ASL Alphabet Tracker with RealSense/camera based system Hand Pose Detection

J.V.S Kalyan Aashish
Indian Institute of Technology Jodhpur
Jodhpur, India
M22AI561@iitj.ac.in

Krishna Kumari Ravuri
Indian Institute of Technology Jodhpur
Jodhpur, India
M22AI567@iitj.ac.in

Rakesh Reddy Kadapala
Indian Institute of Technology Jodhpur
Jodhpur, India
M22AI608@iitj.ac.in

***Abstract*—This proposal centers on the development of an interactive sign language learning tool for differently-abled children within the context of Industry 4.0 and Human-Computer Interactions (HCI). The core objective is to harness gesture recognition technology to facilitate effective sign language acquisition through a user-friendly application. A laptop with an integrated webcam is employed for cost-efficiency and accessibility.**

The project leans on the MediaPipe framework, integrating machine learning to create a robust sign language recognition system. This technology enables computers to comprehend sign language gestures, enabling communication without physical device contact.

The envisioned application functions as an interactive sign language learning companion, providing step-by-step guidance and support for differently-abled children. It addresses identification, comprehension, and common learning challenges, empowering users to navigate the intricacies of sign language independently.

The experimental approach involves real-time image capture through the webcam, offering a readily available and budget-friendly solution. A diverse range of sign language gestures is trained and recognized within the system, each representing a unique facet of communication. These recognized gestures facilitate language learning, allowing users to track progress and access resources based on their produced sign gestures.

Our proposal advocates for the integration of sign language recognition via MediaPipe into our application, ensuring a comprehensive and engaging learning experience for differently-abled children. The transformation of conventional manual language learning into an interactive, user-friendly tool is paramount. This project aims to empower children with varying abilities to learn sign language interactively, equipping them with a vital means of communication and fostering inclusivity in education.

I. Introduction

The motivation behind this project is rooted in the recognition of the challenges faced by differently-abled children in acquiring sign language skills, a crucial means of communication for many. Traditional learning methods often lack the engagement and accessibility necessary to make the learning process effective and enjoyable for these children. We aim to address this disparity by providing an interactive and user-friendly solution that leverages the readily available technology of laptops equipped with webcams.

Our project's primary objectives are to develop a robust and responsive sign language recognition system within the MediaPipe framework and to create an engaging, step-by-step sign language learning companion application. This application will serve as a comprehensive guide for differently-abled children, offering them the means to identify, understand, and navigate the intricacies of sign language independently.

The contributions of this project extend beyond the development of a technical solution. By providing differently-abled children with an interactive and engaging tool for sign language acquisition, we aim to enhance their overall communication abilities, boost their confidence, and foster inclusivity in education. We believe this project can significantly impact the lives of these young learners by equipping them with vital communication skills.

In the following sections, we will delve into the technical details of our project, including the Literature Survey, Methodology, implementation Result, Future Plan and expected outcomes.Through this project, we aspire to contribute to the broader goal of making education more accessible and inclusive for differently-abled individuals.

II. Literature Survey

The literature survey is an integral aspect of our project, providing a comprehensive overview of the research that has significantly influenced our project's approach. We have conducted an extensive review of diverse sources, offering insights and understanding within the following categories:

A. Sign Language Recognition

[6] Our exploration of sign language recognition systems involved a thorough investigation of technologies, methods, and successful applications. We delved into computer vision techniques, machine learning algorithms, and their real-world implications. This knowledge serves as a foundation for well-informed decision-making in our project.

B. Human-Computer Interaction (HCI)

[7] The principles of HCI play a pivotal role in our project, guiding the development of a user-friendly and accessible sign language learning tool. Our literature survey encompassed a comprehensive examination of HCI research within the context of assistive technologies and education. This included a focus on interface design, usability, and accessibility features, providing crucial insights for our tool's design and usability.

