A Project Solution Report

On

Indian Sign Language Recognition

Submitted by:

Team Infinity (AICB1 – AI06)

01	**Hari Lakshmi A. S**
02	**Anushka Bajirao Patil**
03	**Mekapothu Manikanta Reddy**
04	**Pavaneswar Akula**
05	**Vrushali Arun Ghuge**
06	**Sikandar Baksh**

Under the guidance of

Ms. Diana Baby

Project guide cum Mentor

ABSTRACT:

Around 466 million people worldwide have hearing loss, and 34 million are children. `Deaf' people have very little or no hearing ability. They use sign language for communication. People use different sign languages in different parts of the world. Compared to spoken languages they are very less in number. India has developed a sign language by the name Indian Sign Language (ISL). In developing countries, there are only countable schools for deaf students. The Unemployment rate among adults with hearing loss is very high in developing countries. Data from Ethnologue states that among the deaf population in India, about 1% of the total population, the literacy rate and the number of children attending schools are very low. Our project aims to reduce the basic step in bridging the communication gap between normal people and deaf people using Indian sign language. Effective extension of this project to words and common expressions may not only make deaf people communicate faster and easier with the outer world, but also provide a boost in Developing autonomous systems for understanding and aiding them.

KEYWORDS:

Indian Sign Language(ISL), MediaPipe, Long Short Term Memory(LSTM), Gesture Recognition

PROBLEM STATEMENT:

Sign language uses lots of gestures so that it looks like movement language which consists of a series of hand and arm motions. sign language includes specific gestures to each alphabet in the English dictionary and for each number between 0 and 9. Based on these sign languages are made up of two groups, namely static gesture, and dynamic gesture. The static gesture is used for alphabet and number representation, whereas the dynamic gesture is used for specific concepts. Dynamic also includes words, sentences, etc. The static gesture consists of hand gestures, whereas the latter includes the motion of hands, head, or both. Sign language is a visual language and consists of 3 major components, such as finger-spelling, word-level sign vocabulary, and non-manual features. Finger-spelling is used to spell words letter by letter and convey the message whereas the latter is keyword-based. But the design of a sign language translator is quite challenging despite many research efforts during the last few decades. Also, even the same signs have significantly different appearances for different signers and different viewpoints.

SOLUTION:

Gesture Recognition (GR) is one of the most important sub-topics in action recognition. In recent years, it has gained much attention for its role in human-machine interaction. Gesture recognition is also used in sign language recognition which is very important for special people. A lot of research had been done in GR.

We will try to make an AI learning model which will store some data (given at the time of training) and then the module will be trained by giving so many pictorial data, here multiple data will be stored for one gesture (15-25) so that the module can detect the exact sign shown to it from any angle.

The training will be done by acting gestures manually and storing the dataset with a particular name title.

PROPOSED SOLUTION:

Dataset : Creating our own Dataset (ISL Dataset)

Algorithm: Long short Term Memory

Framework :MediaPipe,TensorFlow

1.DATASET CREATION:

- In our project we are going to create our own dataset. The dataset contains representation of Alphabetical ISL sign language, Numerical ISL sign Language and Words that are being in our daily life. Example are Floor, Home.
- Our dataset contains multiple data for one gesture (15-25),which will be adequate to do prediction. The dataset is created manually through web camera.
- Our dataset is created by referring the available ISL Dictionary.

2. DETECTION OF HANDS:

- In our project to detect the hand movements we are going to use MediaPipe, a Framework which is helpful in human face,pose and hand detections.
- For detecting the hand we will wrapping the image from BGR-RGB because MediaPipe supports RGB.
- The extraction and collection of key points will be done.
- After Extraction ,The dataset will be collected manually through web camera and stored in a folder in our local system .The labels will be created for data.

3.TRAINING THE MODEL AND MAKING PREDICTIONS:

- To train the data we are going to use LSTM model to train the dataset.
- Long Short Term Memory is a deep learning model which has feedback connections
- After Training ,Model weights will saved and predictions will be done.

4.WEBSITE:

- A user friendly website will be developed using Django.

DETAILED STUDY:

In our project, we are going to use Mediapipe Framework and Long Short Term Memory (LSTM) deep learning Model.

1.MediaPipe:

MediaPipe is a Framework for building machine learning pipelines for processing time-series data like video, audio, etc. This cross-platform Framework works in Desktop/Server, Android, iOS, and embedded devices like Raspberry Pi and Jetson Nano. Mediapipe is a cross-platform library developed by Google that provides amazing ready-to-use ML solutions for computer vision tasks.

