**INTERNSHIP-II**

Project Report on

# Sign Language Conversion To Text and Audio

By


| | | |
|---|---|---|
| I. MAHIMA | - | 160121729011 |
| UB. POOJITHA | - | 160121729017 |
| E. AKSHAY | - | 160121729028 |
| J. NAVEEN | - | 160121729033 |
| K. MANOJ | - | 160121729040 |
| N. SANJAY | - | 160121729048 |



Branch: **Artificial Intelligence and Machine Learning (J)**

**Submitted to**

**Department of Artificial Intelligence and Machine Learning**

**CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)**

**(Affiliated to Osmania University), Gandipet, Hyderabad - 500075 2023-2024**

Under the guidance of

**Dr. Umakanta Choudhury**

**Professor, Department of EEE, Director- Innovation & Incubation**


# CERTIFICATE

This is to certify that the project entitled **"Sign Language Conversion To Text and Audio"** by the following student has carried out under my mentorship**.**

| | | |
|---|---|---|
| I. MAHIMA | - | 160121729011 |
| UB. POOJITHA | - | 160121729017 |
| E. AKSHAY | - | 160121729028 |
| J. NAVEEN | - | 160121729033 |
| K. MANOJ | - | 160121729040 |
| N. SANJAY | - | 160121729048 |

**Mr. M Asha Kiran**
**Department of Artificial Intelligence and Machine Learning**

**Dr. Umakanta Choudhury**
**Director-Innovation & Incubation**

**Smt.K.Mary Sudha Rani**
**Department of Artificial Intelligence and Machine Learning**

**Dr. J Manoranjini**
**Department of Artificial Intelligence and Machine Learning**

# INDEX

# INTRODUCTION TO AICTE IDEA LAB INTERNSHIP

## What is AICTE Idea Lab?

AICTE has launched a scheme for establishing AICTE-IDEA (Idea Development, Evaluation & Application) Lab in its approved institutions for encouraging students for application of Science, Technologies, Engineering and Mathematics (STEM) fundamentals towards enhanced hands-on experience and learning by doing. Thus, the Idea-Lab serves as a 'Technology Incubator' that aims to support potential ideas with technological merit. Mentoringis provided to develop a potential product, commercially, and connect to angel investors.

## Vision of Idea Lab

To create a culture of start-ups, entrepreneurship, makers and innovators in the higher educational Institute and the state. The aim will be to explore and experiment without the fear of failure with the mission to:

★ Idea Creation to business model.

★ Create engineers, designers, service providers and other stakeholders to provide innovative solutions for the problems of industry and society at large.

★ Minimize barriers for Prototype Development

★ Facilitate transition from prototyping to manufacturing through active collaboration and teamwork.

★ Provide talent development and upskilling opportunities for the MSME industries and students.

## Software Engineering Prototype Development

Software development refers to a set of computer science activities dedicated to the process of creating, designing, deploying and supporting software. Software itself is the set of instructions or programs that tell a computer what to do. It is independent of hardware and makescomputers programmable. Among the various categories of software, the most common types include the following:

- ➔ Application software: The most common type of software, application software, is a computer software package that performs a specific function for a user, or in some cases, for another application. An application can be self-contained, or it can be a group of programs that run the application for the user. Examples of modern applications include office suites, graphics software, databases and database management programs, web browsers, word processors, software development tools, image editors andcommunication platforms.
- ➔ System software: These software programs are designed to run a computer's application programs and hardware. System software coordinates the activities and functions of the hardware and software. In addition, it controls the operations of the computer hardware and provides an environment or platform for all the other types of software to work in. The OS is the best example of system software; it manages all the other computer programs. Other examples of system software include the firmware, computer language translators and system utilities.
- ➔ Driver software: Also known as device drivers, this software is often considered a type of system software. Device drivers control the devices and peripherals connected to a computer, enabling them to perform their specific tasks. Every device that is connected to a computer needs at least one device driver to function. Examples include software that comes with any nonstandard hardware, including special game controllers, as well as the software that enables standard hardware, such as USB storage devices, keyboards, headphones and printers.
- ➔ Middleware: The term middleware describes software that mediates between application and system software or between two different kinds of application software. For example,

middleware enables Microsoft Windows to talk to Excel and Word. It is also used to send a remote work request from an application in a computer that has one kind of OS, to an application in a computer with a different OS. It also enables newer applications to work with legacy ones.

➔ Programming software: Computer programmers use programming software to write code. Programming software and programming tools enable developers to develop, write, test and debug other software programs. Examples of programming software include assemblers, compilers, debuggers and interpreters.

The dimensions of software quality include the following characteristics:

- Accessibility: The degree to which a diverse group of people, including individuals who require adaptive technologies such as voice recognition and screen magnifiers, can comfortably use the software.
- Compatibility: The suitability of the software for use in a variety of environments, OSes, devices and browsers.
- Efficiency: The ability of the software to perform well without wasting energy, resources, effort, time/money.
- Functionality: Software's ability to carry out its specified functions.
- Installability: The ability of the software to be installed in a specified environment.
- Localization: The various languages, time zones and other such features a software can function in.
- Maintainability: How easily the software can be modified to add and improve features, fix bugs, etc.
- Performance: How fast the software performs under a specific load.
- Portability: The ability of the software to be easily transferred from one location to another.
- Reliability: To perform a required function under specific conditions/defined period without any errors.
- Scalability: The measure of the software's ability to increase or decrease performance in response to changes in its processing demands.

- Security: The software's ability to protect against unauthorized access, invasion of privacy, theft, data loss, malicious software, etc.
- Testability: How easy it is to test the software.
- Usability: How easy it is to use the software.

## What Is Commercialization?

- Commercialization is the process or activity through which a good or service is introduced into the market.
- It is also the stage in the product life cycle when a product or service is first made available for purchase.
- Commercialization typically follows research and development (R&D) and can include activities such as marketing, packaging, pricing, and distribution.
- Commercialization is often confused with marketing, but marketing is only one aspect of commercialization.
- Commercialization also encompasses other activities such as product development, packaging, and pricing.
- Commercialization is critical in bringing a new product or service to market and can make or break a business.

Stages of commercialization:

- The first phase of commercialisation brings us to the introduction of the new products in the market. Introducing a product in the market is a massive step for any organization.
- The second stage of commercialisation focuses on the business. It has to be made under the supervision of target customers and must reflect on the excellent research and development for reaching the ultimate business goals.
- The company must satisfy its loyal members when they genuinely succeed in the commercialisation.

## Entrepreneurship Ecosystem

- "Ecosystem" refers to the elements – individuals, organizations or institutions – outside the individual entrepreneur that are conducive to, or inhibitive of, the choice of a person to become an entrepreneur, or the probabilities of his or her success following the launchof the Initiative.
- Organizations and individuals representing these elements are referred to as entrepreneurship stakeholders. Stakeholders are any entity that has an interest, actually or potentially, in there being more entrepreneurship in the region.
- Entrepreneurship stakeholders may include government, schools, universities, private sector, family businesses, investors, banks, entrepreneurs, social leaders, research centers, military, labor representatives, students, lawyers, cooperatives, multinationals, private foundations, and others.

Major Domains:

- Policy covers government regulations and support.
- Finance domain includes the full spectrum of financial services available to entrepreneurs.
- Culture covers societal norms and success stories that help to inspire people to become entrepreneurs.
- Support domain includes non-governmental institutions, infrastructure and the professionals support such as investment bankers, technical experts and advisors.
- Markets cover entrepreneurial networks and customers.
- Human capital includes the education system and the skill level of the workforce.

## Characteristics of Entrepreneurship

Not all entrepreneurs are successful; there are definite characteristics that make entrepreneurship successful. A few of them are mentioned below:

- Ability to take a risk: Starting any new venture involves a considerable amount of failure risk. Therefore, an entrepreneur needs to be courageous and able to evaluate and take risks, which is an essential part of being an entrepreneur.

- Innovation: It should be highly innovative to generate new ideas, start a company and earn profits out of it. Change can be the launching of a new product that is new to the market or a process that does the same thing but in a more efficient and economical way.

- Visionary and Leadership quality: To be successful, the entrepreneur should have a clear vision of his new venture. However, to turn the idea into reality, a lot of resources and employees are required. Here, leadership quality is paramount because leaders impart and guide their employees towards the right path of success.

- Open-Minded: In a business, every circumstance can be an opportunity and used for the benefit of a company. For example, Paytm recognised the gravity of demonetization and acknowledged the need for online transactions would be more, so it utilized the situation and expanded massively during this time.

- Flexible: An entrepreneur should be flexible and open to change according to the situation. To be on the top, a businessperson should be equipped to embrace change in a product and service, as and when needed.

- Know your Product: A company owner should know the product offerings and also be aware of the latest trend in the market. It is essential to know if the available product or service meets the demands of the current market, or whether it is time to tweak it a little. Being able to be accountable and then alter as needed is a vital part of entrepreneurship.