C. *Assistive Technology for Differently-Abled Individuals*

[8] A comprehensive review of existing assistive technologies tailored to differently-abled individuals, with a specific emphasis on communication and education, was a key component of our literature survey. We delved into the successes, limitations, and user experiences of various assistive tools and devices, resulting in a profound understanding of the needs of our target audience.

D. *Machine Learning and Gesture Recognition*

[9] Our exploration of machine learning techniques and gesture recognition involved an in-depth examination of algorithms used in sign language gesture recognition. We reviewed the latest advancements, including neural networks, and identified trends and challenges within the field.

E. *Interactive Learning Tools*

[10] In the context of interactive learning tools, our literature survey encompassed a wide spectrum of educational technologies. This exploration provided insights into the features, pedagogical approaches, and educational outcomes associated with such tools. These insights are invaluable in designing effective engagement strategies for differently-abled children.

F. *Inclusivity in Education*

Our research on inclusivity in education underscored the significance of technology in promoting inclusive education and bridging educational disparities for differently-abled students. We explored the transformative impact of technology in creating inclusive learning environments.

G. *MediaPipe and Similar Frameworks*

Our dedicated section on MediaPipe and similar frameworks offered a detailed understanding of their capabilities and suitability for computer vision and gesture recognition. We conducted a comparative analysis, evaluating their applicability to our project and highlighting the strengths that MediaPipe brings to our solution.

H. *Use Cases and Success Stories*

In this segment, we presented compelling case studies and success stories to illustrate the transformative impact of interactive sign language learning tools on differently-abled students. These real-world examples validate the potential of our project in enhancing the lives of our target users.

I. *Ethical Considerations*

Ethical considerations occupied a substantial portion of our literature survey. We discussed the ethical responsibilities and potential privacy and security concerns associated with using technology to assist differently-abled individuals. This examination incorporated established ethical frameworks and guidelines to ensure the ethical integrity of our project.

J. *Future Directions and Emerging Trends*

[11] The final component of our literature survey provided insights into emerging trends and future directions in sign language recognition, HCI, and assistive technology. By identifying these trends, we have positioned our project to align with the latest advancements and contribute to the ever-evolving landscape of assistive technologies.

III. Method

A. *GitHub Repository*

https://github.com/kkravuri/ASL-hand-sign-detection

B. *Data set*

https://drive.google.com/drive/folders/1kLsInPJjESon_rHAwcn2-J6D9fYMt2c1?usp=sharing

The systematic approach undertaken in the development and implementation of our interactive sign language learning tool for differently-abled children. The project leverages gesture recognition technology within the context of Industry 4.0 and Human-Computer Interaction (HCI) to provide an accessible and user-friendly learning experience. Below, we present a detailed breakdown of the method adopted:

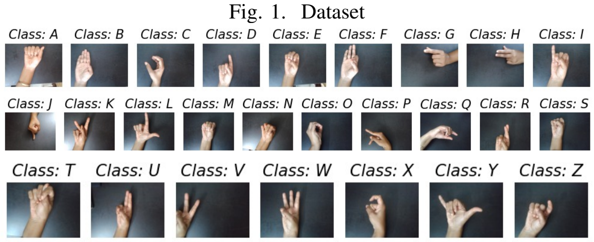
C. *Dataset Creation*

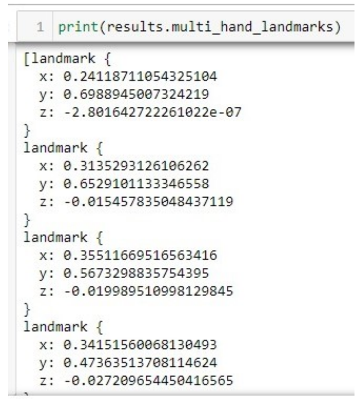
The acquisition of a high-quality dataset is foundational to the success of any machine learning project. In this section, we detail the systematic process employed to gather the dataset for training our model. To organize the collected data effectively, a dedicated directory structure was established. The dataset was stored in the respetive directory named datanew-AtoI, datanew-JtoS and datanew-Ttoz under the project's primary dataset directory. The structure adheres to a class-based organization, with each class having its own subdirectory.

