Sign Language Recognition And Translation To English And Marathi

Pratik Dahatonde
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Prathamesh Khandekar
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Omkar Kharat
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Dr. Saurabh Saoji
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Dr. Naveenkumar Jayakumar
Department of Computational Intelligence
Vellore Institute of Technology
Vellore

Abstract—Sign language serves as a vital mode of communication for the Deaf and Hard of Hearing (DHH) community, yet barriers persist in its recognition and translation. This project addresses these challenges through innovative technological solutions. The primary objective of this study is to develop a robust system for sign language recognition and translation, aiming to enhance communication accessibility for the DHH community. Leveraging advancements in machine learning and computer vision, the project seeks to overcome existing limitations and revolutionize inclusive communication technologies. The project employs a multifaceted approach, integrating Long Short-Term Memory (LSTM) networks and MediaPipe technology to accurately detect and interpret sign language gestures. Through extensive training and validation processes, the system is optimized to achieve high levels of accuracy and efficiency in real-world scenarios. The developed system demonstrates exceptional performance, achieving a recognition accuracy rate of 98% for sign language gestures. Moreover, it seamlessly translates these gestures into both spoken and written English and Marathi, offering real-time, context-aware translations. This project represents a significant advancement in addressing communication barriers for the DHH community. By providing an accessible and inclusive means of communication, the developed system has the potential to revolutionize interactions and promote equality across diverse domains. The success of this endeavor underscores the importance of leveraging technology to foster inclusivity and enhance the quality of life for individuals with hearing impairments.

Keywords–Machine Learning, MediaPipe, OpenCV, LSTM Neural Network, Sign Language, PIL, GoogleTrans

I. Introduction

Communication is the foundation of human interaction; It is easy to exchange ideas, thoughts and words. But for the deaf and hard-of-hearing (DHH) community, traditional means of communication, such as spoken language, are often a big problem. Sign language is a sign language used by the DHH community that plays an important role in communication and teaching. However, there is a lack of disclosure in society that often leaves DHH individuals vulnerable to effective communication and relationships. To address these challenges head-on, the project endeavors to harness the transformative power of technology to bridge the communication chasm between the DHH community and the hearing world. Leveraging cutting-edge advancements in machine learning and computer vision, the system aims to accurately recognize and translate sign language gestures into spoken and written English and Marathi. This innovative approach not only enhances accessibility for DHH individuals but also fosters inclusivity and understanding within society. In this project, a sophisticated blend of technologies is utilized, including Long Short-Term Memory (LSTM) networks for sign language recognition and MediaPipe technology for precise hand and gesture tracking. Additionally, to facilitate translations into Marathi, the Google Trans library is integrated, enabling seamless conversion from English to Marathi text. Through an intuitive and user-friendly interface, the system empowers DHH individuals to communicate effectively in a variety of settings, providing real-time, context-aware translations. The paper presents a comprehensive exploration of the methodology, results, and implications of the project, highlighting its potential to revolutionize communication for the DHH community and beyond. Through collaborative efforts and technological innovation, a world where communication knows no boundaries is strived for, fostering a more inclusive and equitable society for all.

II. Literature Survey

Due to the challenges faced by hearing-impaired individuals in verbal communication, sign language plays a vital role, enabling them to express themselves and integrate seamlessly into society. However, for those unfamiliar with sign language, deciphering these gestures can be difficult. The research aims to bridge this communication gap by proposing a two-way system for sign language understanding. The system leverages MediaPipe technology to extract hand landmark data, which is then fed into a Long Short-Term Memory (LSTM) network for gesture recognition in Indian Sign Language (ISL). Additionally, the system incorporates speech recognition capabilities. It can capture audio in English or Hindi using a microphone and process it using Google Speech API. Alternatively, users can type text, and the system will convert it into corresponding ISL video demonstrations. This two-way approach, combining sign language recognition and text/speech conversion, has the potential to break down communication barriers and foster better interaction between hearing-impaired individuals and others.[1]

