International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

Real-Time Sign Language Recognition Using Advanced Computer Vision

S V Hemanth[1], Valluri Dinesh Ram[2], Raghupathy Aditya Raj[3], Pasula Nithin[4], Vadtya Niharika[5], Tumma Bharath Kumar[6]

[1]Associate professor, Dept. of CSE, Hyderabad Institute of Technology and Management, Hyderabad, Telangana, India.

[2,3,4,5,6]UG Scholar, Dept. of CSE, Hyderabad Institute of Technology and Management, Hyderabad, Telangana, India.

Emails: *hemanth.sandapaga@gmail.com[1], dineshramv13@gmail.com[2], aditya63037@gmail.com[3], nithin.pasula@gmail.com[4], vadtyaniharika@gmail.com[5], bharathkumartumma333@gmail.com[6]*

Abstract

Sign language interpretation tools are effective tools for the deaf or hard of hearing to communicate. The proposed research provides an effective solution to such getting over communication and accessibility barriers in dynamic sign language recognition for people who are both hearing and speaking impaired. Leveraging advanced machine learning techniques, including Random Forest Classifiers, and the Google MediaPipe framework, we propose a robust system capable of interpreting hand gestures into text. Our methodology encompasses data collection under diverse conditions to ensure robustness, preprocessing with geometric feature engineering for enhanced accuracy, and model training on 26 English alphabets. MediaPipe Hand landmark detection provides 21 key points per hand, enabling precise spatial feature extraction, while data augmentation improves adaptability to real-world variations. The system achieved a high classification accuracy of 90.00%, attributed to the ensemble learning capabilities of the Random Forest model. Real-time inference is supported by a seamless integration of MediaPipe Hands for gesture detection and temporal smoothing for stability in dynamic scenarios. Deployed on the Streamlit platform, the user-friendly interface allows for data collection, training, live predictions, constructing words and sentences dynamically based on recognized gestures. This project demonstrates significant potential for applications in education, healthcare, and public services, enhancing accessibility for the hearing impaired. Demonstrating inclusivity through technology establishes a basis for future innovations, such as understanding complete phrases or continuous sign language, further narrowing communication gaps between auditory and non-auditory communities.

Keywords: *Sign language recognition, Google MediaPipe, Machine Learning, Real-time interpretation, Streamlit deployment, Image processing, Computer Vision.*

1. Introduction

For the deaf and the incapable of speaking for themselves, sign language is actually a very important means of communication not only with other impaired people but also with a greater community. However, interpreting sign language if unknown is inherently presented as a limitation. It is currently possible to develop a system now through computer vision and machine learning that can recognize and interpret real-time sign language gestures. This project will develop a friendly sign language recognition tool through Google MediaPipe and machine learning to accurately identify hand gestures that correspond to the alphabets. The sign language recognition tool thus developed can be applied in several real-time scenarios, including educational institutions, workplaces, public services, and health care areas where accessibility is vital. The project also offers an opportunity for further expansion. With additional training and more complex models. By doing so, the tool could significantly enhance the day-to-day interactions of the hearing-impaired, empowering them to engage seamlessly in a world where spoken language is dominant. Ultimately, this sign language recognition project showcases how technology can be harnessed to foster inclusivity and bridge communication gaps, allowing more meaningful and accessible exchanges between the hearing and non-hearing communities.