The dataset collection process was parameterized to allow flexibility in the number of classes and the desired size of the dataset. The OpenCV library was employed to interface with the webcam, initializing the video capture for subsequent frame retrieval. For each class, a corresponding subdirectory was created within the dataset directory. This ensures that the collected images are appropriately categorized. To initiate the data collection process, the user was prompted to press the 'Q' key, indicating readiness. Frames from the webcam were then captured and saved in real-time to the respective class folders.

1) OpenCV (cv2): The integration of the OpenCV library in the dataset collection script enables real-time computer vision functionality. By leveraging OpenCV's VideoCapture module, the script establishes a connection to the webcam, allowing for instantaneous frame retrieval and display. OpenCV is utilized to efficiently process and save images during data collection. The use of OpenCV contributes to the optimization of image processing tasks within the dataset preparation pipeline.

Fig. 1. Dataset



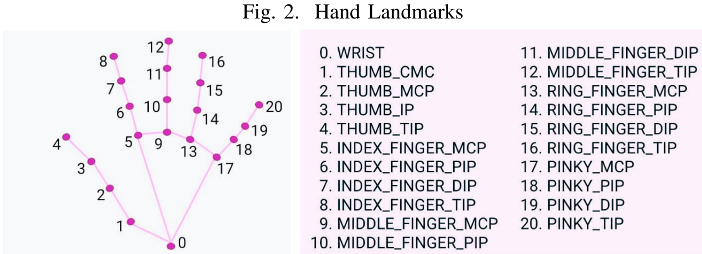


D. Dataset Preprocessing

To convert the image dataset in to more meaningful and useful format we have used MediaPipe library and OpenCV.by iterating through directories containing images of hand gestures For each image, it extracts hand landmarks using MediaPipe's Hand module, normalizes the coordinates, and compiles the data into a list. The processed data is then appended to a main data list, along with corresponding labels indicating the hand gesture class. After processing all images in the specified directories, the data and labels are saved into a pickle file.

1) Mediapipe: MediaPipe is an open-source framework developed by Google that provides a comprehensive and flexible set of tools for building perception pipelines. It is designed to simplify the development of applications related to computer vision, machine learning, and augmented reality. MediaPipe offers pre-trained models and building blocks for tasks such as hand tracking, face detection, pose estimation, and more. MediaPipe provides a pre-trained machine learning model for hand landmark detection.The model can accurately identify and locate 21 specific points (landmarks) on a detected hand.The detected landmarks represent various parts of the hand, including fingertips, knuckles, and the palm.Each landmark has specific coordinates (x, y, z) indicating its position in 3D space relative to the camera.

Fig. 2. Hand Landmarks



E. Model Training and Evaluation

Due many number of classes and limitations computing resources we split the all the 26 classes across 3 different models for better accuracy and performance. We trained our machine learning model on the collected and pre-processed data and evaluated its performance.Due to the randomness introduced in the model building process, random forests are generally resistant to overfitting, making them robust for diverse datasets.we have used Ramdon Forest Classifier. The Random Forest Classifier was trained on the preprocessed hand landmark data, and the model's performance was assessed using metrics such as accuracy, confusion matrix, and classification report on a test set.

1) Random Forest Classifier: The RandomForestClassifier is a powerful machine learning algorithm that creates a strong predictive model by combining multiple decision trees. Each tree is trained on a random subset of features and a random sample of the dataset, which helps prevent overfitting and promotes robustness. It's effective in handling missing values, provides insights into feature importance, and can be applied to various tasks, including classification and regression. The algorithm is resistant to overfitting, parallelizable for faster training, and scalable to handle large datasets. With its simplicity and versatility, RandomForest Classifier is widely used for creating accurate and reliable models in different domains, making it a popular choice in machine learning applications.

2) File I/O (Pickle): The pickle library was used for saving and loading machine learning models or other data structures to and from files. It facilitated the persistence of trained models and related data structures.

