Multilingual Translation of Hand Gestures using Machine Learning

Vaishnavi Karanjkar[1], Rutuja Bagul[2], Rushali Shirke[3], Raj Ranjan Singh[4],

Prof. Pallavi Bhaskare[5]

[1]Department of Computer Engineering, STE'S Smt. Kashibai Navale College of Engineering
[2]Department of Computer Engineering, STE'S Smt. Kashibai Navale College of Engineering
[3]Department of Computer Engineering, STE'S Smt. Kashibai Navale College of Engineering
[4] Department of Computer Engineering, STE'S Smt. Kashibai Navale College of Engineering
[5] Professor, Department of Computer Engineering, STE'S Smt. Kashibai Navale College of Engineering

---***---

Abstract - Sign Language is mainly used by deaf (hard hearing) and dumb people to exchange information between their own community and with other people. It is a language where people use their hand gestures to communicate as they can't speak or hear. The goal of sign language recognition (SLR) is to identify acquired hand motions and to continue until related hand gestures are translated into text. Here, static and dynamic hand gestures for sign language can be distinguished. The human community values both types of recognition, even if static hand gesture recognition is easier than dynamic hand gesture recognition. By creating Deep Neural Network designs where the model will learn to detect the hand motions images throughout an epoch, we are using Deep Learning Computer Vision to recognize the hand gestures. After the model successfully recognizes the motion, an English text file is created that can subsequently be translated to another language. The user can choose from a variety of translations for this paragraph. This application can be used without an internet connection and is entirely offline. With this model's improved efficiency, communication will be easier for the deaf (hard of hearing) and disabled people. We shall discuss the use of deep learning for sign language recognition in this paper.

***Key Words*:** sign language, deep neural network, computer vision, hand gesture.

1. INTRODUCTION

The application of Sign language to multilingual text and is an innovative nexus of technology, linguistic accessibility, and inclusivity is created by the incorporation of sign language into multilingual text. For Deaf and hard of hearing people, sign language is a crucial means of communication that opens up the outside world to them. However, this particular language has frequently encountered difficulties when dealing with spoken and written languages, posing obstacles in daily life and the worlds of education and the workplace.

Using a camera-equipped device, the application procedure starts with the capture of sign language gestures. An advanced backend model created to convert these movements into text uses this image as its input. With the use of cutting-edge methods like MediaPipe framework and Long Short-Term Memory (LSTM) networks, this backend model processes the acquired image with amazing accuracy and efficiency.

This system has the capacity to generate datasets on its own to improve recognition precision and vocabulary coverage. At the moment, the dataset includes frequently used sign language terms including "hello," "thanks," "welcome," "sorry," "foolish," "yes," "I love you," and "deaf." The system's self-learning feature guarantees ongoing enhancement and adjustment to the changing communication requirements of its users.

MediaPipe, a versatile framework for building multimodal applied machine learning pipelines, plays a pivotal role in the initial stages of processing. It is utilized for hand and gesture detection, allowing the system to accurately identify and track the movements of the signer's hands. This step is crucial for understanding the intricate nuances of sign language communication.

Concurrently, the backend model's LSTM network interprets the hand gestures that have been identified. LSTM networks are especially well-suited for sign language identification because of their reputation for being able to represent sequential input and grasp temporal connections. The intended message encoded in the sign language gestures is decoded by the LSTM network by examining the sequence of hand movements over time.

The LSTM network produces matching textual output as it interprets the sign language actions. This product, which was originally given in English, is translated into many Indian communication languages such as Hindi, Marathi, Gujarati and Urdu in order to serve a wider audience. Selected Indian communication languages are included in the translation pipeline, ensuring accessibility for users across different linguistic backgrounds within the Indian context.

The user is presented with the translated text output in real-time, and it is presently available in both English and a few Indian communication languages. Users of sign language and non-sign language speakers can communicate easily thanks to this text output, which can be sent to other devices or applications or shown on the device's screen. The backend approach efficiently bridges the gap between sign language and text by combining LSTM networks with MediaPipe and introducing multilingual translation capabilities, improving accessibility and inclusivity for all users. The incorporation of these cutting-edge technologies highlights the system's capacity to provide precise translations in real-time, promoting smooth communication interactions. The purpose of this work is to add to the expanding corpus of research on

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

improving accessibility and empowering people with different communication needs.

