems and recognized sign language. It is intended to improve communication among people with hearing impairments. Sign Language Recognition Methods It consists of two categories: globe-based and computer vision. The system is based on sensors embedded in gloves. Computer vision-based approaches leverage depth sensors, color cameras, and pose estimation techniques to detect and recognize gestures by analyzing hand movements and body positions [7]. These methods use machine learning and deep learning models such as CNN (fixed neural network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), automatic encoder, hybrid DEP models, and transmission learning [7]. Despite achieving high accuracy rates in some cases, challenges remain in handling complex gestures, reducing error rates.

3. Methodologies

This study focuses on translating Marathi Sign Language into text and speech in real-time through hand gesture recognition. We developed a sign language recognition system using a deep learning algorithm. There is a significant lack of communication between deaf and mute individuals and those who can hear and speak, as most people do not know what deaf and mute individuals are trying to communicate. To address this issue, we created a web application for the Marathi language. We utilized

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

a dataset of images depicting signs from Marathi Sign Language, which we compiled ourselves. Additionally, we processed the data to eliminate any noise. Following this, we employed a deep learning model known as LSTM to predict sign language from the images.

3.1. Marathi Sign Language Dataset

For this study, a custom dataset was developed specifically to assist in building a system for recognizing Marathi sign language. The dataset contains 450 hand gesture images, with 30 distinct hand gesture images provided for each of the 15 frequently used Marathi signs. These images were captured under diverse conditions, including varying lighting, hand orientations, and background environments, to ensure versatility and effectiveness in real-world scenarios. Each image underwent feature extraction to identify crucial elements such as hand contours, surface patterns, and overall structure. To make the images more useful and cleaner, they were processed by removing unwanted backgrounds, reducing noise, and resizing them all to the same size. The result was a refined collection of 15 distinct hand gestures, each representing a specific Marathi word, which were then employed to train and validate the sign language predicting model. [2]

3.2. Marathi Sign Language

The image below displays the signs for the Marathi alphabet, which are part of the Devanagari script employed in writing the Marathi language. Each character has its distinct shape and sound, forming the basis of reading and writing in Marathi. Learning these alphabets helps in understanding the proper pronunciation and formation of words. Visual aids like these are especially beneficial for beginners, helping them recognize and memorize the alphabet more effectively. (Figure 1)

3.3. Dataset Images

We have created our dataset consisting of 15 commonly used Marathi words, with 30 hand gesture images captured for each word. This brings the number of images in the dataset to 450. The dataset has been carefully organized and tagged to support gesture recognition tasks. Marathi language. Each character has its distinct shape (Figure 2,3,4,5) and labeling. This model learns to recognize spatial structures

Figure 1 Marathi Sign Language

Figure 2 Dataset Images 1

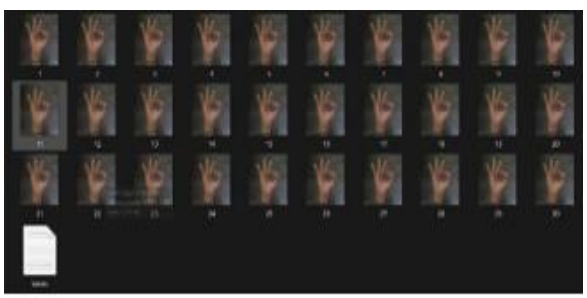

Figure 3 Dataset Images 2

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

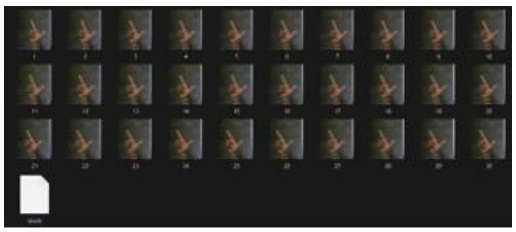

Figure 4 Dataset Images 3

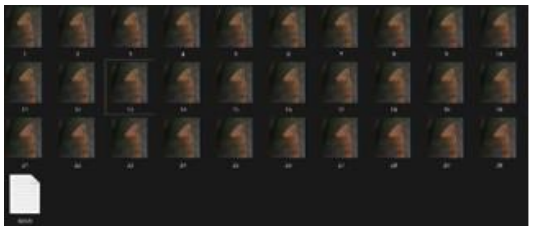

Figure 5 Dataset Images 4

3.4. System Architecture

The system architecture for real-time Marathi Sign Language Translation into Text and Speech Using Hand Gesture Recognition is shown (Figure 6)