The paper tackles the challenge of communication between deaf/hearing-impaired and regular individuals by proposing a real-time sign language translation system. The system relies on deep learning for vision-based recognition, translating signs into text. To achieve this real-time functionality, the approach utilizes You Only Look Once version 4 (YOLOv4), a detector known for its speed. The researchers leverage the Indian Sign Language Dataset for Continuous Sign Language Translation and Recognition (ISL-CSLTR) for experimentation. They further enhance this dataset by incorporating signs from two additional signers, expanding the training data. To improve accuracy, the system employs a pre-trained network that is fine-tuned on the extended dataset. Additionally, data augmentation techniques are used to bolster the system's robustness. The proposed method achieves impressive results with a mean average precision (mAP) of 98.4%, demonstrating real-time translation capabilities. This high accuracy paves the way for real-world applications, which was the primary goal of this research.[2]

Sign language, a complex communication system for deaf individuals utilizing hand gestures, facial expressions, and body language, mirrors spoken languages in complexity but employs a distinct structure. While traditional sign language recognition relies on cameras, these methods suffer from limitations: sensitivity to poor lighting, difficulty handling lengthy video sequences for training, and significant privacy concerns. The study proposes a groundbreaking approach – a contactless and privacy-preserving system for recognizing British Sign Language (BSL) emotions that leverages radar technology. As a proof of concept, the research focuses on identifying six key emotions: confused, depressed, happy, hate, lonely, and sad. The captured radar data is transformed into spectrograms, visualizations depicting signal strength across time and frequency. These spectrograms are then analyzed by three cutting-edge deep learning models – InceptionV3, VGG19, and VGG16 – to extract relevant features. Finally, the system classifies the emotions by categorizing the spectrograms based on the extracted features. The study achieves impressive results, with the VGG16 model demonstrating a maximum classification accuracy of 93.33%. The research presents a promising alternative to camera-based recognition, prioritizing user privacy. While the current focus on emotions lays the foundation for broader BSL recognition in the future, further investigation is necessary to explore the system's ability to identify a wider range of emotions and its effectiveness in real-world applications.[3]

Vision-based sign language recognition (SLR) systems bridge the communication gap between deaf and hearing communities by recognizing signs from video. The review explores advancements in SLR, focusing on techniques for isolating hands (segmentation), capturing sign characteristics (feature extraction), and classifying the signs. Early segmentation methods relied on skin color, but struggled with lighting variations. Deep learning, particularly Convolutional Neural Networks (CNNs), addresses this challenge by learning complex features directly from video data. Feature extraction encompasses geometric (hand shape), motion (movement patterns), and appearance (visual properties) aspects. Geometric features might include fingertip locations and palm orientation, while motion features could be captured by Hidden Markov Models (HMMs). Appearance features might be extracted using wavelet packet decomposition. Traditionally, Support Vector Machines (SVMs) were used for classification due to their effectiveness with complex data. Conditional Random Fields (CRFs) improve accuracy for signs with similar shapes by considering dependencies between signs. Deep Neural Networks (DNNs) like CNNs achieve state-of-the-art performance by learning features and classifying signs simultaneously. The review concludes that deep learning has significantly advanced SLR, with future directions including handling signing variations, achieving real-time recognition, and incorporating facial expressions and body language for a more holistic understanding of sign language.[4]

Recognizing the vital role of sign language in India, especially after the Rights of Persons with Disabilities Act (RPwD Act) mandated its use, researchers are actively developing methods to improve communication accessibility for the deaf and mute community. The study proposes a novel approach for Indian Sign Language (ISL) recognition using deep learning.The system focuses specifically on recognizing static signs, which represent individual alphabets in ISL. Unlike some existing methods, it doesn't require complex hand tracking or pose estimation. Instead, it analyzes the simpler data of hand silhouettes extracted from images. To achieve this, the researchers leverage the power of Convolutional Neural Networks (CNNs), a type of deep learning architecture adept at recognizing patterns in visual data.The approach involves a two-pronged strategy. First, a comprehensive review of existing sign language recognition techniques is conducted to understand the current landscape and potential areas for improvement. Second, the authors build a dataset specifically tailored for training and testing their CNN model. The dataset likely consists of images containing hand silhouettes representing different ISL alphabets.The paper delves into the details of the training and testing phases of the CNN model. Training involves feeding the network with a large number of labeled images (hand silhouettes with corresponding ISL alphabets) to help it learn the distinctive features of each sign. During testing, the trained model is presented with unseen images, and its ability to accurately identify the corresponding ISL alphabet is evaluated.The encouraging outcome of this research lies in the achieved accuracy of 98.64%, surpassing most existing methods for ISL static alphabet recognition. This signifies a significant leap forward in the development of reliable sign language recognition technology. With further refinement and adaptation, such systems have the potential to bridge the communication gap between the hearing and deaf communities, fostering greater social inclusion for deaf and mute individuals.[5]