2. Literature Review

Communication challenges faced by individuals with hearing and speech impairments can be alleviated through the development of technological tools for sign language interpretation. Sign languages, based on hand shapes and movements, are not widely understood by people reliant on spoken and written communication. The approach demonstrated effective recognition of American Sign Language (ASL) alphabets and gestures and offers potential for extending sign language recognition to local Indian languages [1]. Sign language is an essential communication tool for many people worldwide, but recognizing it through visual means is a complex task due to variations in hand shapes, motions, and body positions across different sign languages. Recent advancements in deep learning have significantly improved sign language recognition, especially with vision-based models [2]. A developed system uses highly advanced machine learning algorithms and CNNs to directly translate sign language gestures into text or speech by enabling communication from hearing persons to deaf people and vice versa. In it, the model has been trained using a carefully prepared dataset of sign language gestures for the accurate classification of hand shapes and positions. Data preprocessing techniques are employed to emphasize distinctive hand features, enhancing computational efficiency and classification performance [3]. The recognition of hand signs is very important for communication quality improvements in the hearing-impaired population. The current system had better and, hence more effective performance, compared with other approaches, in recognizing static ISL signs with high precision [4]. Assamese Sign Language is one of India's 22 modern languages that, since its gestures are not widespreadly understood, came across a great challenge on communication issues. An approach was designed using machine learning methods directed at hand gesture recognition in Assamese Sign Language. A dataset consisting of two-dimensional and three-dimensional images of nine static gestures, including vowels and consonants, comprised 2094 data points. The MediaPipe framework detected landmarks, and a feed-forward neural network achieved 99% accuracy during training [5]. Sign languages have played a huge role in communicating for individuals with auditory impairments or deafness. Advancements have affected recognition, visualization, and synthesis of sign languages. This review explores technological developments in these areas, focusing on innovations that aid in communication, learning, and translation of sign languages. The study analyzes over 2000 articles published between 2010 and 2021, highlighting the pivotal role of image processing and deep learning in enhancing sign language applications [6]. In recent years, advances in deep learning for sign language perception focus on CNN automatically extracting features instead of handcrafted ones. Gesture was captured accurately and GPU acceleration for real-time processing helped this experiment achieve 91.7% cross-validation accuracy in classifying 20 Italian gestures in one study. The model exhibited excellent generalization, both over the user and the environment, and the mean Jaccard Index in the gesture spotting competition of ChaLearn in 2014 was 0.789, thus showing efficiency in different conditions [7]. Hand signal recognition is the essential component in context of successful human-computer interaction. The system demonstrates 99.82% recognition accuracy outperforming several advanced techniques achieved in an efficient, cost-effective, and operable manner on standard CPUs [8]. Quantifying fine movements, especially for clinical purposes like assessing hand dysfunctions and rehabilitation outcomes, presents significant challenges. GMH-D demonstrated higher spatial accuracy than GMH with consistent ability in established gold standards both for slow movement and fast movement. The results established GMH-D as an appropriate framework for the evaluation of accurate 3D hand movement in healthcare environments [9]. Sign language recognition systems

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

are to make communication for the hearing impaired more effective. Many traditional approaches like edge detection and motion tracking were so sensitive to light or background changes. Although skin-color modeling based on thresholds in color spaces such as RGB or HSV is successful for hand segmentation, it demands careful tuning for various skin tones [10].
Hand gesture recognition offers solutions for various challenges by enabling machines to interpret human activities, with applications extending to sign language recognition. At each level, the merits of various algorithms are compared as well as the difficulties and constraints present only in gesture recognition and its application to sign language [11]. Effective hand sign detection tools are essential for assisting people with hearing impairments. This reviews automated sign language recognition (SLR) techniques from 2014 to 2021, emphasizing the need for classification approaches that can correctly interpret diverse data. While recent advancements allow for recognizing continuous sign language communication, the generalization potential for commercial use remains a challenge [12]. Significant advancements in sign language recognition and rendering include robotics, virtual reality sensors, and computer vision technologies. Early efforts such as "Ralph" focused on finger-spelling hands, while projects like CyberGloves and camera-based systems like CopyCat emphasized gesture capture and interaction. Hidden Markov Models and neural networks enhanced recognition accuracy [13]. However, the communication between the deaf and society at large has remained a problem because of the lack of knowledge on how to use sign language. This presents the hand gesture to text translation using the method of image processing and CNN. The CNN algorithm forms the classifier. In addition, ensemble was applied to merge the CNN models, which highly improved the accuracy of the system up to 99.4%. This indicates that ensemble techniques do improve classification accuracy. [14]. Unconstrained hand gesture detection in natural video sequences offers extensive applications in human-computer interaction. A database of hand images and background images supports training. Experimental evaluations demonstrated a detection accuracy of 92.23%, highlighting the method's efficacy and robustness in gesture detection and classification [15]. Gesture and sign language recognition system suffers from the effective handling of sequential data in continuous recognition tasks. The traditional approaches based on HMMs rely heavily on pre-defined features. On the other hand, the CNNs provide very strong capabilities in feature extraction while lacking temporal modeling. These approaches show a large reduction of error rate from benchmark datasets [16]. The software system has recognized hand gestures based on features such as orientation, centroid, and finger status. For such natural similarities between human hand structures, the software enables intuitive communication among the auditory and verbal impairment community with others using innovative and user-friendly interfaces [17]. Recognizing Indian Sign Language gestures is a challenging task due to the complexities of head and hand movements. This provides an indication that Convolutional Neural Networks can indeed be applied to this problem. The system employs a selfie mode continuous video capture approach where the deaf independently self-operate the SLR mobile application. A set of 200 signs performed by five different signers from angles and backgrounds was created as a dataset. [18]. One of the reasons why a speech-impaired person would require sign language for expression, this becomes problematic for the non-sign language speaker. Deep learning and computer vision have introduced tremendous promises in motion and gesture recognition. The model used the American Sign Language dataset and provided one of the promising solutions toward sign language translation [19]. It will give a base framework to the identification of alphabets and digits in ISL Indian Sign Language through the live video stream. The system will be working on a segmentation basis of skin color and Background Subtraction, and classification will be done using SVM and CNN respectively. Moreover, an interactive Graphical User Interface (GUI) has been designed so that the recognition tool is accessible easily and produces its outputs both in text and speech [20]. A real-time BdSL recognition system designed with computer vision methods has been proposed. The proposed

