Sign Language to Speech

Sakam Rani, Yarlagadda Rakesh, B Hannah Havilah & Yannam Hema

ABSTRACT

The sign language to speech project is to develop a system that translates gestures from sign language into spoken Telugu using the Random Forest algorithm. It can be challenging for non-speakers to use sign language, which is the primary form of communication for those who are dumb or mute. Our intention in creating this technology is to bridge the communication gap between non-sign language users and sign language users.

Using a machine learning technology called the Random Forest algorithm, sign language motions will be properly recognized and interpreted. As a result, the system will generate Telugu speech output, which indicates that our model will generate a Telugu voice. This will enable speakers and sign language users like dumb or unable to speak people to communicate in real time. This project aims to promote inclusivity and accessibility for people who cannot speak for themselves. An existing project that uses An existing models didn't give the appropriate result. So, overcome those drawbacks to creating a optimize model.

The project aims to improve sign language to speech accuracy by integrating gesture recognition and speech synthesis. Using machine learning algorithms like Random Forest, the project aims to provide real-time translation and seamless communication for individuals with dumb or unable to speak, potentially exceeding current accuracy standards.

Keywords: American Sign language (ASL), Speech synthesis, Telugu language, Random Forest algorithm, Communication, Machine learning, Gesture recognition, Real-time translation, Accessibility and Inclusivity

INTRODUCTION

Bridging the Communication Gap:

Sign language is a vital mode of communication for individuals are who can not speak disabilities, enabling them to express themselves and connect with the world. However, for those who do not understand sign language, a significant communication barrier exists. This project aims to bridge this gap by developing a system that translates sign language gestures into spoken Telugu language, fostering inclusivity and accessibility for the dumb people.

Leveraging Machine Learning for Real-Time Communication:

The system will utilize the Random Forest algorithm, a powerful machine learning method, to accurately recognize and translate sign language gestures. By training the algorithm on a substantial dataset of sign language gestures, the system will be equipped to effectively translate sign language in real-time. This real-time translation capability will enable seamless communication between individuals using sign language (who can not speak) and those who can speak, promoting greater understanding and interaction with society.

Enhancing Accessibility and Inclusivity:

This project, which aims to empower the dumb people, comprises significant components. By encouraging inclusiveness and supporting efficient communication, The technology has the ability to enhance social involvement and quality of life for people who are dumb or unable to speak by promoting inclusion and effective communication.

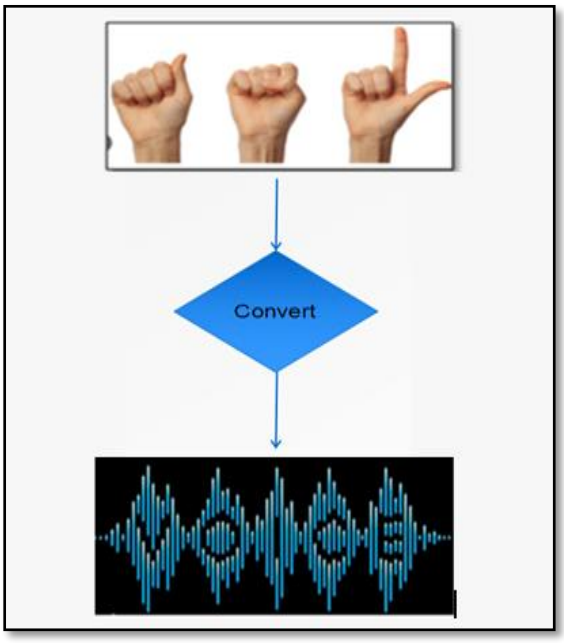

Figure 1: Converting sign language to voice

LITERATURE SURVEY

[1] Akshatha Rani states that the system, The American Sign Language (ASL) dataset, hand tracking methods, and artificial neural network (ANN) architecture are used by the system to identify the alphabet in sign language and translate it into voice. With a 74% accuracy rate, the technology also provides text-to-speech conversion for blind users. By bridging the communication gap between deaf-mute people and others, the suggested approach improves inclusion and accessibility through technology.