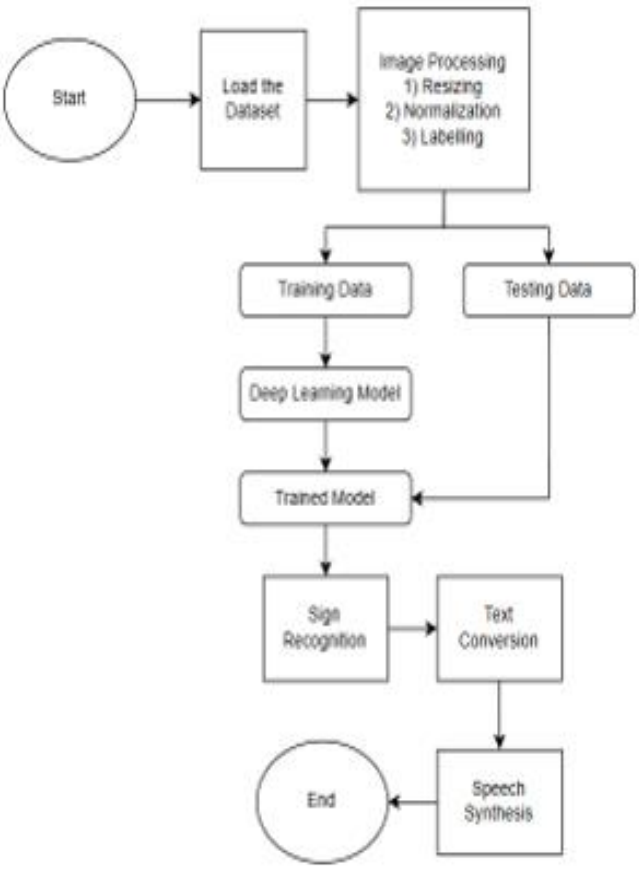

Figure 6 System Architecture

3.5. Applied Deep Learning Algorithms

Deep learning algorithms are powerful models representing complex mathematical relationships, making them ideal for image recognition tasks. After data preprocessing, including resizing, normalization, and labeling. This model learns to recognize spatial structures and visual features directly from image data. The dataset is split into training and testing subsets, where trained data helps the model learn features like edges, shapes of images, and textures. In contrast, the testing data evaluate the ability of the model to recognize and classify unseen images. In this study, we used deep learning architectures such as LSTM for this purpose, each offering different levels of accuracy based on feature extraction techniques and image quality. [3]

3.6. Long Short-Term Memory (LSTM)

LSTM is a specialized form of Recurrent Neural Network (RNN) that is built to capture patterns and relationships across longer sequences in data. Unlike basic RNNs, which may struggle with training problems like vanishing or exploding gradients, LSTMs are designed to preserve important information over time. LSTMs incorporate memory cells along with three control gates as follows - input, forget, and output which manage how information is stored, discarded, or passed through the network. This gate-based mechanism helps LSTMs retain crucial information across longer time intervals and discard irrelevant data. As a result, LSTMs are especially well-suited for handling com plex sequence-based tasks, including time-series forecasting, language understanding, and analyzing dynamic signals. In image processing tasks such as feature detection, LSTMs are frequently employed to process sequentially arranged spatial features, often extracted from Convolutional Neural Network (CNN) layers. These sequences can represent either image patches or encoded spatial-temporal structures. LSTMs analyze these sequences to detect underlying visual patterns and dependencies. One of the most vital operations within the LSTM is the cell state update, which determines how information is carried across timesteps. This dynamic memory management enables LSTMs to focus on relevant

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

features during image feature detection, which is critical in applications such as object recognition, motion analysis, and scene understanding. [4]

3.7. Evaluation Parameter

When it comes to checking how well a deep learning model is doing, we rely on several key metrics like accuracy, precision, recall, and F1-score. These are not just fancy terms they each tell us something different about the model's performance. Accuracy shows how often the model gets things right overall, but it does not always tell the full story, especially if one class appears more often than others. That is where precision and recall come in. Precision helps us understand how many of the positive predictions were correct, while recall tells us how many of the actual positives the model was able to catch. The F1-score blends these two into a single number, offering a clearer picture when we need to balance both correctness and completeness. [5]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