MediaPipe Hand is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame. Whereas current state-of-the-art approaches rely primarily on powerful desktop environments for inference, our method achieves real-time performance on a mobile phone, and even scales to multiple hands

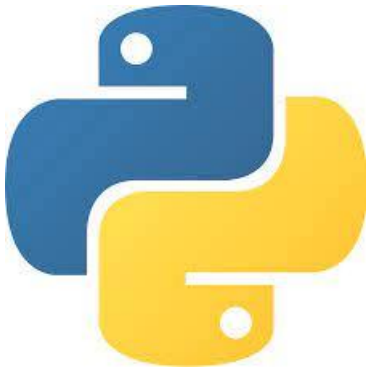
2.Long Short Term Memory:

- Long short-term memory (LSTM), is an artificial neural network used in the fields of artificial intelligence and deep learning. LSTM has feedback connections. LSTM applies to tasks such as unsegmented, connected handwriting recognition, speech recognition, machine translation, robot control, video games, and healthcare. LSTM has become the most cited neural network of the 20th century.
- A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.
- LSTM networks are well-suited to classifying, processing, and making predictions based on time series data,

APPLICATIONS:

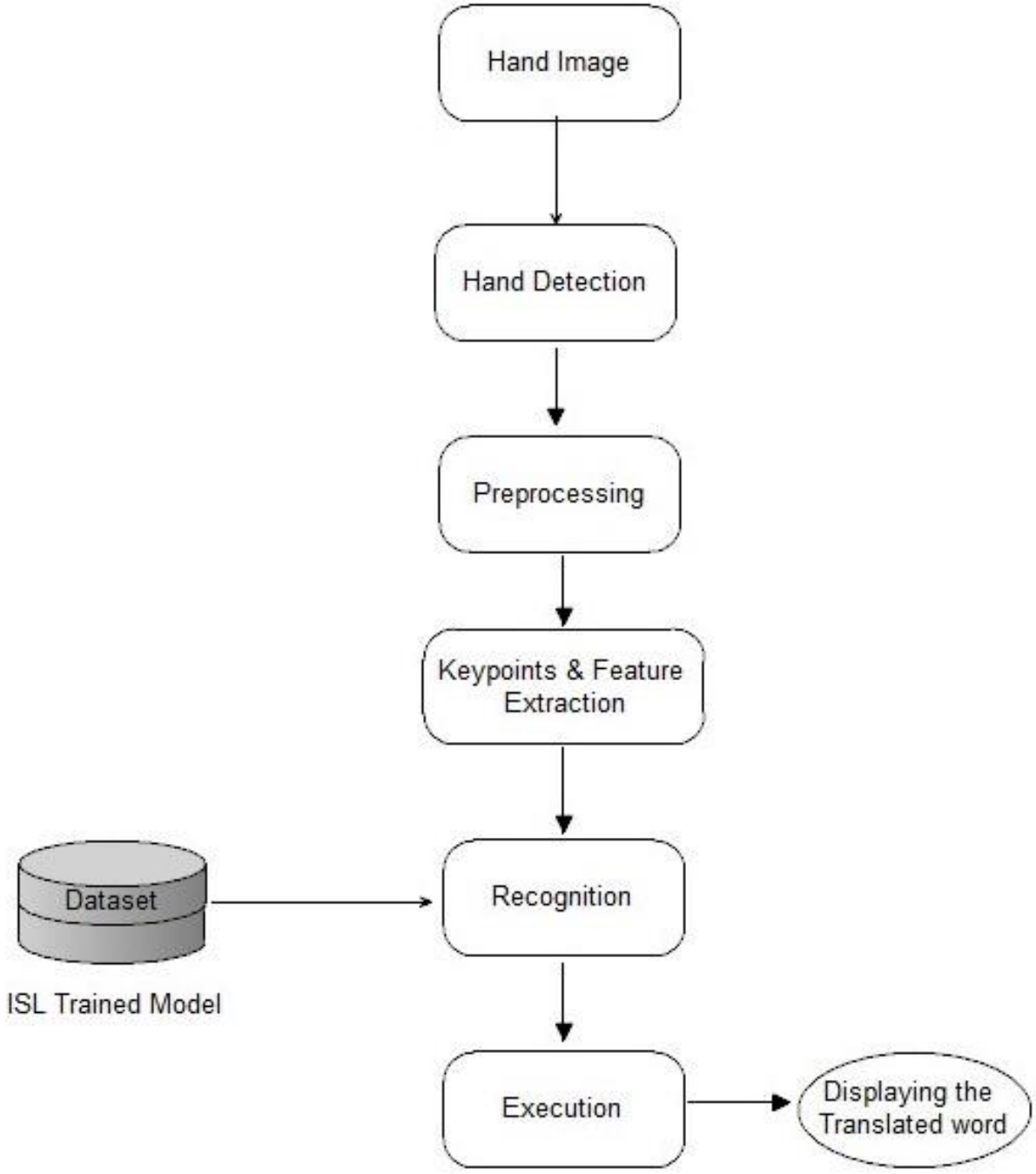
- Robot Control
- Speech recognition
- Sign Language Translation
- Object Co-segmentation
- Time series anomaly detection

