Sign Language Recognition and Training Module

Anjana Devi V (anjanadevi.aby06@gmail.com)
Rajalakshmi Institute of Technology

Charulatha T
Rajalakshmi Institute of Technology

Dharishinie P
Rajalakshmi Institute of Technology

Research Article

Keywords: Open CV, Random Forest algorithm (RF), Support Vector machine (SVM), K-Nearest Neighbor (KNN), Media pipe, American sign language (ASL), Sign language recognition (SLR)

Posted Date: July 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3057185/v1

License: ☛ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.
Abstract

Recognition of sign language has long been essential to communication between the deaf and non-verbal cultures. From early electric signal-based sign language identification to more recent recognition using machine/deep learning techniques, researchers from all over the world have made an effort to automate this process. Key point detection-based sign language recognition (SLR) is the major goal of this study. The focus of this work is American Sign Language (ASL), particularly ASL pickle data. Several machines learning algorithms, including random forest, support vector machine, and k nearest neighbour, were used to train the model. Lastly, the best model is selected from the model testing using assessment parameters including f1score, precision, and recall. To gather user input, a simple GUI is made, and the forecast is then made using the best machine learning model. Additionally, a training tool for learning American sign language has been developed, which will have a significant impact on non-verbal cultures.

1. INTRODUCTION

There are more than 70 million hearing-impaired people worldwide, and about the 80% of them live in developing nations, according to the World Federation of the Deaf. The preferred mode of communication for people who have hearing loss is sign language. To varying degrees, sign language can be recognized using computer vision or any other method. Some claim that sign language is composed of a set of movements, each of which has a specific meaning. The world and other deaf and dumb people can communicate thanks to its meaning. We suggest a mechanism that would facilitate sign language comprehension. To do this, we typically use support vector machines (SVM), and for the classification challenge, we also used random forest algorithms (RF) and K-nearest neighbor (k-NN). Additionally, we are developing a training scheme for anyone who wish to learn how to sign. This will make it easier for many individuals who cannot afford to send their children to school for learning the language, and it will also allow them to speak with those who are hard of hearing without difficulty. By doing this, sign language is being learned and used more widely. This article's goal is to review the techniques for deciphering sign language and create the greatest training aid for those who are visually impaired.

The choice of features is essential for gesture identification since hand movements are so distinctive in terms of shape variation, texture, and velocity. It is simple to determine hand posture by separating out variables such as finger orientations, fingers, skin color, and hand shapes for static hand identification. Due to the lighting and image background, those features are not consistently available and reliable. The silhouette, color, and textures are just a few more nongeometric components that are sufficient for recognition. Since it is challenging to precisely characterize features, the entire frame or converted image is used as the input. The recognizer then implicitly and automatically finds the features.

Developing systems that can identify unique gestures and use them to transmit information or control equipment is one of its main objectives. Yet, hand poses are the static structure of the hand, whereas gestures are the dynamic movement of the hand, and gestures need to be represented in the spatial and temporal domains. The two primary techniques used to analyze the use of hand gestures is a form of
data glove and vision-based technique. The primary objective is to design a vision-based system is capable of real-time sign language recognition. A system based on vision is preferable because it offers a clearer and simpler and realistic means of communication between a human and a machine. The system will take the silhouette of the hand that is placed in front of them as the input. By this means it can recognize the shape of the hand even its bit blurry. Our deepest concern is that sometimes the silhouette of the shape of the hand is not enough to determine the character. For that we have tried our best to debunk the gesture the person is trying to make.

As we can see from Fig. 1.2 that the basic process of SLR is quite recognizable. It follows the basic ideology of machine learning algorithms. The SLR process is first trained with a dataset about how each hand gestures will be made in order to show us a character. This is then processed and the features of the hand is taken as a reference for the model to compare. This is then implemented when the input is acquired by the means of video capturing. The input is taken by studying the features of the hand. Then it compares it with the dataset that is stored in the database and gives us the output as a letter that the person is displaying.

2. LITERATURE SURVEY

The survey was performed on the publications that appeared between 2015 and 2022 on the following platforms:

1. IEEE Xplore, (b) Science direct, (c) ACM journal and (d) Springer. These journal articles help us better understand the computer-aided analytical process that yields a summary of the whole body of research.