Sign language plays a crucial role in enabling communication for deaf and speech-impaired individuals. This research tackles the challenge of real-time communication by proposing a system specifically designed to recognize gestures from Indian Sign Language (ISL). The system relies on computer vision techniques to bridge the communication gap. It employs a method called skin segmentation to identify and track the most relevant parts of the image, which are most likely the signer’s hands (Regions of Interest or ROI). Once these ROIs are identified, the system utilizes a machine learning algorithm known as fuzzy c-means clustering to classify the captured gestures. The approach offers exciting possibilities beyond its core function of sign language interpretation. The ability to recognize gestures in real-time opens doors for various applications, including controlling robots and smart home devices through hand gestures. It can even be used for gesture-based video game control, making the gaming experience more inclusive. Furthermore, the system’s potential for Human-Computer Interaction (HCI) is significant. By facilitating communication between deaf and hearing individuals, this system represents a major leap forward in fostering a more inclusive society. The real-time nature of the system makes it particularly valuable, as it allows for seamless and spontaneous communication. This research holds significant promise for improving the quality of life for deaf and speech-impaired individuals by empowering them to interact more freely with the world around them.[6]

Sign language serves as a vital communication tool for deaf individuals. However, its limited popularity creates a social barrier. The research aims to bridge this gap by proposing a Human-Computer Interaction (HCI) system specifically designed for sign language recognition. Traditionally, sign language recognition research relied on machine learning methods that required significant manual configuration and often lacked adaptability. The work introduces a new approach utilizing Residual Neural Networks (RNNs), a type of deep learning architecture. RNNs offer the advantage of end-to-end recognition, potentially streamlining the process and reducing the need for manual intervention.The system focuses on recognizing static signs from American Sign Language (ASL), including numbers and alphabets. To address potential limitations in training data size, a common challenge in machine learning, the authors implemented data augmentation techniques. These techniques essentially create additional training examples from existing data, enriching the dataset and potentially improving the model’s generalizability. The proposed system achieved a remarkable accuracy of 99.4%, indicating its effectiveness in ASL recognition.The research holds significant promise for the deaf community. By offering a more robust and adaptable approach compared to traditional methods, the system paves the way for more inclusive HCI systems. This can significantly improve communication accessibility for deaf individuals, fostering greater social integration and participation. The real-world applications of such a system extend beyond basic communication, potentially impacting various fields like education and employment opportunities for the deaf population.[7]

Researchers are actively developing computer vision techniques to bridge the communication gap between deaf and hearing individuals through sign language recognition. However, current approaches often face limitations in terms of compatibility. Existing methods might require specific conditions or controlled environments to function accurately. One significant challenge lies in hand segmentation, the process of isolating hand gestures from the face and background in an image. Skin detection algorithms, commonly used for this purpose, can struggle with complexities like overlapping hands or backgrounds with similar skin tones.The passage acknowledges recent progress in sign language recognition with the use of active sensors like depth cameras. These sensors provide more precise hand tracking data, leading to improved recognition accuracy. However, there’s a push for advancements in markerless passive sensors, particularly regular cameras. This would eliminate the need for people to wear special markers while using sign language, making the technology more user-friendly and widely applicable.The core of the issue lies in feature extraction, the process of identifying and extracting key characteristics from sign language gestures that allow the computer to recognize them. By focusing on improving feature extraction techniques specifically for passive camera data, researchers aim to create more robust and adaptable sign language recognition systems. This would not only enhance communication accessibility but also pave the way for broader integration of sign language recognition technology into everyday life.[8]