TOOLS:



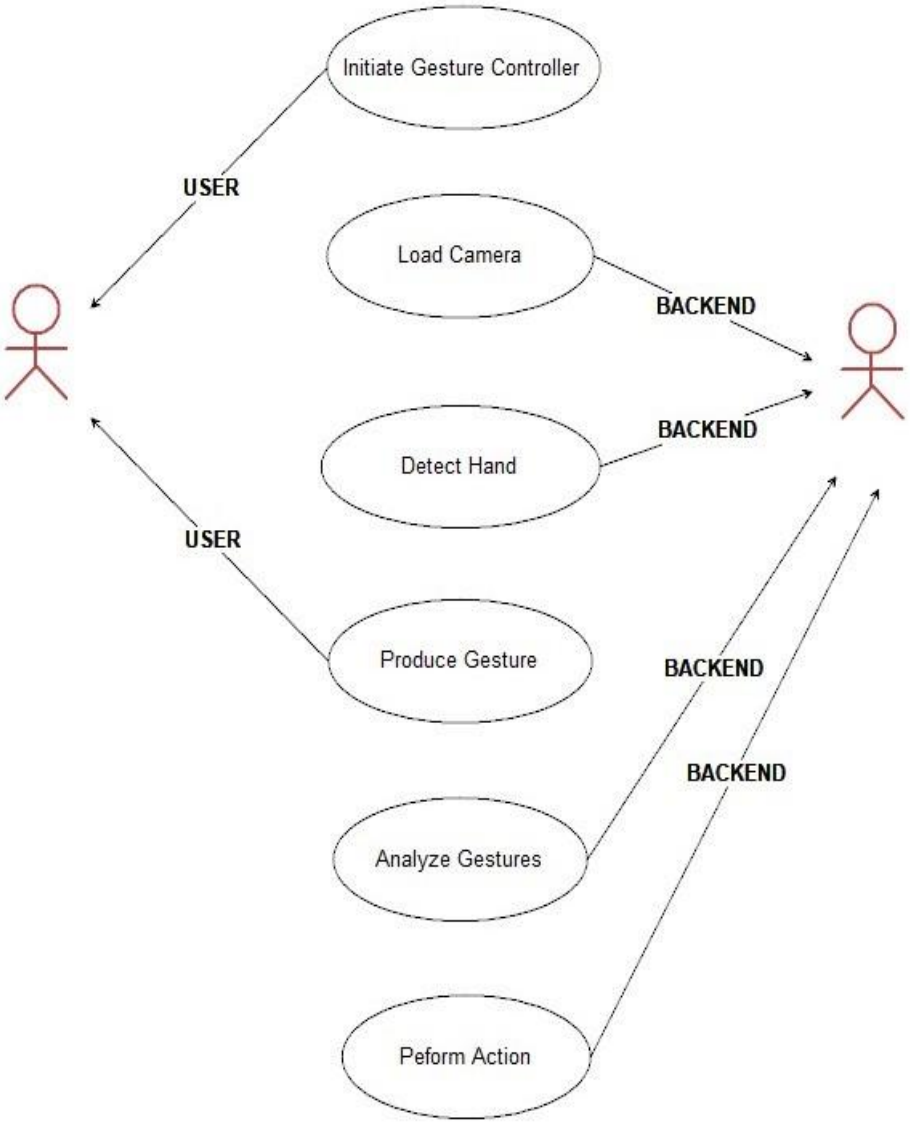
SYSTEM ARCHITECTURE:

A system architecture is a conceptual model that defines the structure, behavior, and views of a system.[1] An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.