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

system makes use of cascaded classifiers based on Haar-like feature detection to distinguish hands in a frame-to-frame manner. Hand signs are distinguished based on Hue and Saturation values assigned to human skin tones and then converted into binary images. The system uses K-Nearest Neighbors (KNN) classification to match these images with pre-trained binary hand sign images [21]. Tradition has it that sign language recognition was accomplished through traditional techniques of image processing and, later on, by deep learning methods, particularly Convolutional Neural Networks (CNNs), to extract features in the process of gesture classification. Those early methods were highly frustrated by background noise and lighting variations; however, CNNs have improved accuracy in the recognition of static and dynamic hand gestures [22]. Speech impairments reduce the effective verbal communication, and sign language has been recognized as one of the primary solutions for such disabled people. However, the general public's inability to understand sign language is an obstacle toward the implementation of sign language recognition systems. The earlier systems had utilized some techniques for image processing like color thresholding and edge detection for extracting hand gestures but still posed a problem under different environmental conditions [23]. Deep learning has led to tremendous growth in the use of target detection, recognition, and tracking in computer vision. Improved applications are found in other related fields, such as sign language recognition. The main reason is for developing sign language recognition systems since techniques in computer vision are widely used and people who do not know the use of sign language still communicate more efficiently [24]. Multimodal sign language recognition is a powerful approach in the improvement of accuracy and robustness of the recognition systems with data coming from different sources, such as image classification, motion capture devices, or even Leap Motion. Early simple recognition systems were based on a single modality working on features extracted from images or motion data, but the limitation was with complex dynamic gestures. It has recently been shown how different modalities, such as vision through cameras and motion data from sensors, can increase the recognition accuracy by significant factors [25].

3. Proposed Work

This sign language recognition system is developed based on the competence of machine learning, real time processing and computer vision in deriving text form hand gestures. These include the key stages such as data collection, preprocessing, training of the model, dynamic inference, integration with the frontend, and performance evaluation of the model.

3.1 Data Collection: Building a Robust Dataset

The process begins with comprehensive data collection, capturing images of hand gestures representing the 26 English alphabets. Using a webcam, gestures are recorded under varying conditions to account for differences in lighting, hand orientation, and background settings. Computer vision techniques ensure efficient capture and organization of images, automating tasks such as hand detection, cropping, and alignment to enhance data consistency. Each image is labeled and stored in alphabet-specific folders. The diversity and size of the dataset are critical to ensuring the model's robustness, with multiple samples collected for each alphabet to handle variability in real-world scenarios.

3.2 Preprocessing and Dataset Creation: Feature Engineering with Computer Vision

Preprocessing and dataset creation convert raw images into meaningful features for model training. Computer vision tools such as MediaPipe Hands detect hand landmarks in each image by identifying 21 key points on the hand, a process that relies on advanced computer vision. These landmarks are normalized relative to the bounding box or the hand's position to ensure scale and positional invariance:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}, y' = \frac{y - y_{min}}{y_{max} - y_{min}}$$

Geometric relationships between key landmarks are captured using Euclidean distance formula:

$$d(p,q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Used to calculate the range between two points (e.g., hand landmarks in MediaPipe). This helps determine the spatial relationships of hand joints. To represent

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

the relative positions of fingers, angles between adjacent joints are calculated:

$$Cosine\ Similarity\ = \cos(\theta) = \frac{A \cdot B}{\parallel A \parallel \parallel B \parallel}$$

This is crucial for identifying the directionality or angles between joints.

$$A \cdot B = (x_2 - x_1)(x_3 - x_2) + (y_2 - y_1)(y_3 - y_2)$$

Detection of joint positions and the directionality or angles between them is fundamentally dependent upon the accuracy of the hand landmark model. Precise hand gesture recognition and movement analysis requires this information.