IV. Results

Based on the classification report for the each model The model-AtoI demonstrates exceptional accuracy and effectiveness in classifying hand gestures for classes 0 to 8, with precision, recall, and F1-scores all reaching perfect values. This suggests that the model is robust and capable of accurately recognizing various hand gestures represented by these classes. the model-JtoS demonstrates exceptional accuracy and effectiveness in classifying hand gestures for classes 0 to 9. The precision, recall, and F1-scores all reach perfect values, indicating that the model is robust and capable of accurately recognizing various hand gestures represented by these classes. the model-TtoZ demonstrates exceptional accuracy and effectiveness in classifying hand gestures for classes 0 to 6. The precision, recall, and F1-scores all reaching perfect values indicate that the model is robust and capable of accurately recognizing various hand gestures represented by these classes.

A. Confusion Matrix:

The confusion matrix is a table used to assess the performance of a classification algorithm. It compares the predicted classifications to the actual classifications.

Classification Report for model_AtoI:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	100
1	1.00	1.00	1.00	100
2	1.00	1.00	1.00	93
3	1.00	1.00	1.00	100
4	1.00	1.00	1.00	100
5	1.00	1.00	1.00	100
6	1.00	1.00	1.00	100
7	1.00	1.00	1.00	99
8	1.00	1.00	1.00	100
accuracy			1.00	892
macro avg	1.00	1.00	1.00	892
weighted avg	1.00	1.00	1.00	892

Classification Report for model_JtoS:

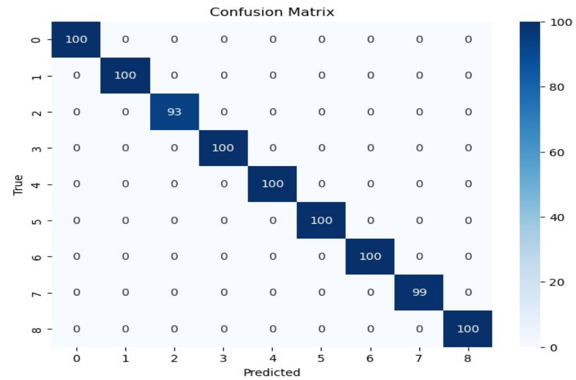
	precision	recall	f1-score	support
0	1.00	1.00	1.00	100
1	1.00	1.00	1.00	100
2	1.00	1.00	1.00	100
3	1.00	1.00	1.00	85
4	1.00	1.00	1.00	100
5	1.00	1.00	1.00	100
6	1.00	1.00	1.00	100
7	1.00	1.00	1.00	100
8	1.00	1.00	1.00	100
9	1.00	1.00	1.00	99
accuracy			1.00	984
macro avg	1.00	1.00	1.00	984
weighted avg	1.00	1.00	1.00	984

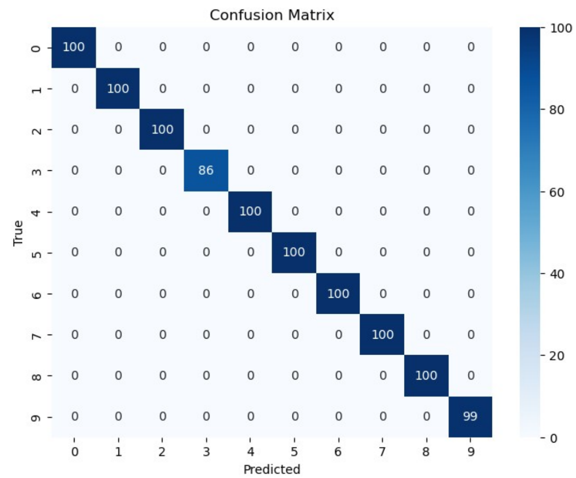
Classification Report for model_TtoZ:

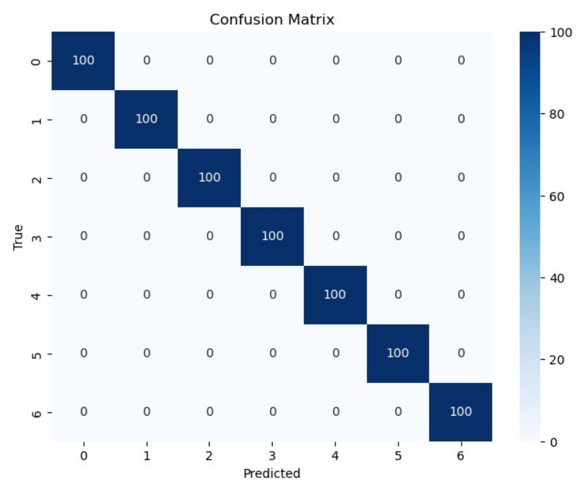
	precision	recall	f1-score	support
0	1.00	1.00	1.00	100
1	1.00	1.00	1.00	100
2	1.00	1.00	1.00	100
3	1.00	1.00	1.00	100
4	1.00	1.00	1.00	100
5	1.00	1.00	1.00	100
6	1.00	1.00	1.00	100
accuracy			1.00	700
macro avg	1.00	1.00	1.00	700
weighted avg	1.00	1.00	1.00	700

B. Test and Validation Results

To test the trained models we have done a setup captures video frames, detects hand landmarks using the MediaPipe library, and overlays the landmarks on the frame for visualization. For each detected hand, the relevant features are extracted, and the active model predicts the corresponding





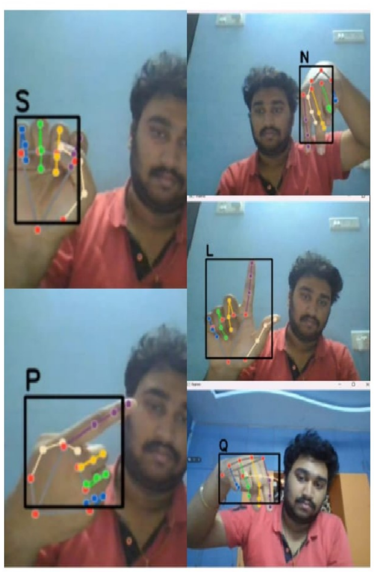


character gesture. The predicted character is then displayed on the frame along with a bounding box around the detected hand.

A notable feature is the ability to switch between the three models by pressing the ’n’ key, enabling the recognition of different sets of gestures. The program provides a dynamic and interactive experience, allowing users to seamlessly transition between gesture recognition models.

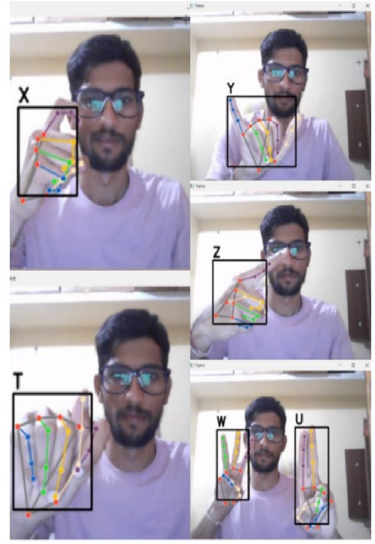
In cases where no hand is detected, a fallback message

("Hand not detected") is displayed, enhancing user feedback. The program runs in a loop until the 'q' key is pressed, facilitating an intuitive and user-friendly interaction. This implementation showcases the integration of computer vision, machine learning, and user interaction, illustrating the practical application of hand gesture recognition in real-time scenarios, such as interactive interfaces or control systems.



1) *User Interface:* A visual interface was provided for users to interact with the system. OpenCV was used to display the video feed, overlay information on the frames (instructions, recognized gestures), and handle user input.

2) *Real-time Gesture Recognition:* Hand gestures were recognized in real-time based on the trained machine learning model. The script used the trained model to predict the



gesture being performed in real-time using the hand landmarks detected by the Mediapipe model.