## Problem Solving

Problem-solving is a cognitive process that involves discovering, analyzing, and solving problems. It is a fundamental skill applicable in various aspects of life, from personal challenges to complex professional situations.

Here's a general framework for effective problem-solving:

1. Define the Problem: Clearly articulate the problem you are facing. Understand the nature and scope of the problem by identifying its key components. This step is crucial for developing a focused approach to finding a solution.
2. Understand the Context: Consider the broader context in which the problem exists.Identify any relevant factors, constraints, or dependencies that might impact the problem or its solution.
3. Generate Possible Solutions: Brainstorm a range of potential solutions. Encourage creativity and consider both conventional and unconventional approaches. Aim for quantity at this stage without immediately evaluating or dismissing ideas.
4. Evaluate and Prioritize Solutions: Assess the feasibility, effectiveness, and potential consequences of each solution. Consider the resources required and any potential risks. Prioritize the solutions based on their merits.
5. Make a Decision: Choose the solution that seems most viable and appropriate based on your evaluation. Consider the potential outcomes and consequences of implementing the chosen solution.
6. Create an Action Plan: Develop a detailed plan outlining the steps required to implement the chosen solution. Define roles and responsibilities, allocate resources, and establish a timeline for implementation.
7. Implement the Solution: Put your action plan into motion. Execute the steps outlined in your plan and monitor progress. Be prepared to adapt and make adjustments if unforeseen challenges arise.
8. Evaluate the Results: After implementing the solution, assess the results. Did it solve the problem as intended? If not, analyze what went wrong and consider alternative approaches or adjustments.

9. Iterate if Necessary: If the initial solution doesn't fully address the problem or if new issues arise, be prepared to iterate. Return to the problem-solving process, reevaluate, andmake necessary adjustments.
10. Reflect and Learn: Reflect on the entire problem-solving process. Consider what worked well, what could be improved, and what lessons you can carry forward to future problem-solving situations.

Key Tips for Effective Problem-Solving:

- Stay Calm and Focused: Maintain a calm and focused mindset. Avoid becoming overwhelmed by the problem and approach it with a clear and rational mindset.

- Collaborate: If the problem is complex, involve others in the problem-solving process. Collaborative efforts often lead to more diverse perspectives and creative solutions.

- Break it Down: Break the problem into smaller, more manageable components. This can make the problem-solving process more approachable and help in identifying specific areas that need attention.

- Learn from Mistakes: View problems as opportunities to learn and grow. Mistakes and setbacks are natural parts of the problem-solving process and can provide valuableinsights for future challenges.

## Brainstorming

Brainstorming is a creative problem-solving technique that involves generating a large number of ideas in a group setting. The goal is to encourage free thinking, promote collaboration, and generate a variety of potential solutions to a problem.

Here's a typical process for conducting a brainstorming session:

1. Define the Problem or Goal: Clearly articulate the problem or goal that the brainstorming session aims to address. This provides a focus for the participants and sets the context for idea generation.
2. Select a Diverse Group: Bring together a diverse group of individuals with different backgrounds, perspectives, and expertise relevant to the problem. Diversity often leads to a wider range of ideas.
3. Create a Comfortable Environment: Ensure that the physical or virtual environment is conducive to open discussion. Encourage a relaxed atmosphere where participants feel comfortable sharing their thoughts without fear of criticism.
4. Set Ground Rules: Establish ground rules to guide the brainstorming process. Common rules include encouraging non-judgmental thinking, welcoming all ideas, and deferring evaluation until later stages.
5. Choose a Facilitator: Designate a facilitator to guide the session and keep it on track. The facilitator's role includes managing time, encouraging participation, and ensuring that the process follows the established rules.
6. Generate Ideas: Encourage participants to freely share their ideas without overthinking or censoring. Quantity is initially more important than quality. Ideas can be written on a whiteboard, sticky notes, or a digital platform visible to all participants.
7. Encourage Wild and Unconventional Ideas: Welcome unconventional and "wild" ideas. Sometimes, the most innovative solutions arise from thinking outside the box. Participants should feel free to contribute even seemingly impractical or unconventional ideas.
8. Build on Others' Ideas: Encourage participants to build on each other's ideas. This can lead to the development of more refined and creative solutions. Participants can add details, combine ideas, or propose modifications.
9. Avoid Evaluation During Ideation: Stress that the goal of the brainstorming session is idea generation, not evaluation. Postpone any critical analysis or judgment until later stages to avoid stifling creativity.
10. Capture and Document Ideas: Record all generated ideas systematically. This can be done using a whiteboard, flip chart, or digital tools. Documenting ideas ensures that nothing is lost and provides a reference for later stages.

11. Review and Organize Ideas: After the initial ideation phase, review and organize the ideas. This may involve grouping similar ideas, identifying patterns, or highlighting key themes.
12. Evaluate and Select: Once a significant number of ideas have been generated, evaluated, and organized, the group can move on to the evaluation and selection phase. Criteria for evaluation may include feasibility, impact, and alignment with project goals.

## Prototyping

Prototyping and testing are critical stages in the design and development process. Whether you are working on a physical product, a software application, or any other project, creating prototypes and conducting tests help validate ideas, identify potential issues, and refine the final solution.

Here's a breakdown of these stages:

1. Define Prototype Goals: Clearly outline what you aim to achieve with the prototype. Whether it's testing a specific functionality, assessing user interaction, or validating a design concept, having clear goals will guide the prototyping process.
2. Choose the Right Prototype Type: Depending on your goals and the nature of your project, you can create different types of prototypes. This could include low-fidelity sketches, wireframes, interactive mockups, or even physical models, depending on the context.
3. Build the Prototype: Develop the prototype based on the defined goals and chosen type. Use the simplest and most cost-effective methods to represent key features or aspects of the final product. The focus at this stage is on functionality rather than aesthetics.
4. Iterate and Refine: Expect that your initial prototype might not be perfect. Iterative refinement is a key part of the process. Gather feedback, identify areas for improvement, and make necessary adjustments to the prototype.
5. Conduct Testing: Testing involves gathering data and feedback to evaluate the prototype's performance. This can include usability testing, user feedback, technical testing, or any other relevant assessments. The goal is to identify both strengths and weaknesses.

6. Gather Feedback: Actively seek feedback from stakeholders, users, or any relevant parties. This feedback is invaluable in understanding how well the prototype aligns with the project goals and user needs.
7. Analyze Results: Evaluate the results of the testing phase. Identify what worked well, what needs improvement, and any unexpected issues that arose during testing.
8. Refine and Iterate Again: Use the insights gained from testing to refine the prototype further. This may involve making changes to the design, functionality, or other aspects based on the feedback received.
9. Scale Up if Necessary: If the prototype performs well in testing and meets the project requirements, you can move forward to scaling up the development process to create the final product.
10. Document Learnings: Document the lessons learned from the prototyping and testing phases. This documentation is valuable for future reference and can inform the development of similar projects.

# Sign Language To Text and Audio


## Abstract

Sign language, a natural communication form for millions, faces accessibility barriers due to interpreter scarcity. We propose a novel real-time method using neural networks for fingerspelled American Sign Language to text conversion. Our approach extracts hand features and classifies them with 95.7% accuracy for individual letters, enabling real-time communication with significantly improved efficiency and reach compared to traditional methods. This deep learning approach paves the way for more accessible and inclusive communication for the deaf and hard of hearing community.


## 1. Introduction:

American sign language is a predominant sign language. Since the only disability Deaf and Dumb (hereby referred to as D&M) people have is communication related and since they cannot use spoken languages, the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. D&M people make use of their hands to express different gestures to express their ideas with other people. Gestures are the non-verbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language. A sign language is a language which uses gestures instead of sound to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. Contrary to popular belief, sign language is not international. These vary from region to region.

Sign language is a visual language and consists of 3 major components [6]:

| Fingerspelling | Word level sign vocabulary | Non-manual features |
|---|---|---|
| Used to spell words letter by letter . | Used for the majority of communication. | Facial expressions and tongue, mouth and body position. |

**Figure -1**

Minimizing the verbal exchange gap among D&M and non-D&M people turns into a want to make certain effective conversation among all. Sign language translation is among one of the most growing lines of research and it enables the maximum natural manner of communication for those with hearing impairments. A hand gesture recognition system offers an opportunity for deaf people to talk with vocal humans without the need of an interpreter. The system is built for the automated conversion of ASL into textual content and speech.

In our project we primarily focus on producing a model which can recognize Fingerspelling based hand gestures and provide audio for the same. The gestures we aim to train are as given in the image below.