[2] Aishwarya Ramesh states that, Understanding which features to extract from static photos is the first step in designing a suitable model. The improved invariance with changes in lighting and shadows is the outcome of this normalization. HOG descriptors are used as the image features for the machine learning algorithm Support Vector Machine (SVM). Therefore, SVM is used to train our model. In this experiment, three distinct array parameters are utilized for SVM, and the outcomes of each are compared. The Detection Method, Kernel, and Dimensionality reduction type are the three array parameters.

[3] Amrutha K states that, The common means of communication for people who are speech and hearing-impaired is sign language. Users can communicate more easily because to the region-specific sign language division. The hearing impaired and others with speech impairments typically rely on human translators because a greater portion of society does not understand sign language. It may not always be feasible to use a human interpreter at a reasonable price or availability. An automated system that can read, understand, and translate sign language into comprehensible form would be the greatest replacement. The communication gap that exists among people in society would be lessened by this translator. For a continuous and fluid sign, the SLR needs be trained with a large amount of sign language data and its syntax.

[4] Salma A. Essam El-Din states that, In order to communicate in daily life, those who are deaf or mute greatly benefit from Sign Language (SL). Instead of using sound patterns to communicate, one might use hand gestures, such as American Sign Language (ASL) or any other SL. In SL, body part movement, orientation, and defined shapes are all done at the same time. The first issue is that most healthy individuals comprehend sign languages very little to nothing. For those who are deaf or mute, effective communication is therefore viewed as a difficulty and barrier in their daily life. When a deaf person converses on one side using sign language (SL) that he is accustomed to and comfortable with, the system converts that SL into sound and pictures that the person can understand.

[5] Ayush Pandey states that, The experiment shows how CNN can be used to solve computer vision difficulties; it can translate sign language with a 95% accuracy in finger typing. By creating datasets and training CNN, it can be expanded to support additional sign languages.

Eliminating the need for interpreters, keeping an eye out for distorted grayscales, and achieving precise prediction while wearing gloves are the major goals.

[6] Ashok Kumar Sahoo states that, The study recommends taking pictures of hand posture, showing text, and improving ISL digit identification with a graphical user interface. Users can add their indications, and the system can forecast outcomes with 100% accuracy. Future studies could investigate classifiers, feature extraction strategies, and combination to build a comprehensive ISL recognition system with 100% real-time interpretation accuracy of ISL signs as the goal.

[7] Bayan Mohammed Saleh states that, For non-speakers in particular, slide interpretation and speech interpretation are crucial to sign language communication. Inaccurate hand segmentation and gesture prediction can result from dim lighting. Inaccuracies can also result from inaccurate peripherals. Sign language technology development is essential for productivity, professional advancement, and improving social interactions. People who are deaf or dumb will benefit from this breakthrough.

[8] Narayana Dharapaneni states that, The history, composition, and significance of American Sign Language (ASL) among the deaf community are covered in this document, with a focus on the language's rich cultural heritage. In order to foster inclusivity and understanding, it draws attention to the importance of interpreters and promotes ASL education and awareness.

[9] Adithya V. states that, In order to reduce computational load and uncover properties that differentiate hand postures, deep learning has been applied to the task of hand posture identification from raw pictures. Experiments on publicly available datasets proved the efficacy of the suggested CNN architecture by showcasing improved recognition performance in terms of accuracy, precision, and recall.

[10] Ankita Saxena states that, The project's main goal is to employ Principal Component Analysis to create a sign language recognition system that will allow hearing-impaired people to communicate. The system uses an android device or webcam to record live video frames, which it then uses to match static hand motions with a database to generate text or speech commands. With successful recognition rates of about 90% according to test results, it is a useful tool for bridging the communication gap between the general public and deaf and mute people.