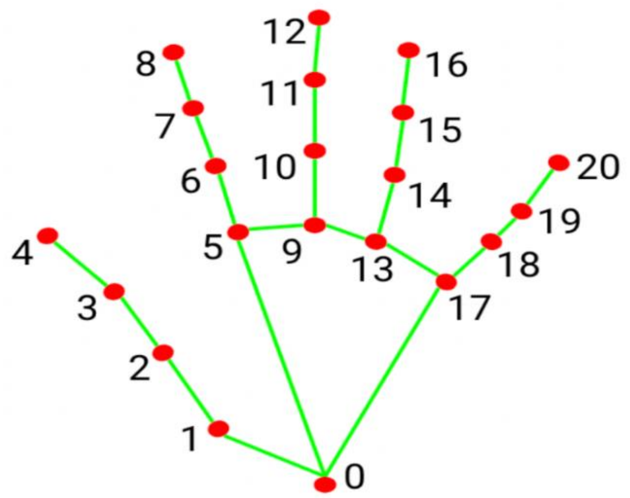

Figure 1 Hand Landmark Model

Figure 1, the hand landmark model bundle predicts the positions of 21 hand and knuckle joint coordinates within the identified hand regions. It was trained on approximately 30,000 real world hand images, supplemented with synthetic hand models over varied backgrounds. Feature vectors are constructed from these normalized coordinates, incorporating additional geometric features such as distances between landmarks and angles between joints. For further generalization, the augmentations of transformations in the form of rotation flip and lighting adjustments are applied on the dataset. Realistic computer vision ensures these augmentations retain real-world variations, thereby making the system better in handling diverse environments.

3.3 Model Training: Machine Learning for Gesture Recognition

The feature vectors resulting from the training of the model are applied to train a Random Forest Classifier mapping hand gestures into the corresponding alphabets. The pre-processing steps feed robust features invariant to the computer vision process, yielding augmentation in the efficiency of learning. The Random Forest Classifier splits data at each decision node based on the Gini Impurity:

$$Gini = 1 - \sum_{i=1}^{C} {p_i}^2$$

where p_i is the fraction of samples assigned to class i and c is the number of classes.
Alternatively, splits are determined using entropy:

$$Entropy = -\sum_{i=1}^{C} p_i \, log_2(p_i)$$

In Random Forest, feature importance estimates the extent to which a particular feature helps to split the data. Additionally, in Random Forest, feature importance refers to the extent to which a certain feature can minimize the impurity measure (Gini, Entropy etc.) in all the decision trees within the forest. It determines the level of participation of each feature in the given classification.

$$FI_j \ = \ \frac{1}{T}\sum_{t=1}^{T}. \sum p(t_k)\Delta i(t_k, j)$$

Where T is number of trees in the Random Forest, $p(t_k)$ is proportion of samples at node k and $\Delta i(t_k, j)$ is decrease in impurity for feature j at node k, indicating feature significance. This ensemble-based machine learning approach improves accuracy and reduces overfitting. In training, the system uncovers deep, complex patterns in the feature vectors by decision trees at every level assisting to a final prediction. Hyperparameters such as number of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

trees and their depth are optimized to provide maximum performance. Once trained, the model is serialized using pickle for deployment in the real-time.

Algorithm Steps:

- **Step 1:** Collect hand gesture images for 26 sign language classes using a webcam and OpenCV.
- **Step 2:** Use MediaPipe Hands to detect 21 hand landmarks and extract their normalized coordinates.
- **Step 3:** Convert hand landmarks into numerical feature vectors representing gesture-specific patterns.
- **Step 4:** Configure a Random Forest Classifier with optimal hyperparameters. Train the Random Forest model on the extracted features and their labels.
- **Step 5:** Compute F1-score, accuracy, precision and recall using a classification report.
- **Step 6:** Use MediaPipe Hands for real-time hand landmark extraction during inference.
- **Step 7:** Pass extracted landmarks to the trained Random Forest model to predict gestures in real-time.
- **Step 8:** Continuously update and display the "current word" and construct sentences dynamically.

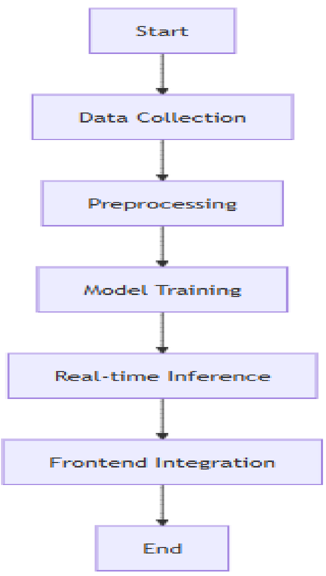

Figure 2 Workflow of A Real-Time Sign Language Recognition System

Figure 2, the flow diagram outlines the Real-Time Sign Language Recognition Application Initiation Captures and Saves Images: Images for Labeled will be captured and saved. The model was trained using landmark data with the use of random forest applying this in preprocessed landmark data. Frames will capture real time. Then hand landmarks will be extracted, and signs predicted by the model that has been trained. The system depicts the results in real-time; the current and known words realize the efficient and interactive sign language recognition before the application ends.