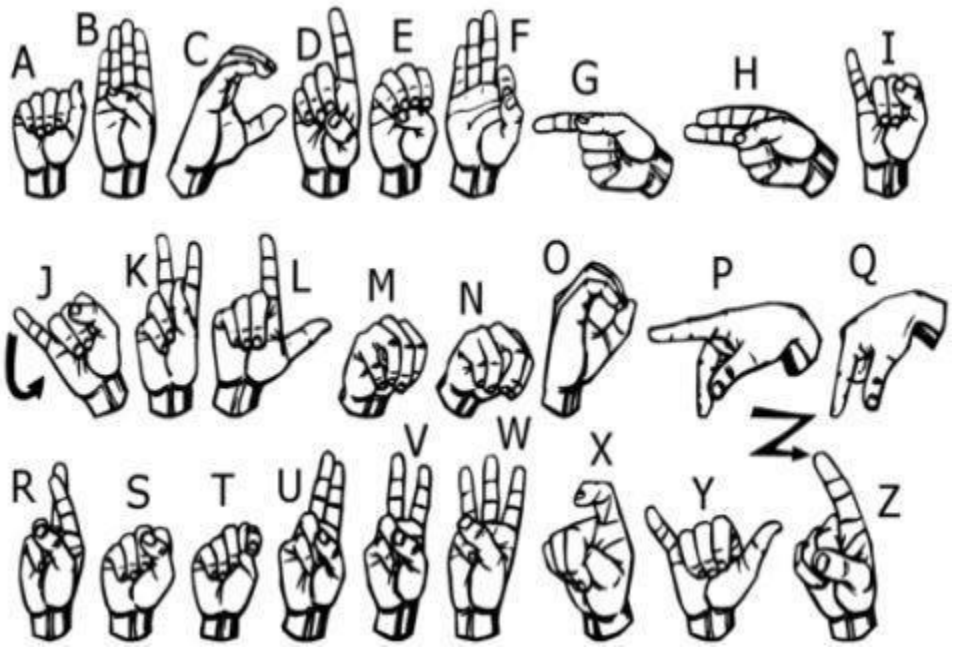


**Figure - 2**

## 2. Motivation:

For interaction between normal people and D&M people a language barrier is created as sign language structure since it is different from normal text. So, they depend on vision-based communication for interaction.

If there is a common interface that converts the sign language to text or audio, then the gestures can be easily understood by non-D&M people. So, research has been made for a vision-based interface system where D&M people can enjoy communication without really knowing each other's language.

The aim is to develop a user-friendly Human Computer Interface (HCI) where the computer understands the human sign language.

There are various sign languages all over the world, namely American Sign Language (ASL), French Sign Language, British Sign Language (BSL), Indian

Sign language, Japanese Sign Language and work has been done on other languages all around the world.

## 3. Literature Survey:

In the recent years there has been tremendous research done on the hand gesture recognition.

With the help of literature survey, we realized that the basic steps in hand gesture recognition are: -

- Data acquisition
- Data pre-processing
- Feature extraction
- Gesture classification

## 3.1 Data acquisition:

The different approaches to acquire data about the hand gesture can be done in the following ways:

1. **Use of sensory devices:**

   It uses electromechanical devices to provide exact hand configuration, and position. Different glove-based approaches can be used to extract information. But it is expensive and not user friendly.

2. **Vision based approach:**

   In vision-based methods, the computer webcam is the input device for observing the information of hands and/or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices, thereby reducing cost. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. The main challenge of vision-based hand detection ranges from coping with the large variability of the human hand's appearance due to a huge number of hand movements, to different skin-color possibilities as well as to the variations in viewpoints, scales, and speed of the camera capturing the scene.

## 3.2 Data Pre-Processing and 3.3 Feature extraction for vision-based approach:

- In [1] the approach for hand detection combines threshold-based color detection with background subtraction. We can use AdaBoost face detector to differentiate between faces and hands as they both involve similar skin-color.

- We can also extract necessary image which is to be trained by applying a filter called Gaussian Blur (also known as Gaussian smoothing). The filter can be easily applied using open computer vision (also known as OpenCV)and is described in [3].

- For extracting necessary image which is to be trained we can use instrumented gloves as mentioned in [4]. This helps reduce computation time for Pre-Processing and gives us more concise and accurate data compared to applying filters on data received from video extraction.

- We tried doing the hand segmentation of an image using color segmentation techniques but skin color and tone

is highly dependent on the lighting conditions due to which output, we got for the segmentation we tried to do were no so great. Moreover, we have a huge number of symbols to be trained for our project many of which look similar to each other like the gesture for symbols 'M' and 'N', hence we decided that in order to produce better accuracies for our large number of symbols, rather than segmenting the hand out of a random background we keep background of hand a stable single color so that we don't need to segment it on the basis of skin color. This would help us to get better results.

## 3.4 Gesture Classification:

- In [1] Hidden Markov Models (HMM) is used for the classification of the gestures. This model deals with dynamic aspects of gestures. Gestures are extracted from a sequence of video images by tracking the skin-color blobs corresponding to the hand into a body– face space centred on the face of the user.

- The goal is to recognize two classes of gestures: deictic and symbolic. The image is filtered using a fast look–up indexing table. After filtering, skin colour pixels are gathered into blobs. Blobs are statistical objects based on the location (x, y) and the colorimetry (Y, U, V) of the skin color pixels in order to determine homogeneous areas.

- In [2] Naïve Bayes Classifier is used which is an effective and fast method for static hand gesture recognition. It is based on classifying the different gestures according to geometric based invariants which are obtained from image data after segmentation.

- Thus, unlike many other recognition methods, this method is not dependent on skin colour. The gestures are extracted from each frame of the video, with a static background. The first step is to segment and label the objects of interest and to extract geometric invariants from them. Next step is the classification of gestures by using a K nearest neighbor algorithm aided with distance weighting algorithm (KNNDW) to provide suitable data for a locally weighted Naïve Bayes‟ classifier.

- According to the paper on "Human Hand Gesture Recognition Using a Convolution Neural Network" by Hsien-I Lin, Ming-Hsiang Hsu, and Wei-Kai Chen (graduates of Institute of Automation Technology National Taipei University of Technology Taipei, Taiwan), they have constructed a skin model to extract the hands out of an image and then apply binary threshold to the whole image. After obtaining the threshold image they calibrate it about the principal axis in order to centre the image about the axis. They input this image to a convolutional neural network

model in order to train and predict the outputs. They have trained their model over 7 hand gestures and using this model they produced an accuracy of around 95% for those 7 gestures.

# 4. Key words and Definitions:

## 4.1 Feature Extraction and Representation:

The representation of an image as a 3D matrix having dimension as of height and width of the image and the value of each pixel as depth (1 in case of Grayscale and 3 in case of RGB). Further, these pixel values are used for extracting useful features using CNN.

## 4.2 Artificial Neural Network (ANN):

Artificial Neural Network is a connection of neurons, replicating the structure of human brain. Each connection of neuron transfers information to another neuron. Inputs are fed into first layer of neurons which processes it and transfers to another layer of neurons called as hidden layers. After processing of information through multiple layers of hidden layers, information is passed to final output layer.

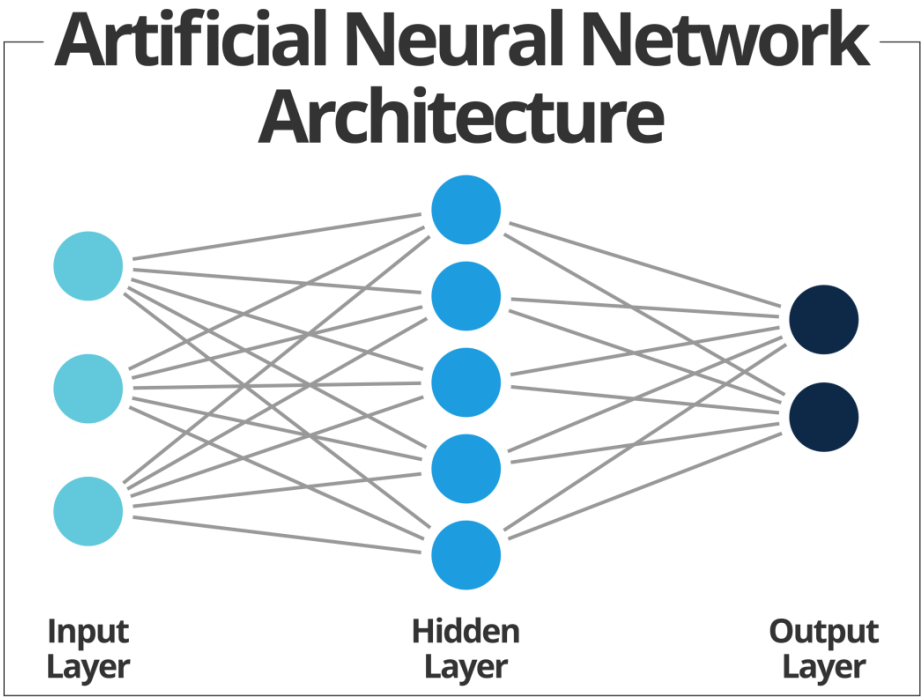



**Figure - 3**

These are capable of learning and have to be trained. There are different learning strategies