2. PROPOSED METHODOLOGY

In the application workflow, the process of capturing, processing, and recognizing hand gestures is further detailed. The CameraX API facilitates the capture of frames, which are saved in the device's internal storage for further processing. Each frame undergoes a series of transformations, including conversion to bitmap format, compression into a byte array output stream, and encoding into base64 for efficient transfer to the Flask backend. Upon decoding in the backend, the frame data is reconstructed into a NumPy array and processed using OpenCV to apply the `mediapipe_detection` function. This function extracts the key points representing the hand gestures, providing the necessary input for the LSTM model. The LSTM model then predicts the gesture label, which is saved in a variable in Java for application use.

Overall, this approach enables real-time hand gesture recognition on mobile devices, offering a seamless and intuitive user experience. By leveraging LSTM networks, the model can effectively capture the temporal dynamics inherent in hand movements, enhancing the accuracy of gesture recognition. The integration of technologies such as CameraX, Mediapipe, and Flask backend facilitates the efficient capture, processing, and transmission of frame data, enabling robust hand gesture recognition within the application. This comprehensive workflow underscores the potential of deep learning techniques in enhancing human-computer interaction, particularly in domains such as sign language interpretation where precise gesture recognition is paramount.

A key component of the sign language recognition system in the proposed research article is the application of the MediaPipe Holistic model, a state-of-the-art framework for thorough human pose estimation. This model provides complex point combinations that are necessary for accurately recognizing both hand and full-body gestures. The MediaPipe Holistic model specifically identifies three crucial configurations:

mediapipe_holistic.LEFT_HAND_CONNECTIONS,
mediapipe_holistic.RIGHT_HAND_CONNECTIONS,
mediapipe_holistic.POSE_CONNECTIONS.

The LEFT_HAND_CONNECTIONS and RIGHT_HAND_CONNECTIONS configurations encompass 63 points each, delineating the connections and coordinates of significant landmarks on the signer's left and right hands, respectively. These points serve as fundamental components in tracking and interpreting the intricate hand movements inherent in sign language communication. Additionally, the POSE_CONNECTIONS configuration comprises 132 points that delineate the connections and coordinates of various landmarks spanning the human body.

These points provide vital contextual information for accurate interpretation of sign language, as they collectively capture the holistic pose and movement dynamics of the signer. By utilizing the predetermined point configurations offered by the MediaPipe Holistic model, the suggested system manages to accomplish reliable real-time analysis and translation of gestural signals into textual output, consequently promoting improved inclusivity and accessibility in communication.

Fig -1: Mediapipe Key points connection

3. MATHEMATICAL MODEL

The LSTM (Long Short-Term Memory) model structure described operates by processing sequences of 30 consecutive frames, each containing 63 key points representing hand gestures. This structured input allows the model to capture temporal dependencies within the gesture sequence, crucial for accurately recognizing complex hand movements. The model's operation begins with real-time frame capture through the phone's camera, displaying the feed on the screen for visualization.

Using the Mediapipe Library, the key points of hand gestures are extracted from each frame and appended to a list. Once this list reaches a size of 30 frames, indicating a complete gesture sequence, it is sent as input to the LSTM model. By learning from the sequential data, the LSTM model can effectively interpret the nuances of hand movements and generate an output label corresponding to the recognized gesture.

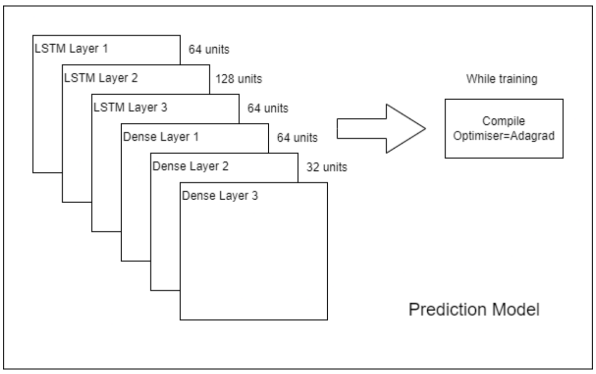

Fig -2: Mathematical Model

4. RESULTS

The proposed method's performance evaluation involves an extensive analysis on the efficacy of the suggested sign language recognition system, which combines cutting edge technologies like MediaPipe and LSTM networks. We analyze in detail the system's capacity to translate sign language gestures captured by image input into precise textual output, including translations into some Indian communication languages. The system performs remarkably well in interpreting a range of sign language gestures by utilizing MediaPipe for accurate hand and gesture detection and LSTM networks for sequential data processing. Moreover, adding self-generated datasets with frequently used sign language terms improves the system's vocabulary coverage and recognition accuracy. The results of extensive testing and analysis confirm that the system is effective in enabling sign language users to communicate easily across linguistic barriers.

