# Sign Language Translator using Machine Learning for Communication with Deaf People

Prabhjot Kaur
Associate Professor
MSIT, GGSIPU
New Delhi

Mohit Kumar
Student,
MSIT, GGSIPU
New Delhi

Shivam Garg
Student,
MSIT, GGSIPU
New Delhi

Sharad Tanwar
Student,
MSIT, GGSIPU
New Delhi

Shivam Kumar
Student,
MSIT, GGSIPU
New Delhi

## ABSTRACT
Sign language is an incredible advancement that has grown over the years. Unfortunately, the language has some drawbacks. Not everyone knows how to interpret sign language when conversing with deaf people. Communication using sign language is always necessary. Communication without an interpreter is difficult. To solve this, there is a need to develop a product that is versatile and robust, which can convert sign language so that ordinary people can understand it and communicate without barriers. The main aim of this paper is to break down barriers between deaf and non-deaf people and propose a system so that deaf and normal people can communicate with each other.

## Keywords
Sign Language, Translator, Gesture Recognition

## 1. INTRODUCTION
American Sign Language (ASL) is a natural syntax with the same roots as the spoken language, but with a completely different grammar. ASL can be expressed in the fate of bodily movements. In Native America, deaf or blind people are a reliable source of absurdity. Sign language has no formal or familiar form. Different signal languages speculate in certain areas. For example, British Sign Language (BSL) is a completely different language than ASL, and people in the US who are familiar with ASL will not easily understand BSL. Some countries have adopted ASL features for their national sign language. Sign language is a method of oral communication for people with speech and hearing impairments.

Worldwide, approximately 360 million people, including 328 million adults and 320 million children, are affected by hearing damage. Disabling hearing loss is defined as hearing loss that exceeds 40 dB in the better ear. As the number of deaf people increases, so does the need for interpreters. Minimizing the language exchange gap between deaf and non-hearing people becomes a desire to have specific and effective conversations between all people. It is one of the current research directions and the most natural way of communication for people with hearing impairments. A hand gesture recognition device can provide deaf people with the ability to speak out loud without the need for an interpreter. The system is designed to automatically convert ASL to text content and speech.

The proposed system is intended to understand some very basic elements of signal language and translate them into text and speech. American Sign Language is a visible language. Along with signatures, thought processes linguistic data through vision. The shape, placement, hand movements, facial expressions, and movement of the frames are all essential elements in conveying the facts. Sign language is not a common language throughout the World. Sign language has its own unique system of signals, and different regions have their own dialects, much like spoken languages have various dialects around the world. The detection rate of American Sign Language is around 90%, though different institutions may use other sign languages, such as the Indian Sign Language. Grammatical accuracy is important in sign language communication, just as it is in spoken language communication. Amazing element of India, it [ISL] has a little difference in sign language, but the grammar is the same in some places in the US. Deaf people in India observe herbal interactions with people around them. I remember it being much better than my own sign language because it is a natural way. The level of mastery of sign language is the same as spoken language. Toddlers start by walking on their hands. Since there are not many institutions for the dissemination of Indian Sign Language in India (except ISLRTC, which was established last year: the future of ISL will be), some people lack understanding and some institutions says that they are voting for ASL over ISL without proper knowledge.

## 2. LITERATURE SURVEY
Rafiqul Zaman Khan and Noor Adnan Ibraheem [1] in their paper surveyed the key issues and challenges faced in hand gesture recognition. One of the key challenges were the extraction methods and image pre-processing If the gesture is a static gesture, the image should only be segmented, but if it is a dynamic gesture, the hand gesture should be located and then tracked. To track the hand, it is much easier to locate the hand by its skin color because it's easy and immutable to scale, shift and rotate. They have implemented various methods such as Gaussian Model (GM) and Gaussian Mixture Model (GMM) for parametric techniques and histogram-based techniques are non-parametric techniques.

Mokhtar M. Hasan [2] using nongeometric features, a multivariate Gaussian distribution can be used to identify hand motions. The input hand image was segmented using two separate approaches: cluster-based thresholding techniques and skin color-based segmentation using the HSV colour model. In order to extract the hand feature, some operations were carried out to capture the hand's shape. The modified Direction Analysis Algorithm was then adopted to determine the relationship between statistical parameters (variance and covariance) from the data, and it was used to compute the object's slope and trend (the hand) by determining the direction of the hand gesture.

