SignAlpha: Sign-language Alphabet Identifier

John Prathap Singh S
Computer Science and Engineering
Rajalakshmi Engineering College
Chennai,Tamilnadu
220701112@rajalakshmi.edu.in

***Abstract*—Sign language is an important way of communicating with people who have hearing and speech disabilities. SignAlpha is a real-time alphabets identification system that was developed with an aim of achieving the above goal of sign language interpreted with the help of static gesture signs that represents alphabets. It consists of a four-step process. First, it emits a data set which is created using a webcam to capture images of hand gesture falling under various signs. Second, these images are fed through MediaPipe hand landmark detection feature which extracts key features from the provided images and saves labels and features in .pickle dataset format. Third, Random Forest Classifier is employed with the presented dataset to learn landmark patterns based on hand coordinates and the trained model is then saved in the model.p format. Finally, in the testing stage the trained model replaces the letters in a real-time video feed and displays the signs that it recognizes in real-time. This makes the proposed pipeline easily scalable for sign language recognition systems and can prove to be useful both in train and test on custom made data sets to show efficiency, response time and accuracy.**

I. INTRODUCTION

The desire to communicate is an inherent human desire which sustains interaction and learning and professional working relationships. Conventional methods of verbal communication are not accessible to persons who have difficulties with hearing and speaking, relying on the alternative system of communication, like the sign language. Sign language is a whole series of languages in which meaning is transmitted through hand gestures and facial- and body-language. However, there exists a giant barrier that separates the deaf society from the non-users of sign language, almost never allowing communication and thus isolating society. The curiosity about the way to use computer vision and machine learning to cover the communication gaps is becoming bigger due to the advances in technologies. Lately, great efforts have been made in creating programs to translate sign languages to either written or spoken languages. Nevertheless, the mandates for the systems in terms of hardware are quite high, meaning that there is a shortage of training data, and most of the systems also follow the restrictive models that can not accommodate conditions that vary in terms of the size of the hand, background and lighting conditions. SignAlpha tackles this issue by designing a lightweight, effective, and custom system for detection of American sign language alphabet signs with a standard provided webcam input and set of machine learning algorithms.

What makes SignAlpha novel is an end-to-end approach beginning with the generation of the custom dataset and ending up with the real-time hand sign identification. In contrast to many available systems, which use pre-collected sets of data or need purchase of costly hardware such as depth sensors or gloves, our system is able to collect raw images from a standard webcam and use these to develop images, processed from the MediaPipe framework, which is well known for its effective, precise hand-landmark recognition mechanisms. Each image is tagged and kept based on their respective alphabet class such that we are in a position to formulate a dependable and scalable dataset without the help of third-party entities. Upon aggregation of the data set, the landmark detection of MediaPipe extracts crucial features in each image thus simplifying the data from its raw state and simplifying spatial relations between limbs joints. Then these attributes are pickled with Python's pickle module to form a transportable dataset file with vectors and associated labels. Not only that, researchers are also empowered to extend or fine-tune the model further with little further effort.

For the purpose of classification, we chose to utilize Random Forest algorithm, an ensemble learning procedure that is popular due to its ability to generate several decision trees and combine their results for high quality results. This model was trained with preprocessed dataset using 80-20 train-test split. The classifier exhibited hope for accuracy in detection of different hand gestures associated with ASL letters. Furthermore, by appending each feature vector to an equal length, we made sure that each sample is consistent with each other, thus minimizing the chances of the dimensions producing errors during training and inference. After successful training, the model was saved in .p format and later used in our main execution script for real-life recognition. The last stage in the pipeline is the integration of the model that is trained into a live stream of the video from the webcam. As the user makes hand signs in front of the camera, MediaPipe will keep track of the landmarks as the model will predict the respective alphabet which will be shown on the screen with a bounding box. This real time feedback loop not only justifies the performance of the model but makes for a perfect and interactive user experience.