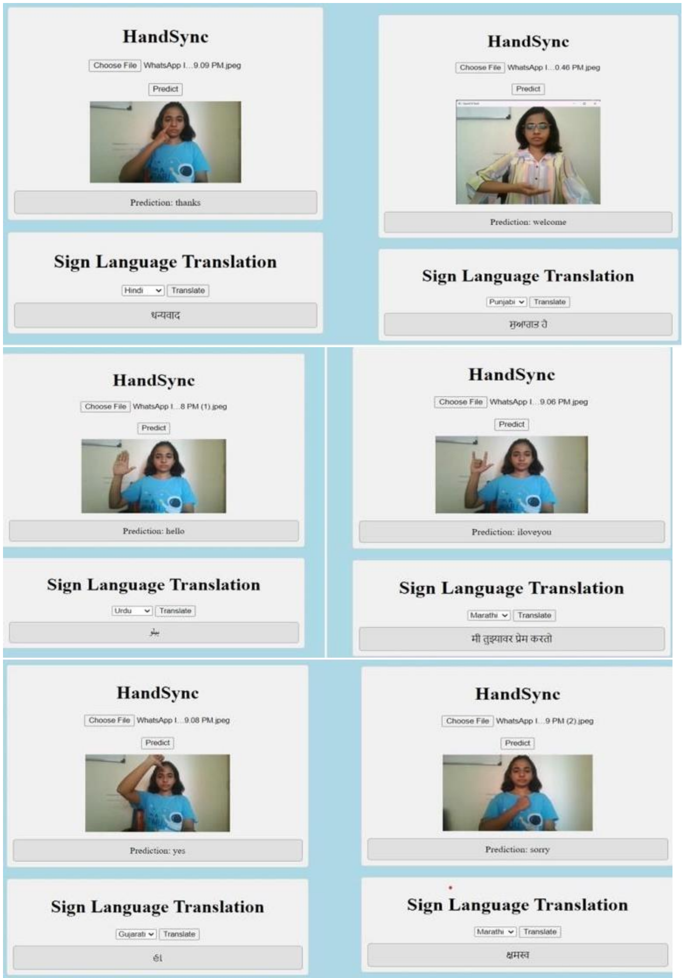

Fig -3: Output

The effectiveness of the suggested sign language recognition system is further clarified by means of quantitative measures like F1 score, accuracy, precision, and recall. When evaluating the system's ability to translate output accurately, these metrics are crucial. by providing an extensive table that shows how well the system performs for a variety of sign language gestures, including translations.

TABLE I. EVALUATION PARAMETERS FOR THE PROPOSED MODEL

Ref	**Accuracy**	**Precision**	**Recall**	**FI-score**
Proposed Model	99.5%	99.4%	99.6%	99.2%

5. CONCLUSION

Sign language recognition has greatly benefited from advancements in machine learning, particularly in computer vision and natural language processing. Deep learning models, such as MediaPipe and Long Short Term Memory (LSTM), have demonstrated impressive results in recognizing signs accurately.

Sign language recognition technology is poised to make a significant impact on the lives of Deaf and hard of hearing individuals. This report underscores the importance of ongoing research and collaboration between experts in machine learning, computer vision, and the Deaf community to drive innovation and make sign language recognition more accurate, accessible, and inclusive. One of the main important future works to be done is to improve the response time and accuracy of the textual outputs. With further advancements and increased awareness, we can look forward to a future where communication barriers are significantly reduced for the Deaf community.