Table 1: References papers details

S.NO	Author	Title	Dataset	Algorithm	Merits	Demerits	Accuracy
1	Akshatha Rani	Sign Language toText-Speech Translator Using ML	American Sign Language (ASL) dataset	SVM, CNN.	Converts sign text to speech, aiding comm -unicaton for deaf and blind	It is less accurate	74%
2	Aishwarya Ramesh	Real-time Conversion of Sign Language to Text and Speech.	OpenCV	RBF, PCA	Development of an Android app for real-time ASL conversion.	Less optimization due to underfitting	N/A
3	Amrutha K	ML Based Sign Language Recognisit i on	ASL Dataset	KNN	Vision based isolated hand gesture detection and recognition in the model.	Less Accuracy	65%

4	Salma Essam El-Din, Mohamed El-Ghany	Sign Language Interpreter System	OpenCV	SLR, ASL, ArSL	higher recognition dynamicaccuracy than ASL	System struggles with gesture-s due to Speed and accuracy	88%
5	Ayush Pandey	Sign Language to Text and Speech Translation	OpenCV	CNN	Real-time Needs to N/Asign language translation using CNN for deaf community	Mainly focusing on single gesture transform	95%
6	Ashok Kumar Sahoo	Indian Sign Language Recognitio-n Using Machine Learning	Sign database with 5000 images, 500 for each numeral sign	KNN	Recognition of ISL immobile numeric signs using classifiers	Less Efficient	N/A
7	Bayan Mohamm-ed Saleh	D-Talk: Sign Language Recognition System for People With Disability Using Machine Learning and Image Processing	Sign Language Image Dataset	SLR	Utilizes machine learning and image processing for communi-cation enhancement	Less Accuracy	60%

8	Narayana Dharapa-neni	American Sign Language Using instance based segmentation	OpenCV	CNN	The study selection highlights the Benefits Of using AI in healthcare	Less Efficient	N/A
9	AdithyaV	A Deep Convolutio nal Neural Network Approach for Static Hand Gesture Recognition	NUS hand posture dataset	CNN	More Optimize	Overfitting	94%
10	Ankita Saxena, Deepak Kumar Jain, Ananya Singhal	Sign Language Recognition Using Principal Component Analysis	Database of 10 sign gestures from Indian sign language	PCA	PCA used for extracting Useful data, Dimension reduction	Performance may decrease due to varying lighting conditions and background noise.	Recognitio n rate of gestures: 70-80%.

ALGORITHM

An ensemble learning technique for both classification and regression tasks is called Random Forest Algorithm.

Working of the Random Forest Algorithm:

1. **Initialization:** Initialization: A dataset comprising pictures or other kinds of data is used to start the Random Forest Algorithm. In the dataset, every data point denotes an instance with associated labels and features.
2. **Random Sampling:** Using a technique known as bootstrapping, the program creates many groups of data in this stage. It chooses subsets of the dataset at random using replacement, thus some data points can appear more than once in the subsets and others might not be at all.
3. **Tree Construction:** The algorithm builds a decision tree for each group of data. At each node, a random collection of features is used to construct these decision trees. Typically, information gain or Gini impurity measures are used to determine the optimum feature and threshold to minimize impurity, and then the nodes are split accordingly.
4. **Ensemble Learning:** By assembling a group of decision trees, the Random Forest Algorithm uses ensemble learning. A predetermined number of decision trees, each trained on a distinct portion of the data, may make up this ensemble.
5. **Voting:** Every decision tree in the ensemble independently predicts values of the unknown data points during the prediction phase. A voting mechanism is then used to combine the predictions made by each tree. The most frequent class, or mode, of the predictions is used as the final prediction in classification tasks. The average forecast of each tree is calculated for regression tasks.
6. **Output:** The Random Forest classifier yields the final prediction produced by the ensemble of decision trees. The best estimate of the label or result for a particular input instance provided by the algorithm is represented by this prediction.
7. **Evaluation:** Lastly, a variety of evaluation metrics, including accuracy, precision, recall, and F1-score, are used to assess the Random Forest classifier's performance. These metrics shed light on how well the classifier generalizes to new data and makes accurate predictions.