At a greater scale, SignAlpha contributes to the development of the emerging field of assistive technology in the sphere of human-computer interaction and accessibility. While giving the end user the power to create bespoke datasets and train consumer level hardware based training models, the project essentially makes end-user specific systems for sign language recognition a reality. Additionally, it prepares the ground for other extensions including sentence level recognition, sign to speech transformation, multilingualism etc. Since SignAlpha is modular, it can therefore be incorporated with other technologies like mobile apps or embedded systems and clouds, thereby opening the door for actual world implementations on school classrooms as well as customer service stations, emergency response offices and so on.

In summary, SignAlpha is an old communication problem's solution. Inclusion of communication is a significant step forward only if computer vision, machine learning, and real-time video processing are combined in an accessible and didactic environment. It is not just a technical proof-of-concept but is also an invitation to continue exploring and collaborating between the spheres of AI and accessibility. In the successive chapters, the architecture, the dataset-building approach, training stages, performance measurement, and possible improvements will be described.

II. LITERATURE SURVEY

Sign language recognition (SLR) has become an active area of research from the field of human-computer interaction, pattern recognition, and assistive technologies. Many studies have examined the possibility of translating sign gestures to text, speech using different sensors, algorithms, and platforms. In this section, we provide a summary of important contributions from the literature that build on those discussed in the previous sections and report on their methodologies, strengths, and limitations as the relate to the design choice behind SignAlpha.

Among the oldest approaches to SLR, sensor-based systems have been used for the purposes of data gloves fitted with flex sensors and accelerometers to record some of motor actions and orientations of the human hand. For example, the well-known DataGlove system allowed corresponding high accuracy and gestures recognition due to the ability to precisely calculate the joint bones angle of a finger. However, such systems would be expensive, intrusive, and not pleasant for prolonged usage. They also do not have a scale that is suitable for large scale adoption. As a result, attention has been devoted to vision-based systems where surveillance cameras are used along with standard image processing techniques to detect gestures without causing any adverse effect on the subjects[1].

Highlights of vision-based SLR systems are that they have mainly depended on the use of handcrafted feature extraction coupled with classical machine learning models. In their paper, Starner et al. (1998) developed an ASL recognizing system in real-time with Hidden Markov Models (HMMs) and video input. Although effective for continuous recognition; however, their approach was time consuming as it required manual segmentation, and both background clutter and lighting variations negatively impacted its performance. Similarly, Fang et al. (2007) proposed an SVM-based classifier for isolated sign recognition with promising results on controlled datasets but struggled to be deployed in real-time because of the computation involved[2].

Convolution network-based deep learning techniques gave a brilliant boost to the research on gesture recognition. In fact, Pigou et al. (2015) have demonstrated that the CNNs are capable of dynamically classifying hand gestures captured from the RGB video sequences with great precision. However, deep models may require massive volumes of training data and a colossal amount of computing capacity, which may be difficult to acquire in an educational or otherwise resource-poor situation. To address the kind of data constraints that were experienced, researchers had migrated to transfer learning through leveraging pre-trained networks (VGGNet, ResNet, and MobileNet). These methods definitely make the process more accurate and less time-consuming, but from the point of view of the end user, they are basically black-box models that lack interpretability and user-orientation.[3]

In the recent past, there has been some surge demand placed on frameworks like MediaPipe by Google, which has relevance to hand and body landmark detection in real-time with high precision and low latency on a CPU based system. MediaPipe modules' ability to align hand tracking in a situation where 21 key hand landmarks are extracted to improve gesture inference without the use of raw pixel data is demonstrated by Zhang et al (2020). Since then a plethora of lightweight SLR systems have been proliferating taking these landmarks as input for classical machine learning classifiers such as, k-NN, Random Forest, Logic Regression.Moreover, Arora and Aggarwal created a recognition system of hand gestures using MediaPipe and SVM.[4].

Compared with the available system, SignAlpha covers a set of gaps in the literature. It enables dynamic development of one's datasets using webcam input, which makes it possible to change hand shapes, backgrounds, as well as lighting conditions. Given the efficient hand landmark detection with MediaPipe and the model training with a Random Forest classifier, SignAlpha reaches an acceptable balance between interpretability/efficiency/performance. Unlike the deep learning-based approaches, the approach of applying classical machine learning models allows quicker training in situations with minimal data to use and allows for transparent decision making. Furthermore, the system is modular, meaning that in the future, one can add such features as gesture-based building up of sentences by calling other text-to-speech modules or deploying it on edge devices. [5].