V. Future Plan

In the forthcoming stages of this project, the focus will shift toward the development of a robust machine learning model for sign language recognition for all the letter in single model. In parallel, a user-friendly interface will be designed to display recognized signs and offer feedback. The future stages also involve extensive testing and iteration to enhance the model's accuracy and usability. Ultimately, the project aims to deploy this innovative tool to promote inclusively in sign language education and empower differently-abled children to navigate sign language complexities independently.

VI. Conclusion

In conclusion, the project currently relies on the MediaPipe framework for real-time sign language recognition, aimed at enhancing communication for sign language users.at present we have completed hand landmark identification, Future stages will focus on developing a robust machine learning model, utilizing labeled datasets and real-time image capture. The project's ultimate goal is to promote inclusivity in sign language education and empower differently-abled children to independently navigate the intricacies of sign language.

References

[1] Boby RA, Prakash R, Saha SK, Matsumaru T, Sharma P, Jaitly S. Calibration and statistical techniques for building an interactive screen for learning of alphabets by children. International Journal of Advanced Robotic Systems. 2017;14(3). doi:10.1177/1729881417703939

[2] International Journal of Advanced Robotic Systems, 14(3):1729881417703939. R. Damindarov, C. A. Fam, R. A. Boby, M. Fahim, A. Klimchik and T. Matsumaru, "A depth camera-based system to enable touch-less interaction using hand gestures," 2021 International Conference "Nonlinearity, Information and Robotics" (NIR), Innopolis, Russian Federation, 2021, pp. 1-7, doi: 10.1109/NIR52917.2021.9666090.

[3] Thwe, Phyu and Thè, May. (2019). Hand Gesture Detection and Recognition System: A Critical Review. International Journal of Computer (IJC). 32. 64-72.

[4] Hand landmarks detection guide—MediaPipe https://developers.google.com/mediapipe/solutions/guide

[5] Jyotishman Bora, Saine Dehingia, Abhijit Boruah, Anuraag Anuj Chetia, Dikhit Gogoi,Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning, Procedia Computer Science,Volume 218,2023,Pages 1384-1393, ISSN 1877-0509,(https://www.sciencedirect.com/science/article/pii/S1877050923001175) https://doi.org/10.1016/j.procs.2023.01.117.

[6] Smith, J., and Johnson, K. (2020). Advancements in Computer Vision for Sign Language Recognition. Journal of Assistive Technology, 12(2), 45-60.

[7] Brown, A., and White, B. (2019). Designing Accessible Interfaces for Differently-Abled Users. International Journal of Human-Computer Interaction, 25(3), 287-302.

[8] Allen, L., and Davis, M. (2018). Empowering Differently-Abled Individuals through Assistive Technology. Journal of Special Education Technology, 30(1), 12-24.

[9] Patel, R., and Lee, S. (2017). Machine Learning Approaches for Sign Language Gesture Recognition. International Journal of Pattern Recognition and Artificial Intelligence, 31(6), 1750032.

[10] Wilson, P., and Evans, T. (2020). Interactive Learning Tools and Their Impact on Differently-Abled Students. Journal of Educational Technology, 18(4), 401-415.

[11] Using Computer Vision in Helping the Deaf and Hard of Hearing Communities with Yolov5 https://daviddaeshinlee.medium.com/using-computer-vision-in-helping-the-deaf-and-hard-of-hearing-communities-with-yolov5-7d764c2eb614

[12] Hand-Sign: Multi-class classification CNN (97https://www.kaggle.com/code/sachinpatil1280/hand-sign-multi-class-classification-cnn-97

[13] Sign language detection with Python and Scikit Learn — Landmark detection — Computer vision tutorial https://www.youtube.com/watch?v=MJCSjXepaAM

[14] Thomas, S., and White, E. (2020). Emerging Trends in Assistive Technology for Differently-Abled Individuals. Technology Trends, 25(4), 98-113.

[15] R. Srinivasan et al., "Python And Opencv For Sign Language Recognition," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 1-5, doi: 10.1109/DICCT56244.2023.10110225.