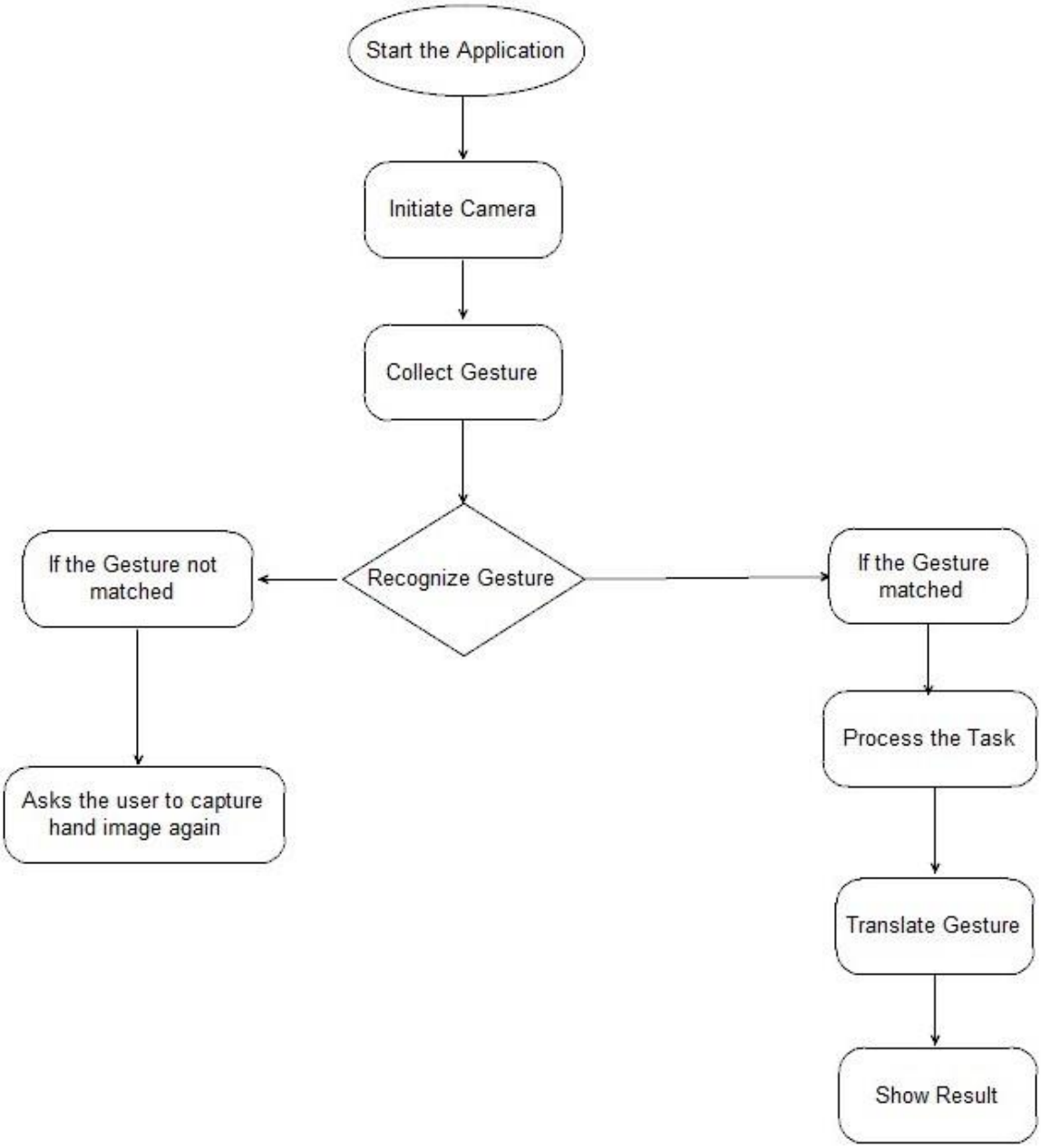
UML DIAGRAM:

A UML diagram is a diagram based on the UML (Unified Modeling Language) to visually represent a system along with its main actors, roles, actions, artifacts, or classes, to better understand, alter, maintain, or document information about the system.