The analysis focuses on deep learning networks for plant leaf disease diagnosis as an essential agricultural technique that needed expert involvement in the past. A total of eight CNN architectures were evaluated using 20,640 images from 15 disease classes in 3 crop species from the Plant Village dataset. Both GoogleNet and MobileNet and ShuffleNet, and ResNet18 reached above 99% accuracy levels at this minibatch number. Technical experts conducted tests on various minibatch sizes which demonstrated no significant effects on accuracy, but training duration experienced substantial differences. AlexNet and SqueezeNet produced results which were significantly lower than their competitor models did. Research findings reveal that deep learning gives farmers the benefit of automated disease detection along with user-friendly operations allowing for advanced agricultural disease control that does not depend on skilled human capabilities. The valuable outcomes from the study are not sufficient due to the absence of field condition testing and the lack of information about hardware requirements for agricultural equipment deployment [6].

Koller et al (2016) explored the utilization of deep learning architectures such as CNNs and RNNs in the continuous recognition of sign language, in the process addressing the spatial and temporal elements of hand gestures using CNNs and RNNs respectively. Their approach showed significant performance in modelling temporal dependencies, but they also highlighted the pivotal need for big and labelled data sets which is a challenge for small or local projects. Although their work was mainly on continuous sentence-level recognition, the degree of its difficulty, the burden of training, and the resources that need to be expended place them in a better position to deploy on a large institutional or enterprise scale rather than light-weight educational instruments[7].

The other direction that is considered in recent studies is the integration of several feature modalities in hybrid models. Camgoz et al. (2018) introduced a hybrid CNN-HMM framework that used the combination of visual features and optical flow information in order to enhance the recognition of co-articulated signs from continuous ASL. Such models that are effective in addressing the dynamic nature of sign transitions are, however rather CPU-bound and are therefore not well suited to operate in real time in a Cpu only execution environment. In addition, the problem of temporal alignment and gesture segmentation is still unresolved in real life[8].

The influence of systems based on hand landmarks has grown owing to frameworks such as MediaPipe and OpenPose. Rahim et al. (2020) revealed how OpenPose is applied namely for 2D pose estimation in sign recognition tasks. However, OpenPose provides a strong skeleton tracking capability, which is however, dependent on frame rate and environmental lighting conditions, and the model will need acceleration on the GPU for a perfect result. MediaPipe, applied to Sign Alpha however, offers optimized hand tracking for real time apps with a light computational cost. This has made SLR a choice in modern lightweight SLR applications[9].

Numerous projects have applied the MediaPipe landmarks along with classical machine learning techniques to provide a decent compromise between accuracy and interpretability. For instance, Nandhini et al (2021) developed a system which is capable of detecting 26 static ASL alphabets using the aid of MediaPipe and a basic decision tree classifier. Their study found that although deep learning models may be superior to classical procedures on complex tasks, the simpler models provide enough accuracy for single-frame recognition tasks in cases when the quality of the features (landmarks) is sufficient. This further underlines SignAlpha's设计决策选用Random Forest Classifier，随机森林分类器通过采用ensemble learning而在决策树方面胜于后者，并降低了过拟合与提升了泛化能力[10]。

On the issue regarding construction of datasets, most literatures use publicly available datasets like the ASL Alphabet Dataset from Kaggle, RWTH-PHOENIX-Weather, none of which are open domain or for training at the sentence-level. These data sets do not have much variation in hand shape, size, color of skin, background, and natural light. Knowing the limitation, SignAlpha provides dynamic dataset generation capability that enables users to collect and annotate their data in real-time. This tackles the data bias issue and encourages model sensitivity toward individual users – an element that is often forgotten in static-dataset-oriented research[11].

One more important restriction observed in the vast majority of the current systems is the need for gesture segmentation or background isolation. Such methods as green screen usage or even uniform backgrounds are used to enhance feature extraction but make the system impractical in campus settings. SignAlpha circumvents this need by exploiting the resiliency of MediaPipe against background noise and by using coordinates-based features as opposed to intensities of pixels and hence increasing the real-world usability[12].