3.4 Real-time Inference: Dynamic Gesture Prediction

The cornerstone of the system is real-time inference, which employs computer vision to process live video feeds and predict hand gestures dynamically. Each frame captured by the webcam undergoes landmark detection and feature extraction using MediaPipe Hands, ensuring consistency with the preprocessing pipeline. The Random Forest Classifier predicts the corresponding alphabet for each frame using majority voting:

$$\hat{y} = mode\{h_1(x), h_2(x), \dots, h_n(x)\}$$

where h_i (x) is the prediction of the i- th tree.

The trained model predicts the corresponding alphabet for each frame in real time, with temporal smoothing mechanisms enhancing stability by aggregating predictions across multiple frames. Computer vision enables efficient tracking of hand movements, ensuring reliable performance even in dynamic conditions.

3.5 Frontend Integration: Real-time Integration and Feedback

Frontend integration ensures the system is accessible and user-friendly, utilizing computer vision for interactive visual feedback. The real-time interface displays the detected hand landmarks, the predicted alphabet, and the constructed sentence, providing immediate insights into the recognition process. The integration of computer vision enhances the user experience by offering visual overlays, such as bounding boxes and landmark points, to guide users during gesture recognition. The design prioritizes

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

simplicity and interactivity, enabling seamless navigation for tasks like restarting sentence formation or correcting errors. This intuitive interface bridges the gap between the model's computational backend and the user's experience.

3.6 Model Evaluation and Improvement: Ensuring Robustness

Extensive testing under diverse real-world conditions ensures robustness, with computer vision providing tools to analyze failure cases, such as misclassified gestures due to occlusions or lighting changes. When errors occur, user feedback is incorporated into an iterative refinement process, with new samples added to the dataset for retraining. This ongoing evaluation loop ensures the model adapts to real-world variations, maintains high reliability, and remains scalable for broader use cases.

4. Result and Discussion

This section shows what the classification model did by using prime metrics, such as feature importance, accuracy, and classification performance. All this now leads to confusion matrix and classification report based on these top features, and it is seen that the model works well in classifying the sign language gestures. Implications of results This section discusses in-depth the implications of results wherein the model relies upon certain features for generalization and on the potential real-time applications. Well, this controlled dataset does demonstrate flawless performance, so it relates back to the model's reliability.

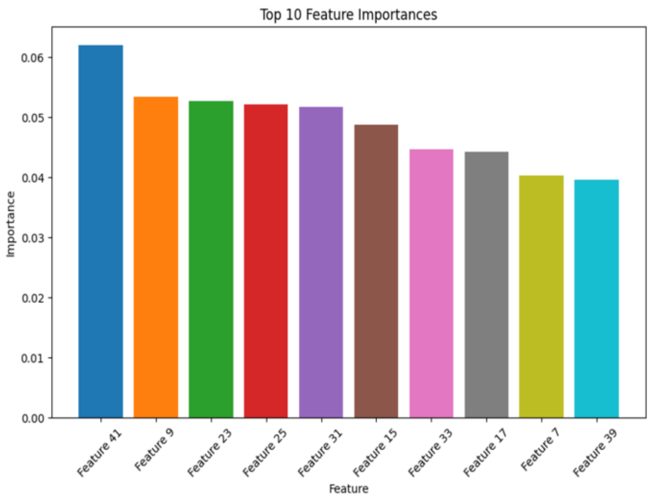

Figure 3 Process of The Dataset

Figure 3, the feature importances describe which features most contributed to this decision process of a classification model. The top 10 features based on their importance values show that Feature 41 is the most important, with an importance of 0.0619, which is 1.6 times greater than the tenth feature, Feature 39. This significant disparity suggests that the model heavily relies on a few specific features to make predictions. The remaining features show decreasing importance values, with Feature 9 at 0.0533, Feature 23 at 0.0527, and so on.