1.1 Random Forest Classifier is commonly used:

Ensemble Learning: The Random Forest Classifier makes use of ensemble learning, which combines several decision trees to produce accurate and dependable predictions. The model's total predictive power is increased by combining the predictions from several trees.

High Accuracy: Random Forest's capacity to deliver great precision is one of its main features. For both classification and regression problems, Random Forest may produce reliable predictions by capturing complex interactions between features without overfitting.

Robust to Outliers: Random Forest is renowned for its stability against outliers. Forecasts from several trees are combined to reduce the effect of outliers on the final predictions, producing more consistent and trustworthy outcomes.

Handles Missing Values: Without the requirement for imputation or removal, Random Forest can manage missing values in the dataset effectively. Surrogate splits and averaging over several trees are used to accomplish this.

Feature Importance: By allowing users to rank attributes for additional analysis and prediction, Random Forest facilitates feature selection and comprehension by offering ratings based on relevance.

Reduced Risk of Overfitting: When compared to individual decision trees, Random Forest's ensemble approach lowers the likelihood of overfitting. Random Forest makes overfitting less likely by averaging predictions from several trees, which enhances generalization to fresh data.

Parallelization: By effectively parallelizing its tree training process, Random Forest can handle big datasets more efficiently and expeditiously, resulting in faster model training times.

No Need for Feature Scaling: Random Forest removes feature scaling by using feature threshold-based node splitting, which makes preparing datasets with different feature sizes easier.

Tuning Parameters: Compared to other algorithms, Random Forest frequently performs well with default settings and is less sensitive to the choice of hyperparameters. Because of this, Random Forest is a viable option for a variety of machine learning tasks and eliminates the need for substantial parameter adjustment.