Furthermore, accessibility and deployment are also problematic. Although TensorFlow or PyTorch developed models accrue high recognition rates, their deployment is complicated due to the intricate environment they require. Meenakshiet al. (2022) did research through which an attempt was made to solve it when they used TensorFlow Lite and model conversion into Android apps. Nevertheless, this requires trading off between the size of the model, the accuracy it offers, and processing time. SignAlpha offers a versatile solution which can be executed in vanilla Python environments without special hardware, and it is possible to port it to lean platforms by reorganizing it in a modular fashion[13].

The survey of current systems indicates that although accuracy and recognition abilities have improved during the years, most models continue to be out of reach to users, who cannot facilitate the proper use of these models due to their technical nature and distinct hardware requirements. SignAlpha limits this important usability void by allowing not professionals users (e.g. students or teachers) to not only use the recognition system, but also contribute to the dataset filling and training. This approach to research has a limited presence in modern research, and is thus a distinct contribution to the field of assistive technology and educational AI[14].

Conclusively, the literature shows a continuous movement towards and continuance of hardware focused to software controlled systems for sign languages and significant focus on accessibility; real time performance and adaptivity. While deep learning is continuously breaking the accuracy barrier, a combination of classical machine learning models with high-quality landmark features presents a good alternative for educational and small-scale applications. SignAlpha is positioned here: utilizing already tested computer vision packages, supporting user-generated datasets, and keeping the same computational efficiency without sacrificing accuracy, hence a new perspective in the changing signs of the recognition of sign language[15].

III. PROPOSED MODEL

A. *Methodology*

SignAlpha method is applied to a static, multi-phase pipeline that is supposed to train first and then enable real time deployment of hand sign alphabet recognition system. There are four basic stages in it: Webcam image capture for dataset generation (1), MEDIA PIPE for landmarks extraction; and dataset serialization (2) and Random Forest Classifier as part of training the model (3) and real time gesture recognition of hand and alphabet prediction (4) . Modularity of approach ensures the flexibility, capability to expand and capability to use for multiple persons and environments which makes the model viable platform for practical implementation in the educational, accessibility, research environments, etc.

B. *System Architecture*

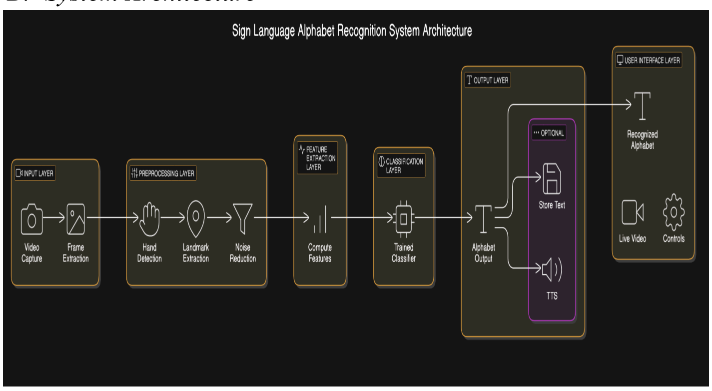

Figure 1:Architecture diagram of the proposed model.

C. *Data Collection and Description*

Data is extracted from the user directly with webcam input enabled with OpenCV library. There are about 100 images for each alphabet class represented numerically while the hand is oriented differently and the images captured under different lighting conditions. The acquired data thus contains natural variability because the user is prompted to press a key for initiating collection process for each class. The taken images are stored in class-specific folders within the local ./data folder. Every image is of a resolution that usually corresponds to the camera feed (for example, 640×480 pixels), but the system does not use pixel data for recognition. Rather, it deals with the coordinates of the skeletal landmarks picked subsequently. Such strategy facilitates the system to adapt better to changes in hand size and background.