1. Unsupervised Learning
2. Supervised Learning
3. Reinforcement Learning

## 4.3 Convolutional Neural Network (CNN):

Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions (number of classes), because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

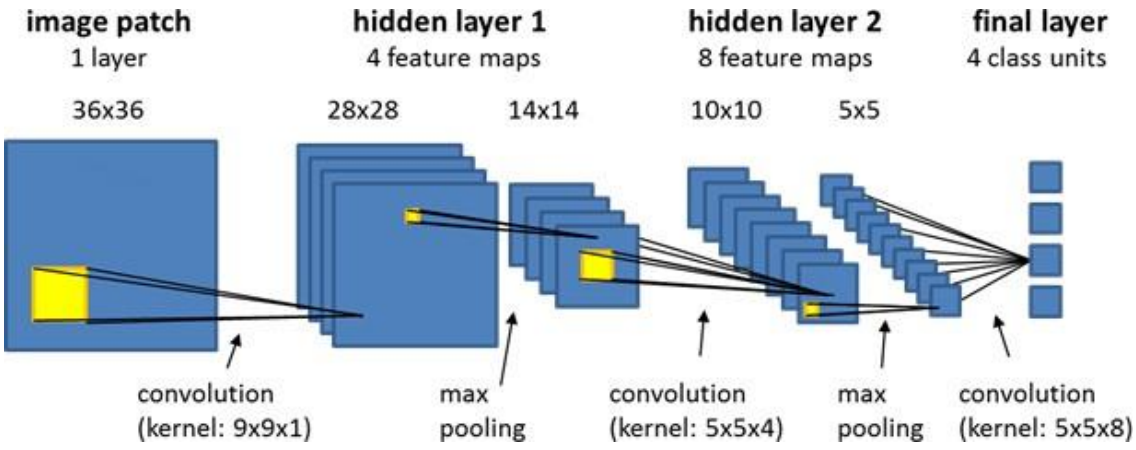



**Figure – 4**

### 1. Convolution Layer:

In convolution layer we take a small window size [typically of length 5*5] that extends to the depth of the input matrix. The layer consists of learnable filters of window size. During every iteration we slid the windowby stride size [typically 1], and compute the dot product of filter entries and input values at a given position.

As we continue this process we will create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position. That is, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some colour.

**2. Pooling Layer:**

We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters. There are two types of pooling:

**a. Max Pooling:** In max pooling we take a window size [for example window of size 2*2], and only take the maximum of 4 values. Well lid this window and continue this process, so well finally get an activation matrix half of its original Size.

**b. Average Pooling:** In average pooling, we take advantage of of all Values in a window.

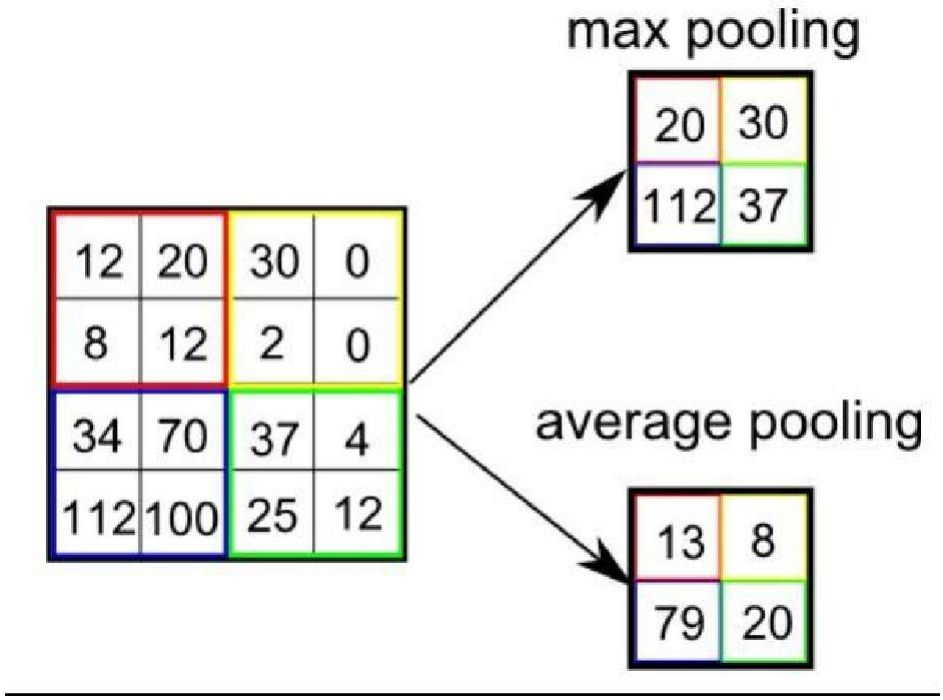



**Figure - 5**

**3. Fully Connected Layer:**

In convolution layer, neurons are connected only to a local region, while in a fully connected region, we will connect all the inputs to neurons.

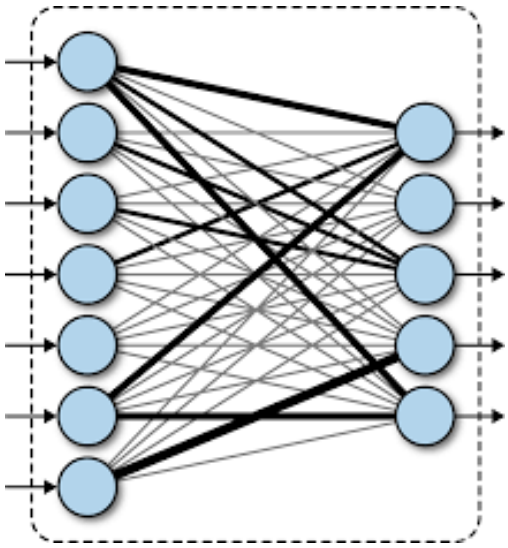


**Figure - 6**

4. **Final Output Layer:**

After getting values from fully connected layer, we will connect them to the final layer of neurons [having count equal to total number of classes], that will predict the probability of each image to be in different classes.

## 4.4 TensorFlow:

TensorFlow is an end-to-end open-source platform for Machine Learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in Machine Learning and developers easily build and deploy Machine Learning powered applications.

TensorFlow offers multiple levels of abstraction so you can choose the right one for your needs. Build and train models by using the high-level Keras API, which makes getting started with TensorFlow and machine learning easy.

If you need more flexibility, eager execution allows for immediate iteration and intuitive debugging. For large ML training tasks, use the Distribution Strategy API for distributed training on different hardware configurations without changing the model definition.

## 4.5 Keras:

Keras is a high-level neural networks library written in python that works as a wrapper to TensorFlow. It is used in cases where we want to quickly build and test the neural network with minimal lines of code. It contains

implementations of commonly used neural network elements like layers, objective, activation functions, optimizers, and tools to make working with images and text data easier.

### 4.6 OpenCV:

OpenCV (Open-Source Computer Vision) is an open-source library of programming functions used for real-time computer-vision.

It is mainly used for image processing, video capture and analysis for features like face and object recognition. It is written in C++ which is its primary interface, however bindings are available for Python, Java, MATLAB/OCTAVE.

# 5. Methodology:

The system is a vision-based approach. All signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

## 5.1 Data Set Generation:

For the project we tried to find already made datasets but we couldn't find dataset in the form of raw images that matched our requirements. All we could find were the datasets in the form of RGB values. Hence, we decided to create our own data set. Steps we followed to create our data set are as follows.

We used Open computer vision (OpenCV) library in order to produce our dataset.

Firstly, we captured around 500 images of each of the symbol in ASL (American Sign Language) for training purposes and around 200 images per symbol for testing purpose.