Sign language recognition has become a prominent area of research in computer vision, particularly for facilitating Human-Computer Interaction (HCI). The study proposes a novel approach that departs from traditional methods. Instead of relying on skin detection, the system focuses on hand shape analysis using Zernike Moments, a mathematical tool adept at describing object outlines.The system operates in a series of well-defined stages. The first stage tackles gesture segmentation, isolating relevant movements by analyzing motion patterns in the video stream. Next, it performs real-time detection of both the signer's hands and their face. This ensures the system focuses on the most critical regions of interest. To optimize processing efficiency, the system then extracts only key frames from the video, eliminating redundant information.The core analysis phase involves two key aspects: tracking hand movements and determining hand postures. The system tracks the trajectory of the hands relative to the face, providing valuable information about the signs being conveyed. For hand posture analysis, the system leverages the power of Zernike Moments. The technique offers a significant advantage – it remains insensitive to hand rotations, ensuring accurate recognition even when hand orientation varies slightly. Finally, the system employs a method called Dynamic Time Warping (DTW) to compare the extracted features and recognize the specific gestures being signed.This research presents a unique contribution to the field of sign language recognition. By utilizing Zernike Moments for robust hand shape analysis, the system offers a promising alternative to traditional approaches. The method holds potential for broader applications beyond HCI, potentially impacting fields like robotics and human-computer interaction in augmented or virtual reality environments. The real-time nature of the system further enhances its value, fostering more natural and seamless communication between humans and computers.[9]

Language differences create significant hurdles in accessing information. Imagine a Marathi speaker wanting to learn about English history or a foreign scholar researching Shivaji – both face challenges due to language barriers. This limited access to information can extend to crucial local news, government documents, and cultural knowledge. To bridge this gap, machine translation offers a powerful tool, translating text, speech, and even images to facilitate information sharing and learning across languages. While significant research has addressed translation between English and major Indian languages like Hindi and Tamil, machine translation remains a complex field due to inherent language variations in grammar, structure, and achieving natural-sounding fluency. The research proposes a novel hybrid machine translation system specifically for translating English to Marathi. The system tackles various document types, including web pages, agricultural texts (beneficial for farmers), medical reports, and tourism information. The proposed system's strength lies in its parallel multi-engine approach, combining the strengths of both statistical and rule-based translation methods. By prioritizing the fluency of statistical translation while ensuring accuracy through rule-based checks, the system aims to achieve optimal results. To further enhance translation quality, the system incorporates a mapper algorithm within the rule-based translation, leverages domain-specific corpora (agriculture, medical, tourism) for statistical evaluation, and utilizes a Marathi wordnet to expand its dictionary and improve translations. While currently designed for text documents, the system has the potential for future expansion into speech translation. The researchers compare their system with Google Translate on a limited set of queries, suggesting their hybrid approach offers superior translation quality. However, the study itself has limitations, including a review of only 10 key research articles and building upon the authors' prior work. Additionally, while an innovative S-measure is proposed for evaluation, details about its functionality are not provided. Overall, the research presents a promising hybrid machine translation system for English to Marathi, with potential applications across various information domains, promoting cross-cultural understanding and knowledge sharing.[10]

III. Methodology

The methodology of Sign Language Recognition and Translation to English and Marathi is implemented in the following steps:

1. Data Collection: Collect a comprehensive dataset comprising video sequences capturing a diverse range of sign language gestures, ensuring adequate representation of different gestures and variations. Employ standardized labeling procedures to assign unique identifiers to each gesture, facilitating data organization and preparation for training and testing phases.
2. Data Preparation: Format the dataset into structured arrays, with each video frame represented as a multidimensional NumPy array (X) and its corresponding label as another NumPy array (Y), ensuring compatibility with machine learning algorithms.
3. Implementation Procedure: Utilize advanced image processing techniques to extract relevant features from video frames, leveraging the Holistic Pipeline and Long Short-term Memory (LSTM) network to capture spatial and temporal patterns in sign language gestures. Employ the MediaPipe solution by Google and Python's OpenCV library to ensure accurate landmark identification of users' hands and body parts, enhancing the robustness and accuracy of the feature extraction process.
4. Model Generation and Model Testing: Develop and train the sign language recognition model using the prepared dataset, optimizing model architecture and hyperparameters to maximize accuracy and generalization performance. Conduct rigorous model testing and validation procedures, including cross-validation and evaluation on independent test sets, to assess the model's performance under various conditions and scenarios.
5. Translation Integration: Integrate the Google Translate API or similar translation services to enable real-time trans-

lation of recognized sign language gestures into Marathi. Develop algorithms to preprocess translated text and ensure its alignment with sign language gestures, preserving semantic meaning and linguistic accuracy during translation.