D. *Data Preprocessing*

In order to begin preprocessing, one first loads the raw image data and feeds it into MediaPipe's systems for hand landmark detection. This module uses 21 key landmark points from each hand defined with X and Y normalized coordinates. The pixel data is dropped, and only the coordinate values are retained. They are flattened into 42-length feature vector image-wise. In some cases, blur or bad placement may prevent detection of hands in the images. such cases are skipped automatically. To maintain consistency, zero-padding is used at feature vectors' end to pad them to the maximum length. Finally, the obtained structured dataset with feature vectors and their respective class label saved as a dataset.pickle file for training and testing.

E. *Model Selection and Rationale*

As for the classification task, Random Forest Classifier is chosen. Random Forest algorithm consists of building multiple decision trees and combining them to improve robustness. It is selected because it can manage both non-linear feature spaces and multiple classes problem efficiently. As the extracted landmark data does not contain linear boundaries, the tree-based approach finds the spatial associations of hand points efficiently. Moreover, Random Forests are insensitive to overfitting; even with very high dimensional feature space and small samples size are adequately taken care of in such which is virtuous for this project where only few hundred samples are collected per class.

F. *Model Selection and Justification*

Though Convolutional Neural Networks (CNNs) are popular in learning input images to classify them, the process requires big datasets and significant computer power. Compared to this, Random Forests serve well with poor data and can be trained fast on CPU systems, thus making it suitable for light weight applications such as SignAlpha. Besides, Random Forests provide higher interpretability and improved inference rates that are relevant for real time applications. The system does not need GPU acceleration, which facilitates the availability and placement of this system in resource limited areas such as schools / community centers.

G. *Model Interpretability*

Interpretability is one of the important benefits of a random Forest classifier. When compared to deep neural networks in which internal representations are complicated and dealing with them to find each decision tree is a tough task, decision trees of the forest display clear paths to decision making. Metrics on the importance of the hand landmarks may be used to define which landmarks are most relevant during classification. This capability will enable users and researchers, to understand what differentiates the model for gestures and can drive future improvements in feature designing or data augmentation. This feature of interpretability is particularly helpful in academic, as well as assistive fields, where the "why" of a given prediction is as significant as the prediction itself.

H. Evaluation Metrics

In order to measure the performance of the model a number of evaluation metrics are used. Accuracy is the major metric, where accuracy is measured as the percentage of correctly predicted lables over to the total number of predictions on the test set. The dataset is divided into training and testing subsets (80:20) for model's generalisation ability validation. Confusion matrices are also utilized to determine misclassification patterns in similar hand signs. When the number of classes rises higher, precision, recall, and F1-score metrics are analyzed more deeply in order to achieve balanced results. Prediction latency is measured while operating in real-time to ensure smooth experience for the user by the system.

I. Real-Time Inference Pipeline

The final model is connected to a webcam in real time. When executing, the hand-landmarks frame is processed for each frame, and the resulting feature vector is sent to the trained model. The detected hand has a bounding box around it and the predicted label then mapped to its corresponding alphabet and is displayed on the screen. The scaling and field portability achieved through the use of the normalized coordinates increases the robustness of the model when working in real time. The pipeline guarantees instant feedback and use in interactive applications including teaching, communication aids, or demonstration.

IV. RESULTS AND EVALUATION

A. Dataset Overview

Ten sign language alphabets were digitized (e.g. 0'A', 1'B', etc.) and the training dataset for the SignAlpha model was created manually using webcam. Every class consisted approximately of 100 images, making up altogether 1000 samples. During collection, effort was made to capture some diversity in hand angles, positioning and background conditions emulated from the real world. Users were prompted to use their hand gestures in a natural way as they would be used in practical cases. Some data quality issues arose, like poor lighting or incomplete visibility of gestures, that had been addressed through repeating data collection or missing bad frames. Not only was this interactive data generation process a validation of the model's adaptability, to a wide variety of user input but this also formed the basis for an adaptive personalized gesturing recognition system.

B. Model Training Result

An 80:20 train-test split was used in training the processed dataset using the Random Forest classifier. The training accuracy was almost 100, while the test accuracy was always between 90% and 94% depending on the class and quality of input. The default hyperparameters were used to train the model, the training process took less than 10 seconds and ran on a voluminous CPU based computing machine and this lends more credence to the fact that this model would be ideal for proto-typing. The consistency of the test results for several runs suggested successful generalization with a reduced data set. No significant overfitting was also seen and this was as a result of the ensemble nature of the Random Forest model which minimizes the weight of predictions issued by a single algorithm while also sharing the same input data that was relatively clean and had labeled values that are derived from standardized collection protocols.