Munir Oudah, Ali Al-Naji and Javaan Chahl [3] in their paper put a light on the limitations of the glove based detection of hand signs in which the computer could be connected to a sensor that is attached to a glove, but this method has limitations that may make it unsuitable for elderly individuals. The wire connection may cause discomfort or confusion, and those with chronic diseases may have difficulty putting on or taking off gloves, leading to discomfort and constriction if worn for extended periods. Additionally, there is a risk of infection, skin damage, or adverse reactions for those with sensitive skin or burns. Furthermore, some sensors can be expensive. Lamberti and Camastra addressed some of these issues in their research by developing a computer vision system that relies on colored marked gloves, eliminating the need for attached sensors. However, the system still requires individuals to wear colored gloves.

In his paper, Sadaoki Furui [4] conducted a comprehensive survey of the major themes and advancements in speech communication research over the past 50 years. This was done in order to gain a more complete technological perspective and to recognize the significant progress that has been achieved in this field. Furui argues that in order to further advance speech processing technology, more research is needed on the human speech process itself.

Norani Mohamed, Mumtaz Begum Mustafa, and Nazean Jomhari [5] conducted a comprehensive review of the vision-based hand gesture recognition system over a period of seven years. They examined the progress made in the field and identified the key issues that needed to be addressed for the technology to be effective in real-life applications. Their research highlighted the significance of data acquisition, features, and the environment of the training data in improving the accuracy and effectiveness of the system. They found that most of the data used in previous studies was collected using a single camera in a restricted environment, which restricted the system's ability to recognize hand gestures in different environments. This suggested the need for sign language databases that are less restrictive and contain different environments to enable the system to reco3. gnize hand gestures in any form of environment. The authors concluded that more attention should be centered on the uncontrolled environment scenario to give researchers the chance to enhance the system's capacity for hand gesture recognition in any kind of environment. Their research highlights the potential for vision-based hand gesture recognition technology and the need for continued research to improve the technology's effectiveness in real-life applications.

D. Raj Reddy [6] analyzed the progress made in the field and identified areas of difficulty that needed to be addressed. He focused on reviewing research progress and explaining the reasons behind the challenges and the methods being used to overcome them. Reddy observed that connected speech recognition was not yet possible and suggested that more work was needed to improve the accuracy and effectiveness of the technology. To address this issue, he proposed implementing techniques for the codification and use of phonological rules in speech recognition systems, which would improve the ability of the technology to recognize and interpret speech patterns. Reddy's research highlights the potential of speech recognition technology and the need for continued research to overcome the challenges and improve its effectiveness.

Viraj Shinde, Tushar Bacchav, Jitendra Pawar, Mangesh Sanap [7] worked on on developing a system that could accurately recognize hand gestures in real-time using affordable and accessible equipment. They used the new technologies and software, and proposed an architecture which consisted of three stages Hand detection, Hand gesture recognition, Finger detection. There methods helped to reduce external interface like mouse and keyboard and thus made it highly portable also. Their conclusion highlighted that their research could lead to a new era of human-computer interaction, where physical contact with devices would no longer be necessary. With the proposed system, anyone would be able to operate a computer seamlessly using gesture commands, making it more accessible and user-friendly.

# 3. SCOPE OF WORK

To accomplish the stated objectives, this study combines qualitative and quantitative research methodologies. Sign language is a system of conventional gestures, mimics, hand signs, and finger spelling, often used to represent complete ideas or phrases. Its main purpose is to provide speech and text output through hand gesture sign language, particularly for individuals who are unable to communicate through traditional means. This proposed system operates without the use of sensors, making it a smart and accessible solution for the deaf and mute community. To simplify things, we have also implemented text-to-speech function so that you can also listen to the message that the user want to say.

Next, a number of machine learning algorithms are used in the research. We assess and contrast each model's performance and look for ways to further boost forecast accuracy. When using different Python libraries, such as openCV, Matplotlib, TensorFlow, SKlearn, etc. from the stage of pre- processing data to the level of model evaluation, we use the Anaconda distribution.

Previous studies did not apply freshly created, superior algorithms to conventional ones like logistic regression in the categorization modelling portion. We build on earlier research by employing these techniques to forecast the word. We provide examples of how crucial features are. We validate the model using a different method and assess its performance using performance indicators.