6. User Interface Development: Design and implement user-friendly interfaces tailored to the needs of diverse user groups, including Deaf and Hard of Hearing (DHH) individuals and users with normal hearing. Prioritize accessibility features such as gesture visualization, intuitive controls, and customizable settings to enhance usability and ensure inclusive communication experiences for all users.

7. Usability Evaluation: Conduct usability testing sessions with representative user groups, soliciting feedback and insights to identify usability issues, user preferences, and areas for improvement. Iteratively refine the user interface based on user feedback, incorporating design enhancements and usability optimizations to enhance user satisfaction and engagement.

8. Continuous Improvement: Implement a scalable architecture that allows for the seamless addition of new sign language gestures as needed, ensuring the system's adaptability to evolving user needs and linguistic diversity. Develop intuitive tools and interfaces for gesture annotation and dataset expansion, enabling users to contribute new gestures easily and efficiently.

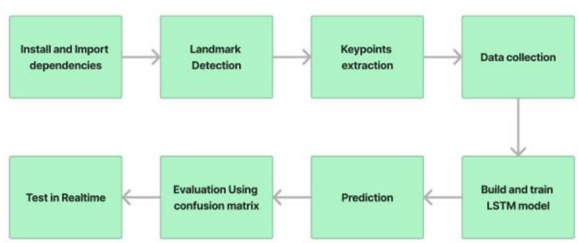

Fig. 1. Block Diagram of Dynamic ISL Gesture Recognition

IV. Proposed Methodology/Project Implementation

A. Dataset Creation and Pre-Processing

Each dynamic sign language gesture was meticulously captured on film 20 times to construct the dataset utilized in this investigation, ensuring comprehensive coverage and variability in gesture representation. Every video sequence comprises 30 frames, capturing the complete motion of the sign gesture. In cases where fewer frames were required for a gesture, zero-padding was applied to extend the sequence to 30 frames, maintaining consistency across the dataset.

B. Feature Extraction

Feature extraction was performed to capture the essential characteristics of dynamic sign gestures, focusing on specific landmarks indicative of both the hands and pose movements. The Mediapipe framework was employed to identify key landmarks for both right and left hands which contain 21 landmarks, as well as pose markers which contain 33 landmarks, resulting in a total of 75 feature landmarks extracted per frame. Utilizing the Holistic model, which integrates hand and pose ensured comprehensive coverage of relevant features for dynamic gesture analysis. Features extracted from each frame were stored in NumPy array files, representing three-dimensional coordinates (x, y, z) for each landmark. Consequently, each NumPy array file encapsulated a total of 258 features.