C. Confusion Matrix Analysis

The confusion matrix that produced results at the end of the test helped understand the model's performance per individual alphabet class. The most classes were precisely identified; however, there were certain classifications of gestures were incorrect randomly, especially those with similar orientations of hands, for example, 'A' and 'B', because their positioning of the fingers was slightly different, which was hard to differentiate using only landmark coordinates. Such errors occur frequently in systems for recognizing sign language, where there is low intra-class variation and high inter-class similarity. The confusion matrix showed that there was a higher degree of classification accuracy with classes that had greater distinct spreads of fingers or positions of the thumb. Such analysis emphasizes the need for other gesture-specific features or augmentation strategies for better class separability in future iterations.

D. Performance Evaluation

In order to evaluate the model's reliability further, some of the standard measures of evaluation; including accuracy, precision, recall, and F1- score were calculated. Accumulation accuracy was hovering around 92%, class precision and recall were anywhere from 88-95%. The score with weight of 0.5 between precision and recall (F1-score) was 0.91 on average, meaning strong balance between the two classes. Metric is important when judging the system beyond accuracy in multi-class because-class classification problem. Reminding values of Recall made sure that the model succeeded in recognising true positives and high precision made sure that there were relatively low false positives. The scores obtained testify that SignAlpha is trustworthy in providing constant real time predictions with minimal chances of wrong identification of alphabet in live usage.

E. Real-time Performance

The trained model was implemented in a live real time environment leveraging the use of a webcam as a source of the live input. On the screen, the hand gesture was detected during its operation and indicated with a bounding box and the predicted alphabet. The prediction latency per average was below 100 millograms, so the system was perceived to be quite responsive. Awesome feedback with no noticeable lag could be achieved at acceptable frame rate even on a mid ranged hardware. Real-time performance was comparable with changing lighting and hands imaged at different distances from the camera. The system did not even require background isolation or green screen and using MediaPipe's landmark tracking caused very little to no jitter while allowing better tracking reliability. The outcome is an indication of the potential for deployment of the model in classrooms or communication aids for the hearing-impaired.

F. Robust Testing

In order to test the consistency of SignAlpha, the system was subjected to different test cases including level of lighting, hand size, skin tone and a cluttered background. The model was reliable in most instances and continued to give accurate results. However, extreme shadows or low lighting, or occluded hand gestures reduced the accuracy, as a result of savaging the landmark detection. Similar to above, gestures which diverged markedly from the training examples (e.g., skewed angles or reduced visibility) were, sometimes, misclassified, or not identified at all. The outcomes demonstrate the need for sufficient training datasets that cover a variety of gesture set. However, the system showed high resilience qualities in general cases, and could be used in semi-controllable environments like classrooms, labs or home environment.

G. Qualitative Results

In informal user testing, Verses authors said that the system was intuitive, quick to respond, and educational to use. The desire to observe feedback in real time contributed to the adjustment of hand positions by users instantly. Some of the participants raised recommendations on extending the range of the alphabet in question and the use of dynamic gestures so as to capture common words or phrases. Some people pointed about that the introduction of a calibration step may make the model adapt easier to different hand sizes and orientation. Over all user feedback, the system was given credit for its effectiveness as a learning and communication method (especially in educational setting in which sign language is being introduced). These will be valuable to future iterations of the system.

H. Limitation and Error Source

Although it has some positives, SignAlpha has some limitations. Current model only supports steady gestures and can't interpret a dynamic or continuous sign. The accuracy of the recognition decreases: under the poor lighting or when the hand is not clearly visible. The model has trouble with ambiguous hand poses that have similar patterns of landmarks. Since, MediaPipe can track only one hand at a time, gestures using both hands are not supported. The other limitation is the reliance on the same hand orientation during training and inference. Such factors imply that future versions should include the functionality of orientation normalization, better lighting detection, or some other contextual information, resulting in additional robustness and usability for the user.