DESIGN

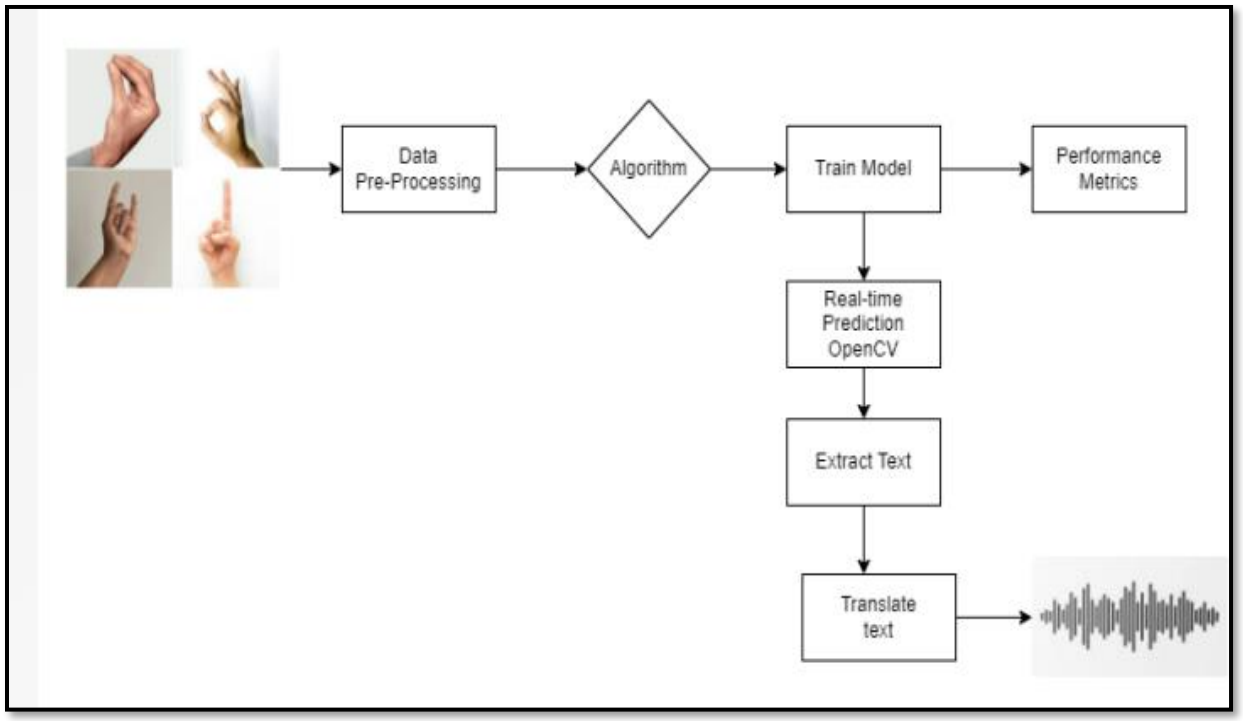

Figure 2: Designing of the sign language to speech

PROPOSED METHADOLOGY

To develop a sign language to speech system, The first step is gathering a dataset of pictures of sign language motions. From these images, we extract hand coordinates using the MediaPipe library. The data is then recorded in a dictionary format along with class labels after the retrieved coordinates are preprocessed and tagged with the appropriate sign language motions. The next stage is to use the preprocessed data to train a machine learning model, like a Random Forest Classifier. To make sure the model is effective at identifying sign language motions, we assess its accuracy on a validation set. When the trained model reaches a high enough accuracy level, it is saved in a pickle file for later use. Next, we build the system in real-time by utilizing OpenCV to record a live video stream and training the model to predict sign language movements in real-time. Characters are taken out of the gestures and combined to form words or sentences. We translate the identified text from English to Telugu or another target language using the Googletrans library. In order to transform the translated text into speech and provide the meaning of the identified sign language movements in the required language, such as Telugu, we finally use the GTTS (Google Text-to-Speech) library. This all-inclusive method bridges the gap between spoken and sign language for those with hearing difficulties, enabling them to communicate effectively.

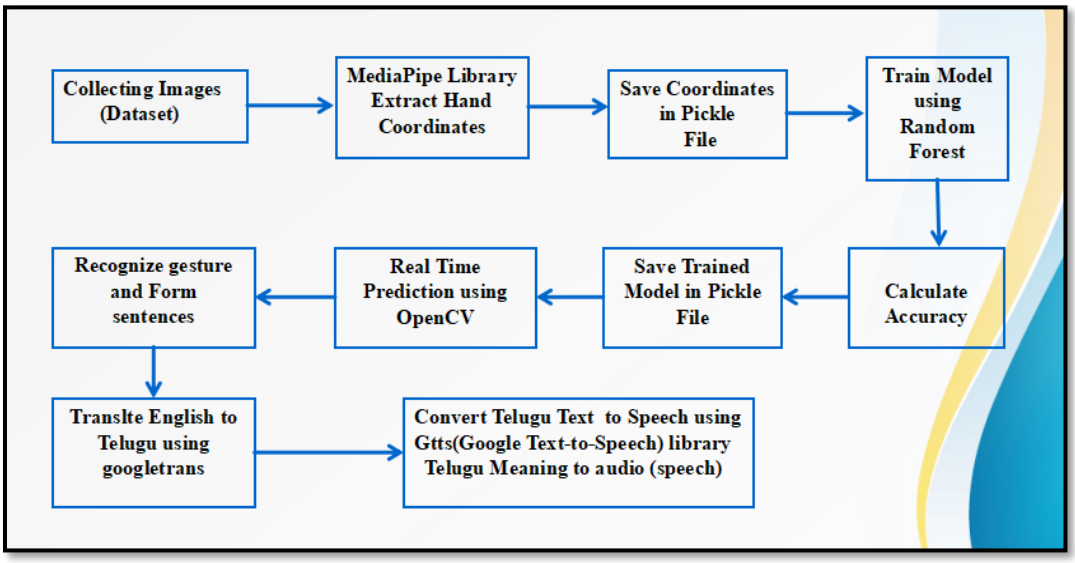

Figure 3: Flow of sign language to speech

PERFORMANCE METRICS

Accuracy: When assessing a model's performance, accuracy is a crucial indicator that shows the percentage of accurate predictions compared to all of the predictions made. It functions as an all-encompassing indicator of how well the model performs overall in accurately identifying occurrences across all dataset classifications.