The results of this research are anticipated to offer communication with dumb and deaf people without any help of the interpreter or any external sensor or device. It helps in minimising the communication gap between the people and communication can be done without any prior knowledge of sign language.

# 4. MATERIALS AND METHODS

## 4.1 Data Pre processing

The methods of data processing here involve using techniques and algorithms to extract information from data, which can vary depending on the information being processed.

Using Image Processing and machine learning methods, the image of the sign is interpreted into allure equivalent discussion. Extraction of the features of the representation is accomplished by way of concept processing and therefore the feature heading is classified utilizing machine learning techniques.

Capturing signs from webcam and translating bureaucracy is the objective concerning this work. The real-world signs are express utilizing a webcam that captures both static and dynamic images using webcam using OpenCV library in this project. The deaf and dumb person individual who is signing is created to signs utilizing skilled hands earlier than the webcam and the concept captured from this is treated accompanying matplotlib study to plan the breakpoints of the person signing.

Matplotlib fundamentally sketches out a stick figure of the material. When the webcam is running the pose estimation treasure labels the indispensable content on the subject's frame in the way that elbow joints, wrist, body part intersections etc and combines bureaucracy all at once skeleton. The indispensable content that is to say completely points of the frame are marked with x, y and z coordinate for each frame rounded up.

As such key points content are recognized from the algorithms. The coordinates of these relates change for various gestures and the relative distance middle from two points the key points is various for various persons. These coordinates are the main component to form the data set for training.

## 4.2 Model Training

Training a machine learning model involves feeding a learning algorithm with training data to enable it to learn from the examples provided. The ML model refers to the resulting artifact generated by the training process, which can be used to make predictions or classify new data.

To train an ML model, the training data must contain the correct answer, referred to as the target or target attribute. The training algorithm analyzes the training data and identifies patterns that map the properties of the input data to the target response. This process results in the creation of an ML model that captures these patterns and can be used to make predictions on new data.

Supervised learning is a subfield of machine learning in which an algorithm attempts to learn a function that maps an input to an output based on labeled training data. The training data consists of input-output pairs, and the algorithm learns to recognize patterns and relationships between the inputs and outputs. Using this knowledge, the algorithm can then make predictions on new, unseen data. In other words, supervised machine learning algorithms are trained on past data in order to identify the optimal input-output relationships that will produce accurate predictions for future data.

There are two types of supervised learning algorithms first is regression algorithms and second one is classification algorithms.

Regression-based supervised learning methods attempt to predict outputs based on input variables. Classification-based supervised learning methods determine the category to which a set of data items belongs.

Classification algorithms are based on probability, which means that the outcome is the type to which the algorithm finds the highest probability that the data set belongs to it. On the other hand, regression algorithms estimate the outcome of problems with infinitely many solutions (continuous set of possible outcomes).

The categorical prediction modeling problem is different from the regression prediction modeling problem because Regression is the challenge of predicting a continuous quantity, while classification is the task of predicting discrete class labels. Yet, there is a substantial overlap between the two models, and both share the same idea of using known factors to produce predictions.

Logistic regression is a type of supervised learning algorithm used for binary classification. It involves using one or more independent variables to predict the outcome of a dependent variable with two possible outcomes. The algorithm aims to find the best fit relationship between the independent variables and the dependent variable, and it is useful in explaining the factors that contribute to the classification. Logistic regression is a popular algorithm due to its simplicity, interpretability, and ability to handle both continuous and categorical independent variables.

The decision tree algorithm builds a tree-like structure that uses if-then rules for classification. It breaks down data into smaller structures and sequentially learns rules until the endpoint is reached. The final structure resembles a tree with nodes and leaves.

The random forest algorithm is an extension of the decision tree algorithm, where you first generate some decision trees using the training data, and then fit your new data into the algorithm. One of the "trees" is generated as a "random forest". It averages the data to connect it to the nearest tree data based on the scale of the data. These models are great for solving the problem of decision trees tying unnecessary data points into a category. Gradient Boosting is a machine learning enhancement system that represents a decision tree for large and complex data. It is based on the assumption that the next possible model will minimize the raw prediction error if combined with the previous set of models.