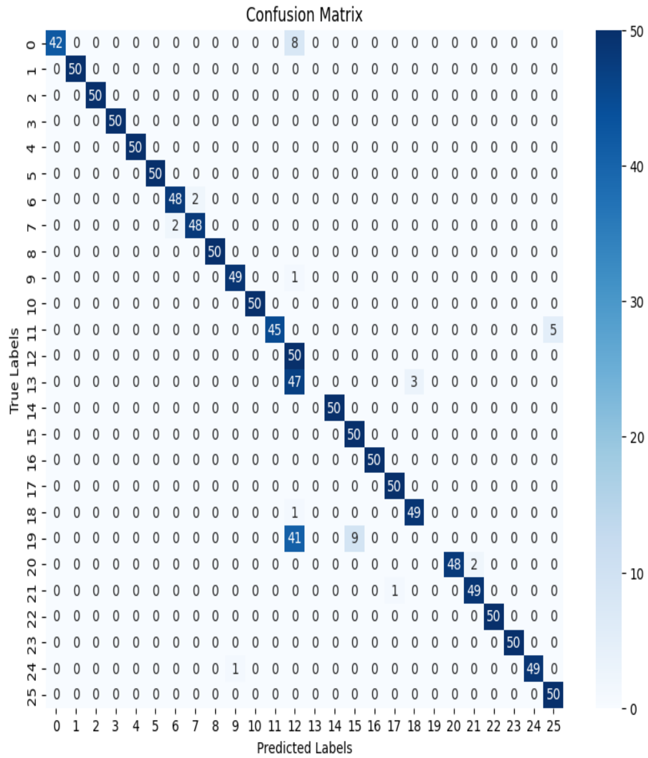

Figure 4 Confusion Matrix

Figure 4, the confusion matrix represents a summary of the model performed on the classification for each of the 26 classes. While many classes achieved strong results, the values on the diagonal are not consistently 50, indicating that some misclassifications occurred. For instance, in class 1, all 50 instances were correctly classified. In class 0, only 42 out of 50 instances were correctly classified, with 8 misclassified into other classes. Class 13 and 19 shows a higher level of misclassification. This result highlights areas where the model could further tune or additional data to improve its robustness.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

	precision	recall	f1-score	support
a	0.98	0.88	0.93	50
b	1.00	1.00	1.00	50
c	1.00	1.00	1.00	50
d	1.00	1.00	1.00	50
e	1.00	0.98	0.99	50
f	1.00	1.00	1.00	50
g	0.96	0.90	0.93	50
h	0.91	0.96	0.93	50
i	1.00	1.00	1.00	50
j	0.89	0.98	0.93	50
k	1.00	1.00	1.00	50
l	1.00	0.86	0.92	50
m	0.36	1.00	0.53	50
n	0.00	0.00	0.00	50
o	0.94	1.00	0.97	50
p	0.79	1.00	0.88	50
q	1.00	1.00	1.00	50
r	1.00	1.00	1.00	50
s	0.94	0.96	0.95	50
t	0.00	0.00	0.00	50
u	1.00	1.00	1.00	50
v	1.00	1.00	1.00	50
w	1.00	1.00	1.00	50
x	1.00	1.00	1.00	50
y	1.00	0.88	0.94	50
z	0.88	1.00	0.93	50
accuracy			0.90	1300
macro avg	0.87	0.90	0.88	1300
weighted avg	0.87	0.90	0.88	1300

Figure 5 Classification Report

Figure 5, the classification report where most classes achieve high scores, however a few classes, like m, n, and t, show misclassifications, with F1 scores as low as 0.53 and 0.00, indicating areas for improvement. Overall, the model achieves an impressive accuracy of 0.90 on a dataset of 1300 samples. The weighted average for precision, recall, and F1-score stands at 0.87, 0.90, and 0.88, respectively, showcasing strong performance. This project demonstrates significant proficiency in sign language recognition using Random Forest classification and MediaPipe hand landmarks. While the confusion matrix shows some misclassifications for certain gestures, the system reliably identifies most of the 26 gestures with high precision and recall.

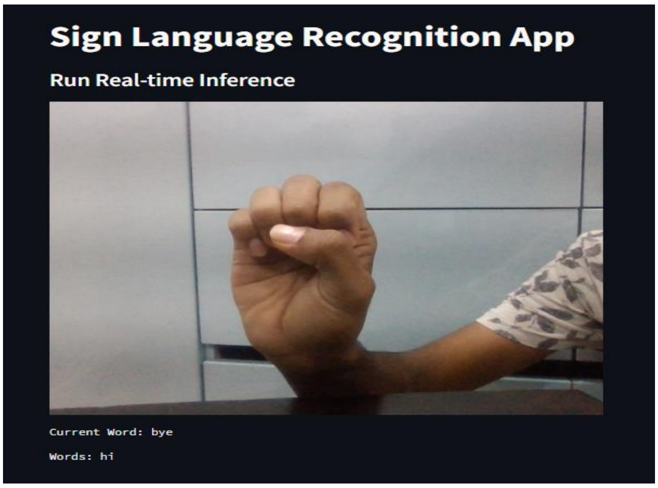

Figure 6 Real-Time Interface

Figure 6 demonstrates the working of sign language recognition interface of the application in real time, which is implemented with a goal to convert sign language gestures into text or spoken words. The interface uses a video camera, translating hand gestures into recognized signs, and presenting them immediately on the screen. This facilitates effective communication between sign language users and non-signers, improving accessibility and engagement.