First, we capture each frame shown by the webcam of our machine. In each frame we define a Region Of Interest (ROI) which is denoted by a blue bounded square as shown in the image below:

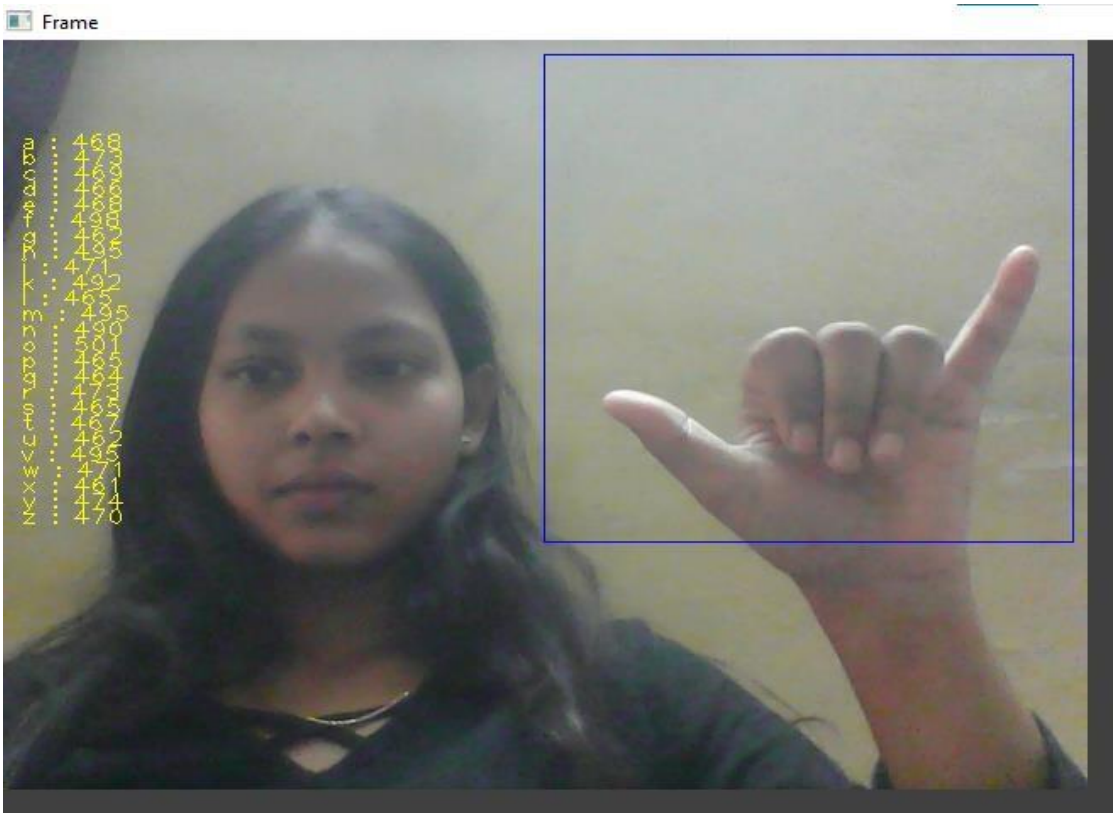



**Figure - 7**

Then, we apply Gaussian Blur Filter to our image which helps us extract various features of our image. The image, after applying Gaussian Blur, looks as follows:

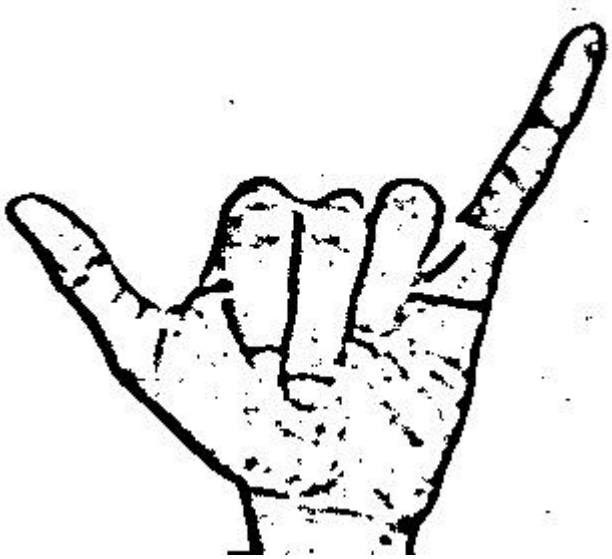


**Figure – 8**

## 5.2 Gesture Classification:

Our approach uses two layers of algorithm to predict the final symbol of the user.

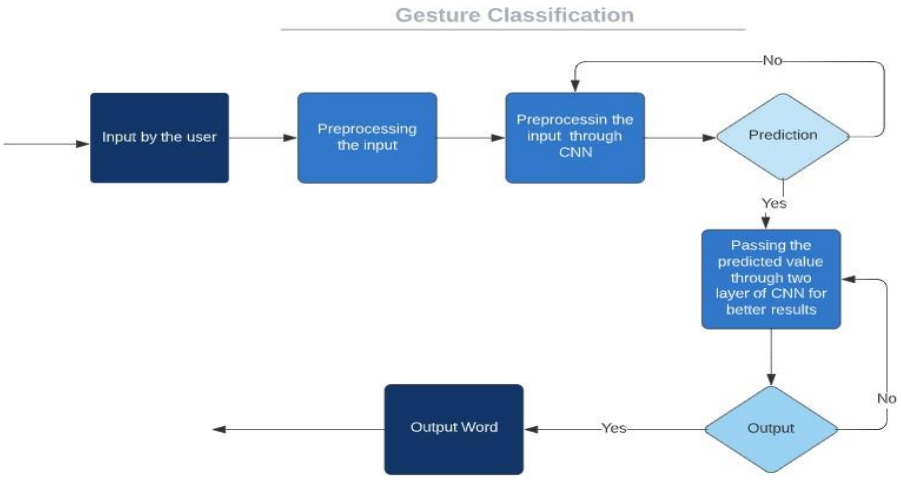



**Figure - 9**

**Algorithm Layer 1:**

1. Apply Gaussian Blur filter and threshold to the frame taken with openCV to get the processed image after feature extraction.
2. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for recognizing the character and converting it to audio.

**Algorithm Layer 2:**

1. We detect various sets of symbols which show similar results on getting detected.
2. We then classify between those sets using classifiers made for those sets only.

## Layer 1:

- ## CNN Model:

1. **1st Convolution Layer:** The input picture has resolution of 128x128 pixels. It is first processed in the first convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 126X126 pixel image, one for each Filter-weights.
2. **1st Pooling Layer:** The pictures are down sampled using max pooling of 2x2 i.e we keep the highest value in the 2x2 square of array. Therefore, our picture is down sampled to 63x63 pixels.
3. **2nd Convolution Layer:** Now, these 63 x 63 from the output of the first pooling layer is served as an input to the second convolutional layer. It is processed in the second convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 60 x 60 pixel image.
4. **2nd Pooling Layer:** The resulting images are down sampled again using max pool of 2x2 and is reduced to 30 x 30 resolution of images.
5. **1st Densely Connected Layer:** Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of 30x30x32 =28800 values. The input to this layer is an array of 28800 values. The output of these layer is fed to the 2nd Densely Connected Layer. We are using a dropout layer of value 0.5 to avoid overfitting.
6. **2nd Densely Connected Layer:** Now the output from the 1st Densely Connected Layer is used as an input to a fully connected layer with 96 neurons.

7. **Final layer:** The output of the 2nd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

- **Activation Function:**

We have used ReLU (Rectified Linear Unit) in each of the layers (convolutional as well as fully connected neurons).

ReLU calculates max(x,0) for each input pixel. This adds nonlinearity to the formula and helps to learn more complicated features. It helps in removing the vanishing gradient problemand speeding up the training by reducing the computation time.

- **Pooling Layer:**

We apply **Max** pooling to the input image with a pool size of (2, 2) with ReLU activation function. This reduces the amount of parameters thus lessening the computation cost and reduces overfitting.

- **Dropout Layers:**

The problem of overfitting, where after training, the weights of the network are so tuned to the training examples they are given that the network doesn't perform well when given new examples. This layer "drops out" a random set of activations in that layer by setting them to zero. The network should be able to provide the right classification or output for a specific example even if some of the activations are dropped out [5].

- **Optimizer:**

We have used Adam optimizer for updating the model in response to the output of the loss function.

Adam optimizer combines the advantages of two extensions of two stochastic gradient descent algorithms namely adaptive gradient algorithm (ADA GRAD) and root mean square propagation (RMSProp).

## Layer 2:

We are using two layers of algorithms to verify and predict symbols which are more similar to each other so that we can get us close as we can get to detect the symbol shown. In our testing we found that following symbols were not showing properly and were giving other symbols also:

1. **For D : R and U**
2. **For U : D and R**
3. **For I : T, D, K and I**
4. **For S : M and N**

So, to handle above cases we made three different classifiers for classifying these sets:

1. **{D, R, U}**
2. **{T, K, D, I}**
3. **{S, M, N}**

## 5.3 Converting to Audio:

We trained our model to turn character into spoken words .Imagine a robot reading aloud anything you type.

We implemented a function which takes the character which is recognized. Think of it like whispering what you want read into its ear.

We used a special tool called "pyttsx3" to prepare itself. It's like the robot warming up its voice.

We can tell the robot how fast you we it to read by choosing a number. Higher numbers mean faster talking, lower numbers mean slower. Then, the code, now ready, tells the robot to read the text you gave it. The robot speaks your words!

Finally, the code waits a bit (like taking a breath) before turning itself off. We can adjust this wait time if we want.

So, this code is like our personal narrator, ready to read anything you give it in any speed you choose. This way we make our model real talking model.