4. Result Analysis

In this section, we showcase the images and graphs that represent key aspects of our model's performance, such as training accuracy, training loss, and a table of evaluation metrics, along with its corresponding graph. These visuals help us track the model's learning journey, illustrating how accurately it's improving over time by minimizing errors. The evaluation metrics provide a clear understanding of the model's performance on unseen test data, helping us assess its overall effectiveness. Together, these results not only highlight the model's strengths but also point out areas where it could be fine-tuned to improve performance even further. (Figure 7,8,9,10,11,12) (Table 1) [6]

Figure 7 Result Images 1

Figure 8 Dataset Images 2

Figure 9 Dataset Images 3

Figure 10 Dataset Images 4

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

Figure 11 Dataset Images 5

Figure 12 Dataset Images 5

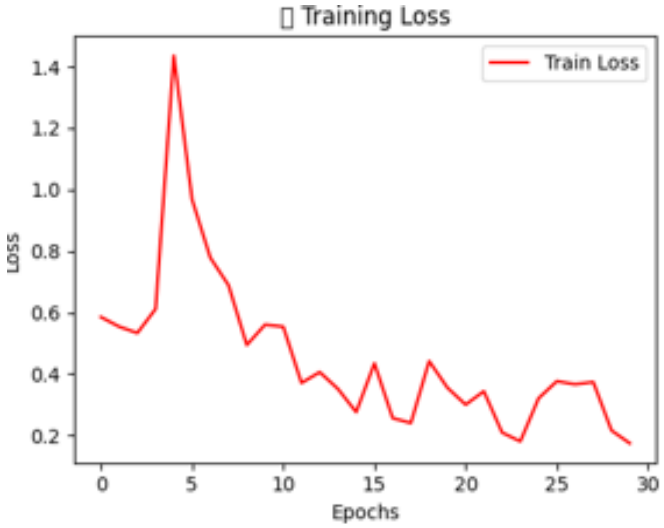

Figure 13 Dataset Images 5

Table 1 Performance Evaluation of Classification Model for 70:30 Split

Model	Accuracy	Precision	Recall	F1 Score
Long Short-Term Memory (LSTM)	90.00%	91.52%	90.00%	89.93%

Conclusion

The project represents a crucial step toward bridging communication gaps by developing a system that converts Marathi Sign Language into text and speech, empowering the people with auditory disabilities to communicate seamlessly with the hearing individuals. This solution emphasizes user friendliness, accessibility, and reliability to ensure accurate, real-time translations suitable for various applications, including daily communication, education, and professional settings. By incorporating ongoing updates and evaluations, the system will adapt to new gestures, regional dialects, and technological advancements, ensuring its relevance and effectiveness over time. Ultimately, this initiative aims to foster inclusivity, enhance accessibility, and create a transformative impact by breaking down communication barriers and strengthening con nections between diverse communities. [7]

References

[1]. Shinde and R. Kagalkar, “Advanced marathi sign language recognition using computer vision,” International Journal of Computer Applications, vol. 118, pp. 1–7, 05 2015.

[2]. E. Rajalakshmi, R. Elakkiya, V. Subramaniyaswamy, L. P. Alexey, G. Mikhail, M. Bakaev, K. Kotecha, L. A. Gabralla, and A. Abraham, “Multi-semantic discriminative feature learning for sign gesture recog nition using hybrid deep neural architecture,” IEEE Access, vol. 11, pp. 2226–2238, 2023.

[3]. S. B. Abdullahi and K. Chamnongthai, “Idf-sign: Addressing inconsistent depth features

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0306
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1946-1953

for dynamic sign word recognition," IEEE Access, vol. 11, pp. 88511–88526, 2023.

[4]. J. Joy, K. Balakrishnan, and M. Sreeraj, "Signquiz: A quiz based tool for learning fingerspelled signs in indian sign language using aslr," IEEE Access, vol. 7, pp. 28363–28371, 2019.

[5]. D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, "Signformer: Deepvision transformer for sign language recognition," IEEE Access, vol. 11, pp. 4730–4739, 2023.

[6]. J. P. Sahoo, S. P. Sahoo, S. Ari, and S. K. Patra, "Hand gesture recognition using densely connected deep residual network and channel attention module for mobile robot control," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1–11, 2023.

[7]. D. Kothadiya, C. Bhatt, H. Kharwa, and F. Albu, "Hybrid inceptionnet based enhanced architecture for isolated sign language recognition," IEEE Access, vol. PP, pp. 1–1, 01 2024