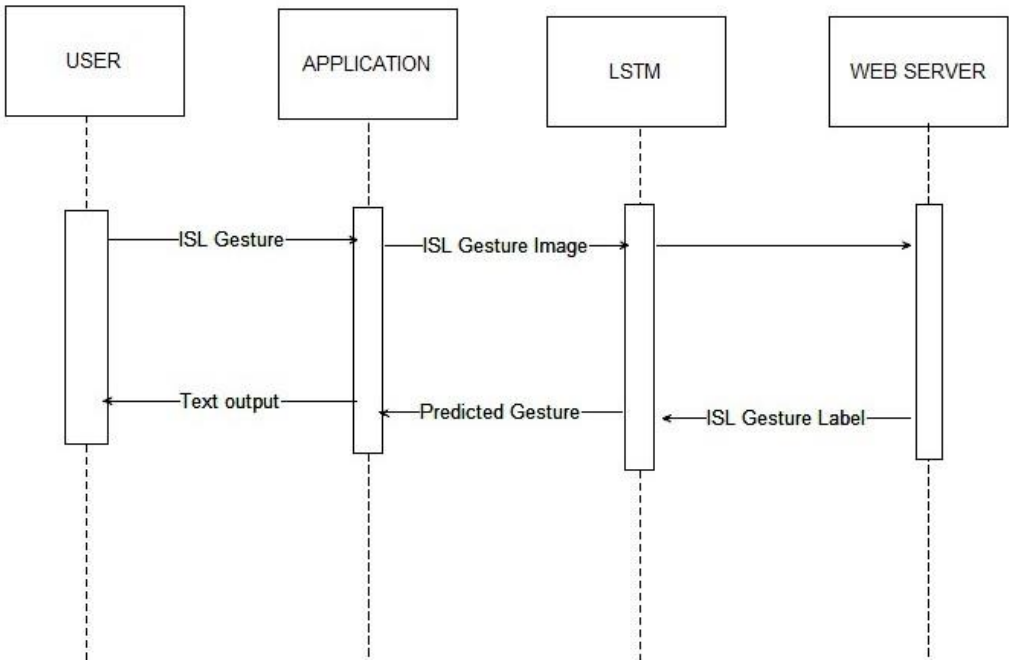
ACTIVITY DIAGRAM:

An activity diagram visually presents a series of actions or flow of control in a system similar to a flowchart or a data flow diagram. An activity diagram shows business and software processes as a progression of actions. These actions can be carried out by people, software components, or computers



SEQUENCE DIAGRAM:

A sequence diagram is a Unified Modeling Language (UML) diagram that illustrates the sequence of messages between objects in an interaction. A sequence diagram consists of a group of objects that are represented by lifelines and the messages that they exchange over time during the interaction.



SYSTEM REQUIREMENTS:

1. SOFTWARE REQUIREMENTS:

- Operating System: Linux/Window/Mac
- Backend: Python
- IDLE :Anaconda/Google Colaboratory
- Framework: TensorFlow, MediaPipe

2. HARDWARE REQUIREMENTS:

- PROCESSOR:i3/ AMD Ryzen
- RAM: 4 GB/6 GB/Higher
- Hard Disk: 6 GB/Higher
- Graphics Card(Optional): RTX / GTX

INDIVIDUALITY REPORT:

COMPARATIVE ANALYSIS:

METHOD 1:

APPROACH:

The approach used here is fourfold cross-validated results for the different approaches. They used Computer vision and Machine Learning algorithms for their research.

METHODOLOGY:

It has been classified into three stages to tackle their classification problem

Stage 1: The first stage is to segment the skin part from the image

Stage 2: Is to extract relevant features from the skin segmented images which can prove significant for the next stage.

Stage 3: Use the extracted features as input into various supervised learning models for training and then finally use the trained models for classification.

FEATURE EXTRACTION:

1. They are used to start SIFT (Scale Inverse Feature Transform) features as it computes the key points in the image.

2. The skin segmented images were obtained using the YUV-YIQ model

ALGORITHMS:

1. Support Vector Machine: The best accuracies observed for this algorithm will be 76% on HOG feature vectors, whereas the 'RBF' kernel failed miserably

2. Random Forest: They used Random Forest with HOG feature vectors on the compressed images and got an accuracy of 46.45 % 4-fold accuracy.

3. Hierarchical Classification: They First trained a linear kernel SVM model to classify alphabets as one-handed or two-handed. The model accuracy was 95% a. Then they trained linear kernel, Multiclass SVM models, to classify the one-handed alphabets (56% accuracy) and two-handed alphabets (60% accuracy) and combined the system. Even though the individual models performed better than the direct multiclass SVM on HOG features, overall, the performance was nearly the same and four-fold CV accuracy of 53.23% was observed.

METHOD 2:

APPROACH:

The Approach was modeled in a way such that Sign Language is captured from a smartphone camera and its frames are transmitted to a remote server for processing. They avoided the use of any external hardware and made it user-friendly. Techniques such as Face detection, Object stabilization, and Skin color Segmentation are used.