Formula:

Accuracy = Number of correct predictions / Total number of predictions.

Error rate: The ratio of the total number of wrong forecasts to the total number of predictions made by the model is the error rate, a statistic that expresses the percentage of incorrect predictions made by a model.

Formula:

$$\text{Error Rate} = 1 - \text{Accuracy}$$

Precision: Precision is a metric that expresses how well the model predicts positive outcomes; it is the ratio of accurately identified positive cases to the total number of positive predictions. It evaluates the model's capacity to reduce false positive mistakes and guarantee that the positive occurrences that are predicted are, in fact, relevant.

Formula:

$$\text{Precision} = \text{True Positives} / (\text{False Positives} + \text{True Positives})$$

Recall: By calculating the percentage of accurately detected positive occurrences among all real positive instances in the dataset, recall, sometimes referred to as sensitivity, assesses the model's capacity to capture all positive cases. It measures how well the model reduces false negative mistakes and guarantees thorough coverage of positive cases.

Formula:

$$\text{Recall} = \text{True Positives} / (\text{False Negatives} + \text{True Positives})$$

F1-score: The model's performance is balanced by combining precision and recall into a single metric called the F1-score. It makes sure that both metrics are taken into account equally. It is the harmonic mean of recall and precision.

Formula:

$$\text{F1-score} = 2 \times (\text{Precision} + \text{Recall} / \text{Precision} \times \text{Recall})$$

EXPERIMENTAL RESULTS AND DISCUSSION

	precision	recall	f1-score	support
0	0.95	1.00	0.97	37
1	1.00	0.97	0.99	40
10	1.00	0.98	0.99	53
11	1.00	1.00	1.00	51
12	1.00	1.00	1.00	50
13	1.00	1.00	1.00	48
14	1.00	1.00	1.00	53
15	1.00	1.00	1.00	51
16	1.00	0.98	0.99	42
17	1.00	1.00	1.00	50
18	1.00	1.00	1.00	50
19	0.98	1.00	0.99	58
2	1.00	1.00	1.00	55
20	1.00	0.98	0.99	53
21	1.00	1.00	1.00	43
22	1.00	0.97	0.99	40
23	0.95	0.98	0.96	41
24	1.00	1.00	1.00	51
25	1.00	0.98	0.99	44
26	1.00	1.00	1.00	44
27	1.00	1.00	1.00	51
28	1.00	1.00	1.00	54
29	0.98	1.00	0.99	55
3	1.00	1.00	1.00	55
30	1.00	1.00	1.00	53
31	1.00	1.00	1.00	52
32	1.00	1.00	1.00	48
33	1.00	1.00	1.00	54
34	1.00	1.00	1.00	52
35	1.00	0.96	0.98	27
36	1.00	1.00	1.00	14
37	1.00	1.00	1.00	12
38	1.00	1.00	1.00	33
39	1.00	1.00	1.00	31
4	1.00	1.00	1.00	48
40	1.00	1.00	1.00	30
41	1.00	1.00	1.00	34
42	1.00	1.00	1.00	30
43	1.00	1.00	1.00	32
44	1.00	1.00	1.00	32
45	1.00	1.00	1.00	35
46	1.00	1.00	1.00	32
47	1.00	1.00	1.00	9
48	1.00	1.00	1.00	2
49	1.00	1.00	1.00	31
5	0.98	1.00	0.99	55
50	1.00	1.00	1.00	35
51	1.00	1.00	1.00	34
6	1.00	1.00	1.00	39
7	1.00	1.00	1.00	54
8	0.98	1.00	0.99	49
9	1.00	1.00	1.00	52
accuracy			1.00	2178
macro avg	1.00	1.00	1.00	2178
weighted avg	1.00	1.00	1.00	2178