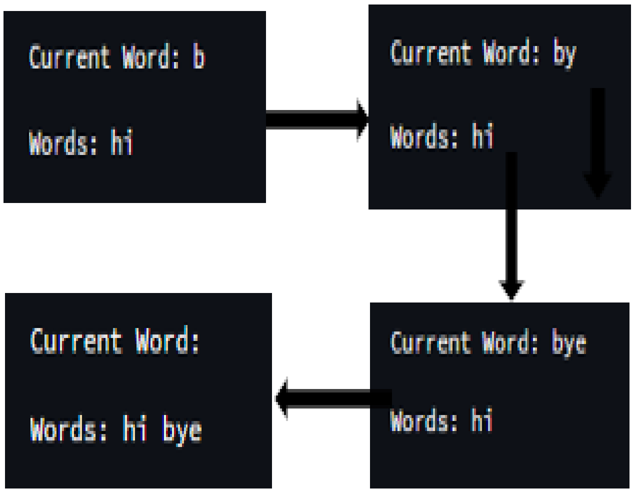

Figure 7 Real-Time Sign Language Recognition with Sentence Construction

Figure 7, this example demonstrates the functionality of the sign language recognition system, particularly its capability to process real-time hand gestures and construct meaningful sentences dynamically. The

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2471 - 2481
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0368

"current word" field represents the word currently being recognized from live gestures, showcasing the system's incremental prediction capability. For instance, as the user signs letters, the system identifies partial or complete words such as "b," "by," and "bye." Simultaneously, the "words" field accumulates previously recognized words to construct a coherent sentence. Initially, the system recognizes "hi" and appends it to the "words" field. After fully recognizing the word "bye," it is added to "words," resulting in the sentence "hi bye." The "current word" resets after a word is fully recognized and transferred to the "words" field, ensuring that the system continuously processes new input gestures while retaining the history of previously recognized gestures. Looking ahead, the model's success in gesture recognition serves as a solid foundation for expanding into real-time applications, dynamic gesture recognition, and integration into practical devices like mobile apps or wearables. While the model demonstrated perfect performance in this controlled dataset, further testing on unseen data can be performed to confirm its generalizability and robustness.

Conclusion

The suggested sign language recognition system aims to be an integration of machine learning, computer vision and its real time processing for effective intervention of the barriers in the social interaction which is faced by the deaf and polluted hearing. By leveraging MediaPipe Hands for hand landmark detection and a Random Forest Classifier for gesture classification, the system achieves exceptional accuracy of 90.00%. This robust framework enables precise recognition of the 26 English alphabets, offering a practical and reliable solution for real-world applications. The data collection process emphasizes diversity and robustness, incorporating variations in lighting, orientations, and augmented samples to ensure adaptability in real-world scenarios. Real-time inference, supported by MediaPipe's hand tracking and temporal smoothing techniques, facilitates dynamic word and sentence construction with remarkable stability and fluidity. This research shows how computer vision and machine learning can reshape communication interfaces and make people and communities understood by others. Besides addressing immediate challenges facing mainstream sign language users, the system lays a foundation for future improvements. By advancing the capabilities of sign language recognition, this project contributes to a more inclusive and accessible society, reinforcing the critical role of technology in breaking linguistic and physical barriers.

Acknowledgements

We would like to express our sincere gratitude to our institute, Hyderabad Institute of Technology and Management (HITAM), for providing the necessary resources and support throughout the course of this project. We are especially thankful to Dr. Hemanth for his valuable guidance, constant encouragement, and insightful feedback, which greatly contributed to the successful completion of our work.

References

[1]. Praiselin, E & Manikandan, Dr & Veronica, Vilma & Hemalatha, Ms. (2024). Sign Language Detection and Recognition Using Media Pipe and Deep Learning Algorithm. International Journal of Scientific Research in Science and Technology. 11. 123-130. 10.32628/IJSRST52411223.

[2]. Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, Sign Language Recognition: A Deep Survey, Expert Systems with Applications, Volume 164, 2021, 113794, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113794

[3]. Bhamare, Pooja & Kumbhakarna, Aishwarya & Shinde, Amruta & Take, Prof. (2024). sign-language Translator using machine learning. Interantional journal of scientific research in engineering and management. 08. 1-5. 10.55041/ijsrem38427.