C. Training the Model

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 30, 64)            82688

 lstm_1 (LSTM)               (None, 30, 128)           98816

 lstm_2 (LSTM)               (None, 64)                49408

 dense (Dense)               (None, 64)                4160

 dense_1 (Dense)             (None, 32)                2080

 dense_2 (Dense)             (None, 12)                396

=================================================================
Total params: 237548 (927.92 KB)
Trainable params: 237548 (927.92 KB)
Non-trainable params: 0 (0.00 Byte)
```

Fig. 2. Model Summary

The preprocessed dataset was partitioned into training and validation sets based on predefined criteria. The model, as depicted in Figure 2, was fine-tuned using hyper-parameters specified in Table I, including sequence length, LSTM layers and neurons per layer, dense layers and neurons per layer, activation functions, optimizer, batch size, and number of epochs. Training involved optimizing the model's parameters and architecture to effectively learn and discriminate between different sign language gestures. Figure 2 provides a visual representation of the model's architecture, illustrating the sequential flow of data through its layers and stages, thereby enhancing understanding of its functionality.

TABLE I
Model Hyper-Parameters

Hyper-Parameters	Values
Training Data	90% (216 videos)
Testing Data	10% (24 videos)
Sequence Length	30
LSTM Layers + Neurons Per Layers	3 Layers 64 neurons 128 neurons 64 neurons
Dense Layers + Neurons per layers	3 Layers 64 neurons 32 neurons 24 neurons
Activation Function - input and hidden layers	Relu
Activation Function - Output layer	Softmax
Optimizer	Adam
Batch Size	64
Epoch	550

D. Model Evaluation

The model's performance was assessed using evaluation metrics, including accuracy, precision, recall, and F1-score, with the validation set. Hyper-parameters and network architecture were adjusted iteratively to optimize performance, ensuring robustness and generalization capabilities of the model. Additionally, the graph Figure 3 below illustrates the accuracy and loss during training on both the training and testing datasets, providing insights into the model's learning progress and performance stability over epochs.

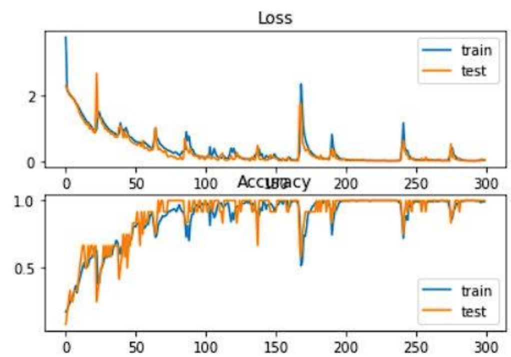

Fig. 3. Accuracy and Loss during training

E. Translation

Integrating the Google Translate API to facilitate dynamic translation of recognized sign language gestures into Marathi, ensuring accurate and real-time linguistic conversions to enhance accessibility for Marathi-speaking users. Utilizing PIL (Python Imaging Library) to seamlessly integrate translated Marathi text into the user interface, allowing for the simultaneous display of both English and Marathi translations alongside the recognized sign language gestures, enhancing user comprehension and inclusivity.

F. Real-Time Testing

The trained model was deployed in real-time scenarios to evaluate its performance in recognizing dynamic sign gestures. Real-world testing enabled the validation of the model's effectiveness and reliability in practical applications, facilitating further refinement and improvement based on observed performance.

V. Results And Analysis

The Sign Language Recognition system developed for dynamic gestures achieved remarkable performance, demonstrating its effectiveness in identifying Indian Sign Language (ISL) gestures.

A. Training Accuracy and Epochs

The SLR system attained a training accuracy of 98.6% within 550 epochs. It was observed that the model's accuracy plateaued after 550 epochs, indicating optimal performance. Batch size experimentation revealed that a batch size of 64 resulted in superior performance compared to smaller batch size (32), highlighting its efficiency in training the model.

B. Validation Accuracy