The k-nearest neighbor algorithm, also known as KNN or k-NN, is a non-parametric supervised classification algorithm that uses close distance to perform classification or prediction on a one-point cluster. of individual data. It is distance-based: it classifies objects based on the classes of their nearest neighbors. Although it can be used for regression or classification problems, it is often used as a classification algorithm, assuming that similar points can be found side by side.

The KNN algorithm in the training phase stores only the data set, and when it receives new data, it classifies those data into a category that is very similar to the new data. The parameter k in kNN refers to the number of labeled (neighborhood) points considered for classification. The value of k indicates the number of points used to determine the outcome. Our task is to calculate the distance and determine the closest categories to our unknown entity.

LDA is a supervised learning algorithm that finds a linear discriminant function to classify two types of data points. It requires labeled training data and aims to maximize the distance between class means while minimizing the variance. It differs from PCA, which reduces the size of data.

# 5. RESULTS AND DISCUSSION

This section discusses overfitting, cross-validation, and evaluation metrics as key elements of model performance evaluation, which can be used in combination with grid search to find the optimal hyperparameters.

Overfitting is a common problem in machine learning where the model fits the training data too closely, resulting in poor performance on unseen test data. This issue arises when the model is overly complex and learns non-representative features from the training data. In contrast, missing pages is when the model is too simple to capture the underlying patterns in the data.

Machine learning models need to be trained to generalize well to new data, and a common issue in achieving this is the bias-variance trade-off, which involves balancing the risk of overfitting or underfitting. Cross-validation is a technique that addresses this challenge by dividing the data into training and validation sets, allowing for reliable estimates of the model's performance. For instance, during k-fold cross-validation, the training data is randomly split into k-folds. The model is trained on k-1 folds and evaluated on the remaining fold, and this process is repeated k times to obtain an average score.

Cross-validation can be a computationally expensive process, especially when used in combination with grid search to tune hyperparameters. However, modern machine learning libraries like sklearn provide easy-to-use functions for performing cross-validation with just a few lines of code.

# 6. EVALUATION METRICS

Choosing the appropriate evaluation metrics is essential in evaluating the performance of a machine learning algorithm. Different evaluation metrics may be more relevant depending on the specific problem, and the criteria used to evaluate the algorithms will influence which algorithm is ultimately chosen. For example, in a binary classification problem, metrics such as accuracy, precision, recall, and F1-score can be used to evaluate the performance of different algorithms.

Accuracy measures the proportion of correctly classified instances, while precision measures the proportion of correctly classified instances among all instances classified as positive. Recall, on the other hand, measures the proportion of positive instances that were correctly classified. F1-score is a metric that combines precision and recall to give a single value that represents the overall performance of the model. In addition to these metrics, other metrics such as ROC curves and AUC (Area Under the Curve) can also be used to evaluate the performance of a classification model.

Accuracy is the number of correct predictions made as the ratio of total of the predictions made. This is the most common measure of classification problems and it is also the most abused one. It is more appropriate to have an equal number of observations in each class (still this is rarely the case) and when all predictions and related prediction errors are of equal importance, this is usually not the case. happen. Accuracy is the percentage of positives out of the total number of predicted positives. Here, the denominator is the model prediction made as positive from the given data set. Accuracy is a good metric for determining when the cost of misinformation is high (e.g. spam detection in email).

Recall is a machine learning evaluation metric that measures the ability of a model to correctly identify all positive cases in the dataset. It is also known as sensitivity or net positive rate. Recall is calculated by dividing the number of true positives by the sum of true positives and false negatives. The denominator represents the actual number of positive cases in the dataset, while the numerator represents the number of positive cases that the model correctly predicted.

Recall is particularly important when there are high costs associated with false negatives. For example, in the case of fraud detection, a false negative could result in a significant financial loss. In this scenario, it is crucial to minimize the number of false negatives, and recall can be used as a reliable metric to measure the model's performance. By maximizing the recall, the model can effectively identify as many true positives as possible, reducing the risk of false negatives and their associated costs.