METHODOLOGY:

Using an Android Smartphone, gestures and signs performed by the person using ISL are captured and their frames are transmitted to the server for processing. To make the frames ready for recognition of gestures and hand poses, they need to be pre-processed. The pre-processing first involves face removal, stabilization, and skin color segmentation to remove background details and later morphology operations to reduce noise. Then the images are fed into the classifier. This is encoded for HMM and fed to it. The gesture whose HMM chain gives the highest score with the forward-backward algorithm is determined to be the recognized gesture for this pattern.

DATASET:

The authors used a dataset that consists of 0 to 9 Sign language representation, A to Z alphabets representation, and some basic words like good morning, good night, etc. In total the dataset consisted of 24624 images.

ALGORITHMS:

They applied 6 different grid sizes – 5x5 10x10, 10x15, 15x15, 15x20 and 20x20 to extract features from the same training data. These features were then fitted into a k-NN classifier. The features extracted from the testing data were then classified using the k-NN classifier trained previously. They inferred that the grid size of 10x10 gives the highest accuracy of 99.714%.

ANDROID APPLICATION:

The system is implemented as an Android application. The application uses the Smartphone's camera to capture the sign language used by the person. The frames were captured at a rate of 5 frames per second and they were connected to a remote server. After each pose or gesture is classified, the result is sent back to the application which is displayed in the top portion. They used Sockets to stimulate a client-server connection.

INDIVIDUALITY OF THE PROJECT:

1. There are more than 300 different sign languages. There is no universal sign language. Different sign languages are used in different countries or regions. For example, British Sign Language (BSL),American Sign Language(ASL),India Sign Language(ISL) etc. In our project, we are going to use the Indian sign languages (ISL) dataset. Our dataset consists of ISL Alphabets, and ISL Words. We are going to create a dataset by referring ISL Dictionary.

2. There have been many projects implemented on Sign Languages recognition. Some projects recognize Alphabets sign and translate them and make a sentence, Some projects recognize Signs (Thank you, Hi) and translate them, they were the example of Frame at a point.

   The approach we are going to use is to take more frames and predict the action and translate them from ISL to English Languages. This can be done using the MediaPipe framework.

PROJECT OUTCOMES:

3. Reduces the bridge between the Normal people and Deaf &Dumb people.
4. Helps Normal people to have healthier communication with Deaf &Dumb people.
5. Develops an interest in the Indian community to learn Indian Sign Language.
6. Helps the Deaf & Dumb community to excel in their Academics
7. Inspires the Indian citizens to create more ISL Learning websites, Application
8. Helps Deaf & Deaf community to achieve greater heights in their Dream, Passion, and Career

RECOMMENDATIONS:

Hearing and speaking loss is the most common sensory deficit in humans today. As per WHO estimates in India, there are approximately 63 million people, who are suffering from Significant Auditory Impairment; this places the estimated prevalence at 6.3% of the Indian population. As per the NSSO survey, currently, 291 persons per one lakh population are suffering from severe to profound hearing loss (NSSO, 2001). Of these, a large percentage are children between the ages of 0 to 14 years.

1. Global Statistics - As per the World Health Organization (WHO) survey, over 460 million individuals which constitute about 5% of the world population have hearing disabilities.

2. Indian Statistics - Of these 460 million individuals about 12.3 million belong to India.

Our recommendation would be to develop an improved Android Application that has features like an ISL Learning Portal, a News section about the Deaf &

Dumb community, an ISL Translation section, and Details of Nearby helpers, and Hospitals. This application can be used for 12.3 Million deaf & dumb people in India.

REFERENCES:

1. Official Website of Ministry of Health &Family Welfare Government of India
2. Articles from The Times of India
3. Persons with Disabilities (Divyangjan) in India - A Statistical Profile: 2021 under Ministry of Statistics and Programme Implementation.
4. Articles from Hindustan Times
5. Sign Language prediction using Machine learning algorithm published by International Research Journal of Engineering and Technology (IRJET)
6. Referred Journal paper: Real-time Indian Sign Language (ISL) Recognition published by Institute of Electrical and Electronics Engineers
7. https://google.github.io/mediapipe/solutions/hands.html
8. https://en.wikipedia.org/wiki/List_of_sign_languages
9. https://www.bbc.com/news/world-asia-india-39101899
10. https://en.wikipedia.org/wiki/List_of_sign_languages