Testing the model on the validation dataset yielded an accuracy of more than 99%, affirming the model's robustness and generalization capability. The confusion matrix depicted in Figure 4 provided insights into the model's performance for each dynamic gesture, while precision, recall, and F1-score metrics (as shown in Figure 5) quantified the model's performance for individual gestures.

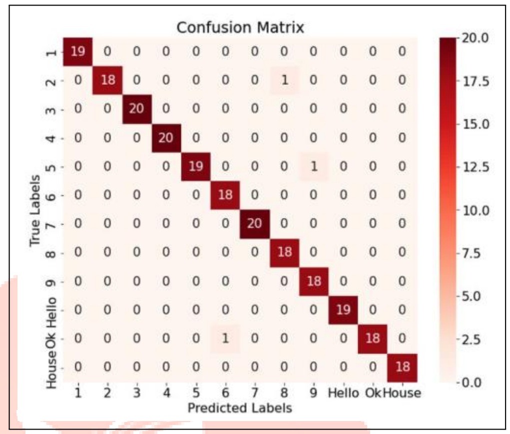

Fig. 4. Confusion Matrix

C. Average Metrics

The average precision, recall, and F1-score for the validation set were calculated to be 99%, 99%, and 99%, respectively as shown in Figure 5, underscoring the overall effectiveness of the Sign Language Recognition system in recognizing dynamic ISL gestures.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	19
2	1.00	0.95	0.97	19
3	1.00	1.00	1.00	20
4	1.00	1.00	1.00	20
5	1.00	0.95	0.97	20
6	0.95	1.00	0.97	18
7	1.00	1.00	1.00	20
8	0.95	1.00	0.97	18
9	0.95	1.00	0.97	18
Hello	1.00	1.00	1.00	19
Ok	1.00	0.95	0.97	19
House	1.00	1.00	1.00	18
accuracy			0.99	228
macro avg	0.99	0.99	0.99	228
weighted avg	0.99	0.99	0.99	228

Fig. 5. Model Metrics

D. Real-Time Testing

Real-time testing of the model using a standard laptop camera demonstrated its practical applicability. In addition to accurately recognizing dynamic signs, the model seamlessly translated the recognized gestures into Marathi in real-time, enhancing accessibility for Marathi-speaking individuals. However, occasional false positives were observed during transitions between gestures, likely due to the camera's continuous detection of changes. This indicates a potential area for future research, particularly in refining successive dynamic sign recognition techniques. The images below provide visual examples of the real-time testing process and demonstrate the effectiveness of the Sign Language Recognition system in recognizing and translating dynamic sign gestures. Figure 6 illustrates the recognition of the sign for 'Hello' and its translation into Marathi, while Figure 7 showcases the recognition of the sign for 'House' and its Marathi translation. Furthermore, Figure 8 and Figure 9 depict the recognition of the numbers 1 and 5, respectively.

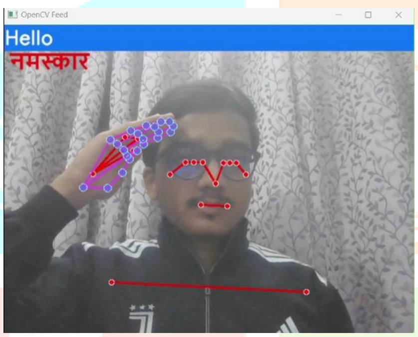

Fig. 6. Hello Gesture

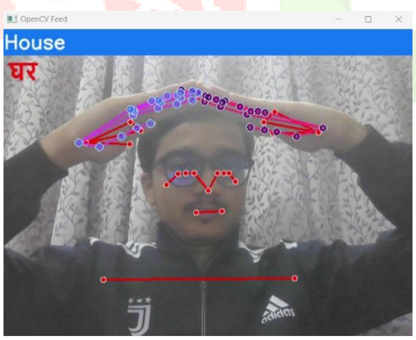

Fig. 7. House Gesture

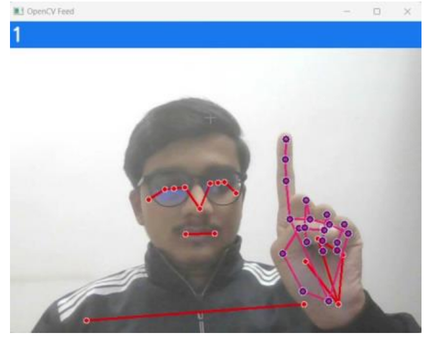

Fig. 8. Number 1 Gesture

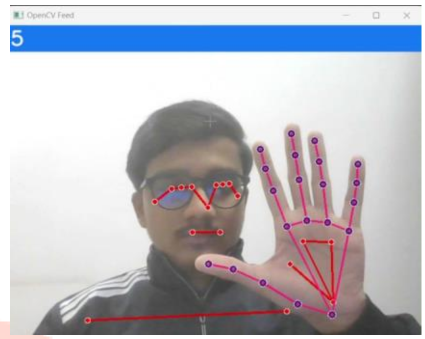

Fig. 9. Number 5 Gesture

These findings showcase the potential of Sign Language Recognition systems for dynamic gestures in ISL identification, laying the groundwork for further advancements in this field. The Sign Language Recognition system's robust performance in both offline validation and real-time testing scenarios underscores its viability for practical applications in facilitating communication for individuals with hearing impairments.

VI. Conclusion And Future Scope

The development and evaluation of the Sign Language Recognition system for dynamic gestures signify a significant advancement in facilitating communication for individuals with hearing impairments. Through meticulous dataset creation, feature extraction, and model training, the Sign Language Recognition system demonstrated exceptional accuracy and robustness in recognizing Indian Sign Language (ISL) gestures in real-time scenarios. The achievement of a training accuracy of 98.6% within 550 epochs underscores the effectiveness of the model architecture and training methodology. Furthermore, real-time testing of the model using a standard laptop camera highlighted its practical applicability, with seamless translation of recognized gestures into Marathi enhancing accessibility for Marathi-speaking individuals.