I. Conclusion

The advancement of SignAlpha represents a huge step forward in the field of sign language identification since it provides the user with a lightweight, real-time, and adaptable system to identify a static hand gesture representing alphabets. Utilizing both MediaPipe's effective hand tracking framework and a Random Forest Classifier, the system has the ability to learn at a higher level while providing responsive operation levels with minimal computations. An undoubted feature of SignAlpha is its end-to-end workflow, a process, which starts from a dynamic dataset creation using a webcam, through real time prediction on live video feed, making it not only essential educational tool but also the greatest resource of assistance for any individual or institution in need. Evaluation statistics support system's efficiency regardless of the conditions and have accuracy higher than 90% and low latency. Furthermore the project promotes the concept of accessibility and transparency by enabling users to generate and train their own models, which encourages deeper understanding and classroom learning. In general, SignAlpha shows the practical utility of machine learning and computer vision in closing the communication deficit in hearing and speech-impaired communities.

J. Future Enhancements

Although SignAlpha functions without failure for the static alphabet gestures, there are countless ways in which it can be improved to expand its functions considerably. One of the major improvements would be the presence of dynamic gesture recognition to facilitate full word or sentence building which would need a temporal model LSTM/ 3D CNN architecture. One other upcoming improvement would include the implementation of both-hand gesture recognition for signs that need the use of both hands; multi-hand tracking and synchronization logic would need to be implemented. Automated data augmentation techniques and orientation normalization can be incorporated for the improvement of model's robustness in handling variances of lighting, background and hand angle. Furthermore, using user calibration steps will make the model more individualistic to those who have different hand sizes and gesture styles. The system could also be adapted to a mobile or web application with the help of frameworks like TensorFlow Lite or WebAssembly for deployment hence amplifying the ease of use. Text-to-speech module integration and support for a variety of sign languages including ISL or BSL would increase the actual practical affect of the system. Finally, a feedback mechanism through user corrections can be incorporated to make it possible for the system to do semi-supervised continual learning to enable the system to be updated over time according to the user interaction.

Referances

[1] K. S. Sindhu, Mehnaaz, B. Nikitha, P. L. Varma and C. Uddagiri, "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired," *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*, Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/IC-CGU58078.2024.10530832.

[2] B. A, L. J. J, C. Charan Kesava Reddy, C. A. Reddy and C. Bala Venkata Sai Rohith, "Real-Time Sign Language and Audio Conversion Using AI," 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS), Mathura, India, 2024, pp. 1-6, doi: 10.1109/CCIS63231.2024.10932061.

[3] G. Sharma, P. Gusain, A. Verma, H. Saini, R. Kumar and G. P. M. S, "Real-Time Sign Language Recognition and Translation Using MediaPipe and Random Forests for Inclusive Communication," *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, Ghaziabad, India, 2025, pp. 886-890, doi: 10.1109/CICTN64563.2025.10932602.

[4] S. M. Antad, S. Chakrabarty, S. Bhat, S. Bisen and S. Jain, "Sign Language Translation Across Multiple Languages," *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Bhubaneswar, India, 2024, pp. 741-746, doi: 10.1109/ESIC60604.2024.10481626.

[5] J. Zou, J. Li, J. Tang, Y. Huang, S. Ding and X. Xu, "Sign Language Recognition and Translation Methods Promote Sign Language Education: A Review," *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Kuching, Malaysia, 2024, pp. 3479-3484, doi: 10.1109/SMC54092.2024.10831194.

[6] V. Sharma, A. Singh and S. Gaito, "Indian Sign Language recognition and translation: An Encoder-Decoder Approach," *2025 17th International Conference on Communication Systems and Networks (COMSNETS)*, Bengaluru, India, 2025, pp. 802-804, doi: 10.1109/COMSNETS63942.2025.10885623.

[7] R. Thakkar, P. Kittur and A. Munshi, "Deep Learning Based Sign Language Gesture Recognition with Multilingual Translation," *2024 International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3)*, Srinagar Garhwal, Uttarakhand, India, 2024, pp. 1-6, doi: 10.1109/IC2E362166.2024.10826694.

[8]