[4]. Wadhawan, A., Kumar, P. Deep learning-based sign language recognition system for static signs. Neural Comput & Applic 32, 7957–7968 (2020).

[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.

[6]. B. Joksimoski et al., "Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities," in IEEE Access, vol. 10, pp. 40979-40998, 2022, doi: 10.1109/ACCESS.2022.3161440.

[7]. Pigou, L., Dieleman, S., Kindermans, PJ., Schrauwen, B. (2015). Sign Language Recognition Using Convolutional Neural Networks. In: Agapito, L., Bronstein, M., Rother, C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science (), vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40

[8]. Barbhuiya, A.A., Karsh, R.K. & Jain, R. CNN based feature extraction and classification for sign language. Multimed Tools Appl 80, 3051–3069 (2021). https://doi.org/10.1007/s11042-020-09829-y

[9]. Gianluca Amprimo, Giulia Masi, Giuseppe Pettiti, Gabriella Olmo, Lorenzo Priano, Claudia Ferraris, Hand tracking for clinical applications: Validation of the Google MediaPipe Hand (GMH) and the depth-enhanced GMH-D frameworks, Biomedical Signal Processing and Control,Volume 96, Part A, 2024, 106508, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.106508.

[10]. Ali Erol, George Bebis, Mircea Nicolescu, Richard D. Boyle, Xander Twombly, Vision-based hand pose estimation: A review, Computer Vision and Image Understanding, Volume 108, Issues 1–2, 2007, Pages 52-73, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2006.10.012.

[11]. Cheok, M.J., Omar, Z. & Jaward, M.H. A review of hand gesture and sign language recognition techniques. Int. J. Mach. Learn. & Cyber. 10, 131–153 (2019). https://doi.org/10.1007/s13042-017-0705-5

[12]. M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

[13]. Becky Sue Parton, Sign Language Recognition and Translation: A Multidisciplined Approach from the Field of Artificial Intelligence, The Journal of Deaf Studies and Deaf Education, Volume 11, Issue 1, Winter 2006, Pages 94–101, https://doi.org/10.1093/deafed/enj003

[14]. C. Suardi, A. N. Handayani, R. A. Asmara, A. P. Wibawa, L. N. Hayati and H. Azis, "Design of Sign Language Recognition Using E-CNN," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Surabaya, Indonesia, 2021, pp. 166-170, doi: 10.1109/EIConCIT50028.2021.9431877.\

[15]. Zhao, X., Song, Z., Guo, J., Zhao, Y., Zheng, F. (2012). Real-Time Hand Gesture Detection and Recognition by Random Forest. In: Zhao, M., Sha, J. (eds) Communications and Information Processing. Communications in Computer and Information Science, vol 289. Springer, Berlin,Heidelberg. https://doi.org/10.1007/978-3-642-31968-6_89

[16]. Koller, O., Zargaran, S., Ney, H. et al. Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs. Int J Comput Vis 126, 1311–1325 (2018). https://doi.org/10.1007/s11263-018-1121-3

[17]. A. S. Nikam and A. G. Ambekar, "Sign language recognition using image-based hand gesture recognition techniques," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/GET.2016.7916786.

[18]. G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing and Communication Engineering Systems (SPACES), Vijayawada, India, 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.

[19]. 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, doi: 10.1109/BigData.2018.8622141.

[20]. Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, Indian Sign Language recognition system using SURF with SVM and CNN, Array, Volume 14, 2022, 100141, ISSN 2590-0056, https://doi.org/10.1016/j.array.2022.100141.

[21]. M. A. Rahaman, M. Jasim, M. H. Ali and M. Hasanuzzaman, "Real-time computer vision-based Bengali Sign Language recognition," 2014 17th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2014, pp. 192-197, doi: 10.1109/ICCITechn.2014.7073150.

[22]. N. Buckley, L. Sherrett and E. Lindo Secco, "A CNN sign language recognition system with single & double-handed gestures," 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2021, pp. 1250-1253, doi: 10.1109/COMPSAC51774.2021.00173.

[23]. J. J. Raval and R. Gajjar, "Real-time Sign Language Recognition using Computer Vision," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 542-546, doi: 10.1109/ICSPC51351.2021.9451709.

[24]. T. Li, Y. Yan and W. Du, "Sign Language Recognition Based on Computer Vision," 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2022, pp. 927-931, doi: 10.1109/ICAICA54878.2022.9844497.

[25]. Bird, J.J.; Ekárt, A.; Faria, D.R. British Sign Language Recognition via Late Fusion of Computer Vision and Leap Motion with Transfer Learning to American Sign Language. Sensors 2020, 20, 5151. https://doi.org/10.3390/s20185151