However, the occasional occurrence of false positives during transitions between gestures points towards areas for improvement, particularly in refining successive dynamic sign recognition techniques. Future research endeavors could focus on enhancing the model's robustness to transitions and exploring novel approaches for dynamic sign recognition. Additionally, there is a scope for expanding the dataset to include a wider variety of sign gestures, encompassing regional variations and complex gestures, to further enhance the model's versatility and inclusivity.

Moreover, the integration of advanced techniques such as deep learning architectures, multimodal fusion, and attention mechanisms could potentially improve the Sign Language Recognition system's performance and address existing challenges. Collaborative efforts with sign language experts, linguists, and stakeholders from the deaf and hard of hearing community can enrich the development process, ensuring the model's effectiveness and relevance in real-world settings.

In conclusion, the successful implementation and evaluation of the Sign Language Recognition system lay a strong foundation for future advancements in sign language recognition technology, fostering inclusivity and accessibility for individuals with hearing impairments.

ACKNOWLEDGMENT

The authors wish to thank their parents and associates for their valuable support and encouragement throughout the development of the project work, and they would also like to thank their guide Dr. Saurabh Saoji for guiding them throughout the project work.

REFERENCES

[1] T. Nehra, D. Saisanthiya and A. Modi, "Indian Sign Language (ISL) Recognition And Translation using Mediapipe and LSTM," 2023 World Conference on Communication Computing (WCONF), RAIPUR, India, 2023, pp. 1-5.

[2] S. Sharma, R. Sreemathy, M. Turuk, J. Jagdale and S. Khurana, "Real-Time Word Level Sign Language Recognition Using YOLOv4," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-7.

[3] H. Hameed et al., "Privacy-Preserving British Sign Language Recognition Using Deep Learning," 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 4316-4319.

[4] S. Sharma and S. Singh, "Vision-based sign language recognition system: A Comprehensive Review," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 140-144.

[5] S. C.J. and L. A., "Signet: A Deep Learning based Indian Sign Language Recognition System," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0596-0600.

[6] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6.

[7] M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1629-1633.

[8] Suharjito, F. Wiryana, G. P. Kusuma and A. Zahra, "Feature Extraction Methods in Sign Language Recognition System: A Literature Review," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), Jakarta, Indonesia, 2018, pp. 11-15.

[9] S. Mathur and P. Sharma, "Sign Language Gesture Recognition using Zernike Moments and DTW," 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2018, pp. 586-591.

[10] P. Salunkhe, A. D. Kadam, S. Joshi, S. Patil, D. Thakore and S. Jadhav, "Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator)," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 2016, pp. 924-931.

[11] M. S. Alom, M. J. Hasan and M. F. Wahid, "Digit Recognition in Sign Language Based on Convolutional Neural Network and Support Vector Machine," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-5.

[12] H. Bhavsar and J. Trivedi, "Hand Gesture Recognition for Indian Sign Language using Skin Color Detection and Correlation-Coefficient algorithm with Neuro-Fuzzy Approach," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 2019, pp. 1-5.

[13] Y. Yan, Z. Li, Q. Tao, C. Liu and R. Zhang, "Research on Dynamic Sign Language Algorithm Based on Sign Language Trajectory and Key Frame Extraction," 2019 IEEE 2nd International Conference on Electronics Technology (ICET), Chengdu, China, 2019, pp. 509-514.

[14] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899.

[15] S. Lahoti, S. Kayal, S. Kumbhare, I. Suradkar and V. Pawar, "Android Based American Sign Language Recognition System with Skin Segmentation and SVM," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-6.