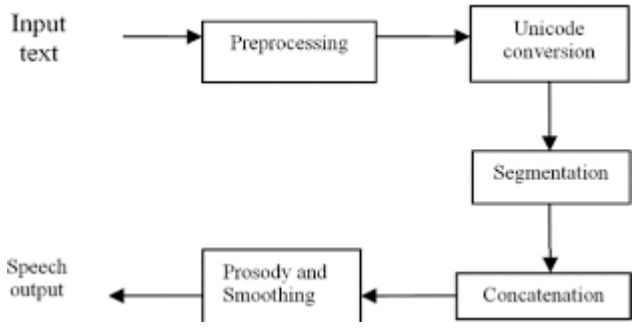



**Figure - 10**

## 5.4 Training and Testing:

We convert our input images (RGB) into grayscale and apply gaussian blur to remove unnecessary noise. We apply adaptive threshold to extract our hand from the background and resize our images to 128 x 128.

We feed the input images after pre-processing to our model for training and testing after applying all the operations mentioned above.

The prediction layer estimates how likely the image will fall under one of the classes. So, the output is normalized between 0 and 1 and such that the sum of each value in each class sums to 1. We have achieved this using SoftMax function.

At first the output of the prediction layer will be somewhat far from the actual value. To make it better we have trained the networks using labelled data. The cross-entropy is a performance measurement used in the classification. It is a continuous function which is positive at values which is not same as labelled value and is zero exactly when it is equal to the labelled value. Therefore, we optimized the cross-entropy by minimizing it as close to zero. To dothis in our network layer we adjust the weights of our neural networks. TensorFlow has an inbuilt function to calculate the cross entropy.

As we have found out the cross-entropy function, we have optimized it using Gradient Descent in fact with the best gradient descent optimizer is called Adam Optimizer.

# 6. Challenges Faced:

There were many challenges faced during the project. The very first issue we faced was that concerning the data set. We wanted to deal with raw images and that too square images as CNN in Keras since it is much more convenient working with only square images.

We couldn't find any existing data set as per our requirements and hence we decided to make our own data set. Second issue was to select a filter which we could apply on our images so that proper features of the images could be obtained and hence then we could provide that image as input for CNN model.

We tried various filters including binary threshold, canny edge detection, Gaussian blur etc. but finally settled with Gaussian Blur Filter.

More issues were faced relating to the accuracy of the model we had trained in the earlier phases. This problem was eventually improved by increasing the input image size and also by improving the data set.

## 7. Results:

We have achieved an accuracy of **95.8%** in our model using only layer 1 of our algorithm, and using the combination of **layer 1 and layer 2** we achieve an accuracy of **98.0%**, which is a better accuracy then most of the current research papers on American sign language.

Most of the research papers focus on using devices like Kinect for hand detection.

In [7] they build a recognition system for Flemish sign language using convolutional neural networks and Kinect and achieve an error rate of **2.5%.**

In [8] a recognition model is built using hidden Markov model classifier and a vocabulary of 30 words and they achieve an error rate of **10.90%**.

In [9] they achieve an average accuracy of **86%** for 41 static gestures in Japanese sign language.

They also used CNN for their recognition system. One thing should be noted that our model doesn't uses any background subtraction algorithm whiles some of the models present above do that.

So, once we try to implement background subtraction in our project the accuracies may vary. On the other hand, most of the above projects use Kinect devices but our main aim was to create a project which can be used with readily available resources. A sensor like Kinect not only isn't readily available but also is expensive for most of

audience to buy and our model uses a normal webcam of the laptop hence it is great plus point.Below are the

| | | | | | | P | r | e | d | i | c | t | e | d | | V | a | l | u | e | s | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
| | A | 147 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| | B | 0 | 139 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 |
| | C | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 0 | 145 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | E | 0 | 0 | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | F | 0 | 0 | 0 | 0 | 0 | 135 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 10 | 0 | 0 | 0 |
| C | G | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | H | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 143 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| r | I | 0 | 0 | 0 | 33 | 0 | 0 | 0 | 0 | 108 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 1 | 0 | 0 | 0 | 0 |
| r | J | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| e | K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| c | L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| t | M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 154 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| a | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| l | Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 147 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| u | R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| e | S | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 0 | 0 | 0 | 132 | 0 | 0 | 0 | 0 | 8 | 0 |
| s | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 | 0 | 0 | 0 | 0 | 0 |
| | U | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 35 | 0 | 0 | 115 | 0 | 0 | 0 | 0 |
| | V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 | 1 | 0 | 0 |
| | W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 149 | 0 | 0 |
| | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 148 | 0 |
| | Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 |
| | Z | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

***Algo 1***

confusion matrices for our results.

**Figure - 11**

| | | | | | | P | r | e | d | i | c | t | e | d | | V | a | l | u | e | s | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
| | A | 147 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| | B | 0 | 139 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 |
| | C | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | D | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | E | 0 | 0 | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | F | 0 | 0 | 0 | 0 | 0 | 135 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 10 | 0 | 0 | 0 |
| C | G | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | H | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 143 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| r | I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| r | J | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| e | K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| c | L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| t | M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 152 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 154 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| a | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| l | Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 147 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| u | R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| e | S | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 133 | 0 | 0 | 0 | 0 | 8 | 0 |
| s | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 | 0 | 0 | 0 | 0 | 0 |
| | U | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 |
| | V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 | 1 | 0 | 0 |
| | W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 149 | 0 | 0 |
| | X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 148 | 0 |
| | Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 |
| | Z | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

***Algo 1 + Algo 2***

**Figure – 12**

## 8. Future Scope:

We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms.

We are also thinking of improving the Pre Processing to predict gestures in low light conditions with a higher accuracy.

This project can be enhanced by being built as a web/mobile application or a website for the users to conveniently access the project.

Also, the existing project only works for ASL; it can be extended to work for other native sign languages with the right amount of data set and training. This project implements a finger spelling translator; however, sign languages are also spoken in a contextual basis where each gesture could represent an object, or verb. So, identifying this kind of a contextual signing would require a higher degree of processing and natural language processing (NLP).

## 9. References:


[1] T. Yang, Y. Xu, and "A., Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, May 1994.

[2] Pujan Ziaie, Thomas M uller, Mary Ellen Foster, and Alois Knoll "A Na¨ive Bayes Munich, Dept. of Informatics VI, Robotics and Embedded Systems, Boltzmannstr. 3, DE-85748 Garching, Germany.

[3]https://docs.opencv.org/2.4/doc/tutorials/imgproc/gausian_median_blur_bilateral_filter/gausian_median_blur_bilateral_filter.html

[4] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.

[5]aeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural Networks-Part-2/

[6] http://www-i6.informatik.rwth-aachen.de/~dreuw/database.php

[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

[8] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011).

[9] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," *2017 Nicograph International (NicoInt)*, Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[10] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen" Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[11] https://opencv.org/

[12] https://en.wikipedia.org/wiki/TensorFlow

[13] https://en.wikipedia.org/wiki/Convolutional_neural_nework

# 10. Appendix:

## 1. openCV:

openCV (Open-Source Computer Vision Library) is released under a BSD license and hence it's free for both academic and commercial use.

It has C++, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android. OpenCV was designed for computational efficiency and with a strong focus on real-time applications.

Written in optimized C/C++, the library can take advantage of multi-core processing. Enabled with OpenCL, it can take advantage of the hardware acceleration of the underlying heterogeneous compute platform.

Adopted all around the world, OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 14 million. Usage ranges from interactive art, to mines inspection, stitching maps on the web or through advanced robotics.

## 2. Convolutional Neural Network:

CNNs use a variation of multilayer perceptron's designed to require minimal pre-processing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage.

They have applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing.

## 3. TensorFlow:

TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.

It is used for both research and production at Google.

TensorFlow was developed by the Google brain team for internal Google use. It was released under the Apache 2.0 open-source library on November 9, 2015.

TensorFlow is Google Brain's second-generation system. Version 1.0.0 was released on February 11, 2017. While the reference implementation runs on single devices, TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units).

TensorFlow is available on 64-bit Linux, macOS, Windows, and mobile computing platforms including Android and iOS.

Its flexible architecture allows for the easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices.

# Business Plan Template


# SignSenseTech

# Business Plan

**2023**


# Table of Contents

## Executive Summary

“SignSenseTech” introduces a revolutionary AI-powered system that bridges this gap. Our machine learning model recognizes hand gestures corresponding to ASL alphabets, translating them into text and spoken audio in real-time. This empowers seamless communication between ASL users and non-signers, fostering greater inclusivity and accessibility.

SignSenseTech stands out in the market due to its proprietary dataset, meticulously crafted for training the ASL recognition model. Our utilization of Convolutional Neural Network (CNN) technology ensures high accuracy and efficiency in recognizing hand gestures. The seamless integration of character and audio output further sets us apart, providing a comprehensive solution for communication.