ACKNOWLEDGEMENT

The present world of competition there is a race of existence in which those who have the will to come forward succeed. Project is like a bridge between theoretical and practical work. With this willing we joined this particular project. First of all, we would like to thank the supreme power the Almighty God who is obviously the one who has always guided us to work on the right path of life. We sincerely thank Prof. R. H. Borhade sir, Head of the Department of Computer Engineering of Smt Kashibai Navale college of engineering, for all the facilities provided to us in the pursuit of this project.

We are indebted to our project guide Prof. P. V. Bhaskare, Department of Computer Engineering of Smt. Kashibai Navale college of engineering. We feel it's a pleasure to be indebted to our guide for his valuable support, advice and encouragement and we thank him for his superb and constant guidance towards this project.

We are deeply grateful to all the staff members of the computer department, for supporting us in all aspects. We acknowledge our deep sense of gratitude to our loving parents for being a constant source of inspiration and motivation.

REFERENCES

[1] Carmela Louise L. Evangelista, Criss Jericho R. Geli, Marc Marion V. Castillo: Long Short-Term Memory-based Static and Dynamic Filipino Sign Language Recognition (2023)

[2] Roli Kushwaha, Gurjit Kaur, Manjeet Kumar: Hand Gesture Based Sign Language Recognition Using Deep Learning (2023)

[3] Rinki Gupta, Roohika Manodeep Dadwal: Deep Learning based Sign Language Recognition robust to Sensor Displacement (2023)

[4] Subhangi Kumari, Ernest Tarlue, Aissatou Diallo, Megha Chhabra, Gouri Shankar Mishra, Mayank Kumar Goyal: A Review of Segmentation and Recognition Techniques for Indian Sign Language using Machine Learning and Computer Vision (2023)

[5] Jashwanth Peguda, V Sai Sriharsha Santosh, Y Vijayalata, Ashlin Deepa R N, Vaddi Mounish: Speech to Sign Language Translation for Indian Languages (2022)

[6] Dr. Aruna Bhat, Vinay Yadav, Vishesh Dargan, Yash: Sign Language to Text Conversion using Deep Learning (2022)

[7] Jaya Nirmala: Sign language translator using machine learning (2022)

[8] Mrs.Aerpula Swetha, Vamja Pooja, Vundi Vedavyas, Challa Datha Venkata Naga Sai Kiran, Sadu Sravan: SIGN LANGUAGE TO SPEECH TRANSLATION USING MACHINE LEARNING (2022)

[9] Tuan Linh Dang a,∗ , Sy Dat Tran a , Thuy Hang Nguyen a , Suntae Kim b , Nicolas Monet b: An improved hand gesture recognition system using keypoints and hand bounding boxes (2022)

[10] Sang-Geun Choi , Yeonji Park and Chae-Bong Sohn : Dataset Transformation System for Sign Language Recognition Based on Image Classification Network (2022)

[11] E. B. Villagomez, R. A. King, M. J. Ordinario, J. Lazaro and J. F. Villaverde, "Hand Gesture Recognition for Deaf-Mute using FuzzyNeural Network," 2019 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2019, pp. 30- 33, doi: 10.1109/ICCEAsia46551.2019.8942220.

[12] G. K. R. Madrid, R. G. R. Villanueva and M. V. C. Caya, "Recognition of Dynamic Filipino Sign Language using MediaPipe and Long ShortTerm Memory," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984599.

[13] M. B. D. Jarabese, C. S. Marzan, J. Q. Boado, R. R. M. F. Lopez, L. G. B. Ofiana and K. J. P. Pilarca, "Sign to Speech Convolutional Neural Network-Based Filipino Sign Language Hand Gesture Recognition System," 2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC), Rome, Italy, 2021, pp. 147-153, doi: 10.1109/ISCSIC54682.2021.00036.

[14] K. E. Oliva, L. L. Ortaliz, M. A. Tobias and L. Vea, "Filipino Sign Language Recognition for Beginners using Kinect," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 2018, pp. 1-6, doi: 10.1109/HNICEM.2018.8666346.

[15] M. Allen Cabutaje, K. Ang Brondial, A. Franchesca Obillo, M. Abisado, S. Lor Huyo-a and G. Avelino Sampedro, "Ano Raw: A Deep Learning Based Approach to Transliterating the Filipino Sign Language," 2023 International Conference on Electronics, Information, and Communication (ICEIC), Singapore, 2023, pp. 1- 6, doi: 10.1109/ICEIC57457.2023.10049890.