Figure 4: Performance metrics results for model

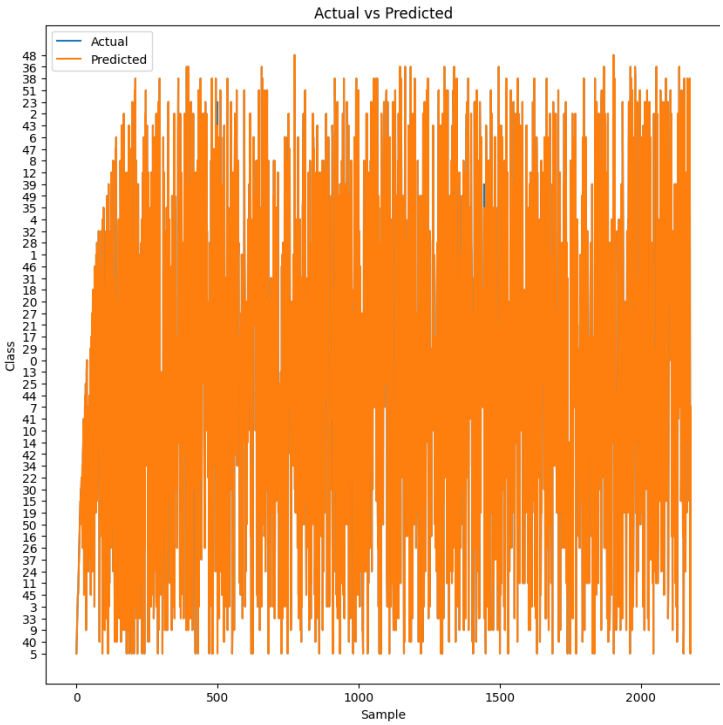

Figure 5: Actual & Predicted results

CONCLUSION

Our sign language to speech project successfully developed a system that translates sign language gestures into spoken Telugu language, facilitating seamless communication between who cannot speak or dumb people and those who speak. By leveraging the Random Forest algorithm for gesture recognition and speech synthesis, we have created a user-friendly solution that promotes inclusivity and accessibility for people who can not speak or dumb. This model has the accuracy is 99%. This project not only addresses the communication barriers but also fosters understanding and social interaction, ultimately enhancing the quality of life for this demographic.

BIBLIOGRAPHY

Machine Learning Libraries:

cv2 (OpenCV)
mediapipe
os
random
pickle
scikit-learn
numpy (np)
time
gtts
pygame
googletrans

Paper References:

Google Scholar: https://scholar.google.com/
Sci-hub: https://sci-hub.hkvisa.net/
Dataset: ASL (own dataset)

8.CONCLUSION

Our sign language to speech project successfully developed a system that translates sign language gestures into spoken Telugu language, facilitating seamless communication between who cannot speak or dumb people and those who speak. By leveraging the Random Forest algorithm for gesture recognition and speech synthesis, we have created a user-friendly solution that promotes inclusivity and accessibility for people who can not speak or dumb. This model has the accuracy is 99%. This project not only addresses the communication barriers but also fosters understanding and social interaction, ultimately enhancing the quality of life for this demographic.

BIBLIOGRAPHY

Machine Learning Libraries:

cv2 (OpenCV)
mediapipe
os
random
pickle
scikit-learn
numpy (np)
time
gtts
pygame
googletrans

Paper References:

Google Scholar: https://scholar.google.com/
Sci-hub: https://sci-hub.hkvisa.net/
Dataset: ASL (own dataset)