Other commonly used measures of classification performance include F1 score, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The F1 score is the harmonic mean of precision and recall and provides a balanced measure between the two. Specificity measures the proportion of true negatives to the total number of negatives in the data set. It is a useful metric when the cost of false positives is high, such as in medical diagnoses. AUC-ROC is a measure of the model's ability to distinguish between positive and negative classes and provides a single scalar value that represents the overall performance of the model. AUC-ROC is especially useful when the positive and negative class sizes are imbalanced. Choosing the right evaluation metrics depends on the specific classification problem and the trade-offs between different types of classification errors.

**Table 1. Classification results for threshold = 100**

| S.No | Classifier | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| 1 | Logistic Regression | 0.38 | 0.54 | 0.43 | 0.58 |
| 2 | Decision Tree | 0.86 | 0.86 | 0.85 | 0.86 |
| 3 | Random Forest | 0.89 | 0.89 | 0.88 | 0.89 |
| 4 | Gradient Boosting | 0.86 | 0.86 | 0.86 | 0.86 |
| 5 | KNN | 0.83 | 0.82 | 0.82 | 0.82 |
| 6 | LDA | 0.7 | 0.75 | 0.71 | 0.75 |

## 7. CONCLUSION

Machine learning has the potential to greatly improve the accuracy and speed of sign language translation. With enough data and the right algorithms, a machine learning model can learn to accurately recognize and translate sign language gestures into written or spoken language. This can help to bridge the communication gap between deaf individuals and those who do not know sign language, and can also make it easier for deaf individuals to access written or spoken information. However, it is important to note that machine learning is not a perfect solution and there may still be some errors or misunderstandings in the translation process. It is also important to consider the ethical implications of using machine learning for sign language translation, including the potential for biases in the data and the potential for replacing human interpreters with machines.

It is likely that the use of machine learning for sign language translation will continue to improve and become more widespread in the future. As the technology and algorithms advance, the accuracy and speed of sign language translation is likely to increase, making it a more viable option for a wider range of applications. In addition, the increasing availability of data and the development of more sophisticated machine learning models will likely lead to the creation of new and innovative solutions for sign language translation.

There are also likely to be ongoing debates about the ethical implications of using machine learning for sign language translation. Some people may be concerned about the potential for replacing human interpreters with machines, while others may see it as a way to improve access and communication for deaf individuals. It will be important to consider these ethical issues carefully as the use of machine learning in this area continues to develop.

The future of sign language translation using machine learning is bright. As machine learning technology continues to advance, it is likely that the accuracy and speed of sign language translation will improve. This can help to further bridge the communication gap between deaf individuals and those who do not know sign language, and can also make it easier for deaf individuals to access written or spoken information.

There are also potential applications for sign language translation beyond just facilitating communication between deaf individuals and those who do not know sign language. For example, sign language translation technology could be used to create captions for video content, or to enable deaf individuals to interact with voice-based artificial intelligence assistants.

However, it is important to consider the ethical implications of using machine learning for sign language translation. There may be concerns about the potential for biases in the data, or the potential for replacing human interpreters with machines. It will be important for researchers and developers to carefully consider these issues as they continue to work on improving sign language translation using machine learning.

## 8. REFERENCES

[1] Khan, Rafiqul Zaman, and Noor Adnan Ibraheem. "Hand gesture recognition: a literature review." *International journal of artificial Intelligence & Applications* 3, no. 4 (2012): 161.

[2] Hasan, Mokhtar M., and Pramod K. Mishra. "Hand gesture modeling and recognition using geometric features: a review." *Canadian journal on image processing and computer vision* 3, no. 1 (2012): 12-26.

[3] Oudah, Munir, Ali Al-Naji, and Javaan Chahl. "Hand gesture recognition based on computer vision: a review of techniques." journal of Imaging 6, no. 8 (2020): 73.

[4] Furui, Sadaoki. "50 years of progress in speech and speaker recognition research." ECTI Transactions on Computer and Information Technology (ECTI-CIT) 1.2 (2005): 64-74.

[5] Mohamed, Noraini, Mumtaz Begum Mustafa, and Nazean Jomhari. "A review of the hand gesture recognition system: Current progress and future directions." IEEE Access 9 (2021): 157422-157436.

[6] Reddy, D. Raj. "Speech recognition by machine: A review." Proceedings of the IEEE 64, no. 4 (1976): 501-531.

[7] Shinde, Viraj, et al. "Hand gesture recognition system using camera." Int. J. Eng. Res. Technol.(IJERT) 3.1 (2014).