SignSenseTech anticipates robust growth in the coming years, driven by increasing demand for assistive technologies and a commitment to ongoing innovation. Initial projections indicate a positive financial trajectory, with revenue streams coming from both product sales and potential partnerships for broader deployment. Our Convolutional Neural Network (CNN) model boasts high accuracy in recognizing American Sign Language (ASL) alphabets, even in real-time scenarios. The integrated text-to-audio feature eliminates the need for separate sign language interpreters, increasing accessibility and convenience. Pioneer advancements in assistive technology, collaborating with research institutions and industry leaders to create a more inclusive world for the D&M community and Expanding our product offerings to cover a broader range of sign languages and gestures, increasing our impact on global communication accessibility and finally Establishing SignSenseTech as a leader in ASL recognition technology, securing partnerships for pilot programs and early adoption are the goals of our company.

SignSenseTech is seeking strategic investors who share our vision and are passionate about making a positive impact on the lives of the D&M community. We are looking for partners who not only provide financial support but also bring expertise and networks that can accelerate our growth. SignSenseTech presents a unique opportunity to invest in a socially impactful and commercially viable technology with the potential to revolutionize communication for the Deaf and Mute community. We invite you to join us in bridging the gap and creating a more inclusive world. Investment funds will be allocated towards research and development, market expansion, and enhancing our technological infrastructure.

## Company & Business Description

### Company Purpose

SignSenseTech is a pioneering technology company focused on revolutionizing communication accessibility for the Deaf and Mute (D&M) community. Our flagship product utilizes a custom-built dataset and Convolutional Neural Network (CNN) technology to recognize and translate American Sign Language (ASL) gestures into both text and audio, facilitating seamless communication. We offer a comprehensive solution: our software recognizes hand gestures in real-time, converts them into text, and subsequently speaks the text aloud, bridging the gap between the Deaf and Mute community and the hearing world.We aim to break down barriers by providing innovative solutions that empower individuals with hearing and speech impairments to communicate effectively in a digital world.

Our Target Market:

- Educational institutions: Enhancing learning and inclusivity for Deaf and Mute students.
- Healthcare facilities: Improving communication with patients and promoting accessibility.
- Government agencies: Providing accessible services and fostering community engagement.
- Individuals: Empowering Deaf and Mute individuals with independent communication tools.

### Mission/Vision Statement

Mission: SignSenseTech is on a mission to bridge the communication gap for the Deaf and Mute community through advanced technology, fostering inclusivity and empowerment.

Vision: We envision a world where technology ensures that no individual is left unheard, where communication is accessible to all, regardless of hearing or speech abilities.

### Core Values

**Inclusivity:** We believe in creating technology that includes everyone. Our solutions are designed to empower individuals with diverse abilities, ensuring that no one is left behind.

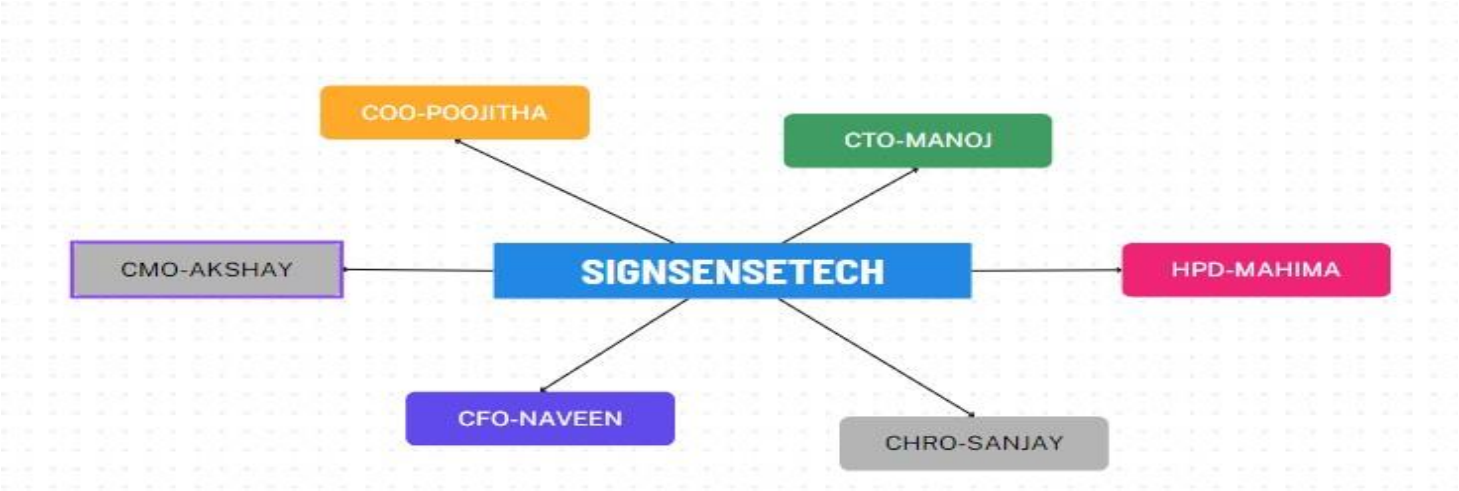
**Innovation:** We embrace a culture of continuous innovation, pushing the boundaries of technology to develop cutting-edge solutions that positively impact the lives of our users.

**Empathy:** Understanding the unique needs of the D&M community is at the heart of what we do. We approach our work with empathy, striving to make a meaningful difference in the lives of those we serve.

**Collaboration:** We value collaboration, both internally and externally. By partnering with the D&M community, researchers, and industry experts, we can collectively create more effective and impactful solutions.

**Quality:** We are committed to delivering high-quality products that exceed expectations. Our dedication to excellence ensures that our solutions are reliable, accurate, and user-friendly.

## Team & Org Structure




- CTO-**Chief Technology Officer : Manoj**

  Responsible for leading technological advancements and plays a pivotal role in driving innovation, overseeing the development of cutting-edge solutions, and ensuring the technical excellence of our products.
- COO-Chief Operations Officer : **Poojitha**

  Oversees the operational aspects of SignSenseTech, ensures smooth day-to-day functioning, manages partnerships, and focuses on scaling operations to meet the demands of our growing user base.
- CMO-Chief Marketing Officer : Akshay

  Leads the marketing efforts, and is responsible for developing and executing strategies to promote SignSenseTech's products, emphasizing their impact on communication accessibility.
- CFO-Chief Financial Officer : Naveen

  Heads the financial operations, oversees budgeting, financial planning, and reporting to support the company's growth and sustainability.
- CHRO-Chief Human Resources Officer : Sanjay

  Responsible for the human resources aspect of SignSenseTech, focuses on talent acquisition, employee development, and creating a positive workplace culture.
- HDP-Head of Product Development : **Mahima**

  Leads the product development team, ensuring the seamless integration of text and audio features, plays a crucial role in delivering high-quality and user-friendly applications.

## Products Line

### Product Offerings

**ASL Recognition System:**

- *Functionality:* Our core product utilizes a Convolutional Neural Network (CNN) to accurately recognize and translate American Sign Language (ASL) gestures into written text. Additionally, an integrated audio feature converts the recognized text into spoken language, facilitating communication for individuals with hearing and speech impairments.

### Service Offerings

**Custom Dataset Development:**

- We provide the service of creating custom datasets for organizations and researchers working in the field of sign language recognition. This service ensures that the recognition model is tailored to specific needs, enhancing accuracy and inclusivity.

**Integration and Training Services:**

- For businesses and institutions seeking to integrate our technology into their existing systems, we offer integration services and training programs. This ensures a seamless transition and optimal utilization of our ASL Recognition System.

### Pricing Model

**ASL Recognition System:**

- *Charge:* The ASL Recognition System will be priced as a one-time purchase or subscription model, depending on user preferences and usage patterns.
- *Markup:* The pricing will reflect the value of our technology, considering development costs, ongoing support, and continuous improvement. Subscription models will include regular updates and access to new features.

**SignSense Mobile Application:**

- *Charge:* The mobile application will be available for free download, with optional in-app purchases for premium features and customization options.
- *Markup:* In-app purchases will be reasonably priced, providing users with the flexibility to enhance their experience without creating financial barriers.

**Custom Dataset Development:**

- *Charge:* Custom dataset development services will be priced based on the scope and complexity of the project.
- *Markup:* The pricing will reflect the expertise and effort required to create high-quality datasets tailored to the specific needs of the client.

**Integration and Training Services:**

- *Charge:* Integration and training services will be offered as customizable packages, with pricing determined based on the level of support required.
- *Markup:* The pricing will consider the expertise of our team and the value-added by ensuring a smooth integration process for our clients.

The pricing model is designed to be competitive while ensuring sustainability and continued innovation. We aim to provide affordable access to our technology while maintaining the high standards of our products and services.

## Market Analysis

### Target Market:

- Primary: Deaf and Hard of Hearing individuals (DHH) of all ages and backgrounds.
- Secondary: Professionals who interact with DHH individuals, such as educators, healthcare providers, customer service representatives, and family members.

### Market Size:

- Global DHH population: 466 million (WHO, 2021)
- Indian DHH population: 18 million (WFD, 2022)
- Growing awareness and accessibility focus create significant potential for SignSenseTech's technology.

### Location Analysis:

- India: Large DHH population, growing tech sector, and government initiatives promoting accessibility, making it a prime market for SignSenseTech.
- Focus on urban centers: Higher DHH population density, better access to resources and technology infrastructure.

### Competitor Analysis:

| Competitor | Comparative Strength(s) | Comparative Weakness(es) | Counterpoint(s) |
|---|---|---|---|
| Competitor A (Global Company) | Established brand, advanced technology, large user base | High cost, limited focus on Indian market, lack of customization options | SignSenseTech's focus on affordability, Indian language support, and customizable features provides a competitive edge. |
| Competitor B (Indian Startup) | Affordable solution, Indian market focus | Limited language support, less advanced technology | SignSenseTech's superior language recognition engine and focus on user experience offer a clear advantage. |
| Competitor C (Open-Source Project) | Free and open-source, customizable | Limited features, lack of user support | SignSenseTech's user-friendly interface, comprehensive features, and dedicated customer support differentiate it from open-source options. |

.

This analysis provides a starting point for understanding your target market, buyer personas, and competitive landscape. Further research and refinement will be crucial for developing a successful marketing and sales strategy for SignSenseTech.

## Marketing Plan

### Positioning Strategy

SignSenseTech removes communication barriers by accurately recognizing ASL gestures and converting them into text and audio, empowering D/m individuals and fostering inclusivity. Our AI-powered CNN model sets us apart with high accuracy and real-time functionality, exceeding traditional sign recognition methods. SignSenseTech caters to various applications, from education and healthcare to workplace communication and entertainment. Our goal is to make SignSenseTech accessible and affordable for individuals, organizations, and educational institutions.

SignSenseTech empowers D/m individuals by providing a voice, enhancing their independence and social interactions. Our technology facilitates effective ASL learning and teaching, improving communication skills and accessibility in educational settings. SignSenseTech aids in better diagnosis and communication with D/m patients, enhancing healthcare quality and patient experience. Foster inclusivity and accessibility in the workplace by removing communication barriers for D/m employees and customers.

### Acquisition Channels

**Targeted online advertising:**

Utilizing search engine marketing (SEM), social media advertising (Facebook, Instagram, LinkedIn), and targeted banner ads on relevant websites.

**Partnerships and co-marketing:**

Collaborating with D/m organizations, educational institutions, healthcare providers, and assistive technology companies for cross-promotion and outreach.

**Content marketing:**

Creating informative blog posts, articles, and videos about ASL, accessibility, and the benefits of SignSenseTech.

**Event marketing:**

Participating in relevant conferences, workshops, and trade shows to showcase the technology and connecting with potential buyers.

## Sales Plan

### Sales Methodology

SignSenseTech will primarily adopt an inbound sales strategy, leveraging digital marketing to attract potential customers interested in sign language communication solutions. We'll utilize content marketing, social media, and targeted online advertising to generate leads. The focus will be on showcasing the unique features and benefits of our Sign Language Recognition System.

### Sales Organization Structure

**Sales Team:**

A dedicated sales team will handle customer interactions, inquiries, and conversions.

**Marketing Collaboration:**

Sales and marketing will work closely together to align messaging, ensuring that the unique selling propositions are effectively communicated.

**Roles Breakdown:**

The sales team will include Sales Representatives, Account Managers, and a Sales Manager overseeing the entire process.

### Sales Channels

**Online Platform:**

SignSenseTech will primarily sell its products through an online platform, providing easy access for customers to browse, purchase, and receive support.

**Distribution Partnerships:**

Exploring partnerships with educational institutions, assistive technology providers, and retailers to expand the reach of our products.

### Tools and Technology

**Website and E-commerce:**

A user-friendly website with an integrated e-commerce platform for seamless online transactions.

**CRM Software:**

Implementing a Customer Relationship Management (CRM) system to manage leads, customer interactions, and track sales performance.

**Live Chat:**

Utilizing live chat for real-time customer support and engagement.

**Analytics Tools:**

Employing analytics tools to track website traffic, user behavior, and sales metrics for continuous improvement.

**Call Software:**

Implementing a reliable call software solution for effective communication with potential clients.

## Legal Notes

### Legal Structure

SignSenseTech will be structured as a Limited Liability Company (LLC). This structure offers a balance between liability protection and operational flexibility, making it suitable for a technology-focused venture like ours.

### Legal Considerations

**Business Registration:**

Will register SignSenseTech as a LLC in accordance with local regulations, ensuring compliance with business registration requirements.

**Intellectual Property Protection:**

We have taken measures to protect our machine learning model and related technologies through patents and copyrights where applicable, safeguarding our intellectual property.

**Data Protection:**

Given the sensitive nature of user data involved in sign language recognition, SignSenseTech adheresto data protection laws and regulations, implementing robust security measures to safeguard user information.

**Health and Safety Regulations:**

SignSenseTech complies with health and safety regulations relevant to our industry. Our technology focuses on enhancing communication with the Deaf and Hard of Hearing community, ensuring it is safe and accessible.

**Insurance Coverage:**

We have acquired business insurance coverage to protect against potential risks, including liability and property insurance, to ensure financial stability in unforeseen circumstances.

**Zoning Laws:**

SignSenseTech has identified and adhered to zoning laws, ensuring that our business activities align with local zoning regulations.

**Accessibility Compliance:**

As our technology aims to enhance communication with the Deaf and Hard of Hearing community, we ensure compliance with accessibility standards, both in terms of our product and our physical business premises.

**Employee Regulations:**

SignSenseTech adheres to employment laws, providing fair wages and maintaining a healthy working environment. We have also implemented measures to support diversity and inclusion within our workforce.

**Tax Obligations:**

We are compliant with all tax obligations, including income tax, sales tax, and other relevant taxes. Our financial records are well-maintained to facilitate accurate reporting.

**Customer Data Consent:**

To maintain transparency, SignSenseTech has implemented clear privacy policies and obtains user consent for data collection and usage, aligning with privacy regulations.

## Financial Considerations

### Startup Costs

| Description | Cost |
|---|---|
| Research & Development (initial model development) | ₹25,00,000 |
| Software Development (recognition engine & mobile app) | ₹40,00,000 |
| Marketing & Branding | ₹13,30,000 |
| Legal & Accounting Fees | ₹5,00,000 |
| **Total Startup Costs | ₹83,30,000** |

Cost Reduction Strategies:

- Utilize open-source libraries and data sets for R&D.
- Partner with universities or research institutions for grants and resources.
- Outsource non-critical software development tasks to cost-effective options.
- Focus on organic marketing strategies and partnerships with Deaf & Mute groups.
- Consider cloud-based solutions for cost-effective server and infrastructure requirements.

| Year | Q1 Revenue | Q2 Revenue | Q3 Revenue | Q4 Revenue | Annual Revenue |
|---|---|---|---|---|---|
| 1 | ₹11,55,000 | ₹16,92,500 | ₹23,10,000 | ₹34,65,000 | ₹86,22,500 |
| 2 | ₹53,06,250 | ₹63,75,000 | ₹74,43,750 | ₹85,12,500 | ₹276,37,500 |
| 3 | ₹1,06,12,500 | ₹1,17,50,000 | ₹1,28,87,500 | ₹1,40,25,000 | ₹492,75,000 |

**Break-Even Analysis**

SignSenseTech expects to break-even within approximately 18 months of launch, achieving total sales of about ₹1,29,37,500.

**Projected P&L**

| Year | Revenue | Cost of Goods Sold | Operating Expenses | Gross Profit | Net Income |
|---|---|---|---|---|---|
| 1 | ₹86,22,500 | ₹25,87,500 | ₹47,43,750 | ₹13,00,250 | ₹-34,43,500 |
| 2 | ₹276,37,500 | ₹82,91,250 | ₹90,62,500 | ₹102,83,750 | ₹-12,18,750 |
| 3 | ₹492,75,000 | ₹147,82,500 | ₹138,37,500 | ₹206,55,000 | ₹68,17,500 |

Funding Requirements

SignSenseTech seeks ₹83,30,000 in seed funding to cover initial costs for R&D, marketing, and key personnel hiring.

# Conclusion:

In this report, a functional real time vision based American Sign Language recognition for D&M people have been developed for asl alphabets.

We achieved final accuracy of **98.0%** on our data set. We have improved our prediction after implementing two layers of algorithms wherein we have verified and predicted symbols which are more similar to each other.

This gives us the ability to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate.

In transforming this project into a business plan, SignSenseTech envisions providing a scalable and sustainable solution. The intersection of machine learning, accessibility, and communication technology positions SignSenseTech as a socially responsible venture with the potential to make a positive impact on the lives of many. The commitment to ongoing research, user feedback, and technological advancements will be integral to the company's growth and success in delivering inclusive communication solutions.