In this study,[1] CNN is used for HGR, which simultaneously takes into account ASL's alphabets and numerals. Additionally addressed are the benefits and drawbacks of CNNs utilized for HGR. Modified Alex Net and modified VGG16 models for categorization form the foundation of the CNN architecture. A multiclass support vector machine (SVM) classifier is employed after feature extraction, which is done using modified pre-trained VGG16 and Alex Net architectures. Without any human oversight, it automatically recognizes crucial features. It handles picture classification effectively and precisely. high cost of calculation. To obtain good accuracy, a large amount of training data is necessary. lack of spatial invariance with respect to the input data. The object's position and orientation are not encoded. Without any human oversight, it automatically recognizes crucial features. It handles picture classification effectively and precisely

There is an increasing need to integrate the enormous number of deaf-mute persons in China into society through the use of effective sign language processing technologies [7]. This review emphasizes the aspects studied in sign language recognition research, discusses categorization, offers the datasets available, and identifies trends for future research. It also gives a summary of the most significant work on Chinese sign language recognition and translation. Users merely need to use their hands inside the camera collection range and are not need to wear any other external equipment. low price. convenient
and simple to operate. It is greatly affected by environmental factors such as light, skin color, and occlusion. It requires several image processing techniques which might affect the recognition accuracy.

Introduces the software, data mining, and knowledge finding skills necessary for the data scientist to proceed towards meta-analytics, or the upcoming generation of analytics, in the systems that are hybrid by design and use multiple analytics to derive relevant information about the data [4]. At the end of the chapter, two larger parts will demonstrate how to construct a classifier from scratch using many of the statistical techniques. The average prediction performance is enhanced. High precision and a more stable model are provided. The variation of prediction mistakes is decreased. It might be more challenging to interpret. When utilizing the ensemble learning method, the model may occasionally be either overfitted or underfit.

The project creates a multilingual sign language framework that also models hand movement by deriving hand movement components from target sign language independent data [6]. We validate the proposed methodology and demonstrate that sign language recognition systems may be successfully created by utilizing multilingual sign language resources through research on Swiss German Sign Language, German Sign Language, and Turkish Sign Language. It does admirably when it comes to recognizing. It is easier to implement and analyses. The issue of label bias is resolved. There are frequently many unstructured parameters in HMMs. To get better results, extensive training is necessary. For training, a sizable dataset is needed.

Image processing, pattern recognition, and artificial intelligence are all regarded to be part of the sign language recognition (SLR) field of study. The blockage of one hand on another is the main obstacle for an SLR. As a result, the feature vector is formed with weak segmentations, which leads to incorrect classifications of the signs and a low recognition rate. It is helpful in situations where the learned target function needs to be evaluated quickly. The training dataset's noise doesn't affect it very much. It can tolerate flaws. Finding a suitable network structure is complex and computationally costly.

When directly applied to these issues, original SVM exhibit poor performance. We suggest using weighted support vector machines (WSVM) in this paper for automated process monitoring and early defect detection. We demonstrate the advantages of WSVM over conventional SVM and contrast them under various fault circumstances. We test the suggested approach using the most common anomalous quality control patterns as well as a genuine application from the wafer manufacturing sector in binary and multi-class settings. When dealing with continuous features and multidimensions, it performs better. It can be used in many different fields. Tolerance towards unimportant qualities. To obtain its highest level of prediction accuracy, a sizable sample of the dataset is necessary. Hyperparameters are frequently difficult to interpret when they have an influence.

3. PROPOSED SYSTEM

The proposed system of our model contains the following modules:
3.1 Data collection and pre-processing:

This proposed system is passed on the dataset of ASL (American Sign Language). The main idea behind using ASL is that it is one of the most common Sign Language in the world. Due to its communication which translates to English language its known world-wide which leave us the perfect dataset for all people to understand and learn. Image processing is a method for converting a physical image to a digital one so that a person can edit, add to, or remove information from it. A video frame or photograph serves as the input for this type of signal distribution, and the output might either be another image or attributes associated to that image. In the system Feature Extraction becomes the main job for image pre-processing. For this first the image is captured using live cameras and taken as input which undergoes a major background check from BGR to RGB for OpenCV () to work. Then captured images were passed to a Hand tracking model from the Media Pipe framework which was deployed using the library. This is the main step where the feature of the model has been extracted. Using media pipe coordinates has been given for the images. These coordinates serve as the data values for further carrying out the task of Sign Language recognition.

3.2 Algorithms Applied:

Random Forest

The algorithm constructs a collection of decision trees and combines their predictions to make a final prediction. Random Forest is a popular choice for sign language recognition due to its ability to handle complex, non-linear relationships between features and outputs. One advantage of Random Forest for sign language recognition is its ability to handle high-dimensional data, such as sign language data that includes multiple modalities. Performance of Random Forest, like any machine learning algorithm, will depend on the quality of the data and the appropriate selection of features and parameters. High recognition rates and accuracy are advantages of utilizing the Random Forest algorithm for sign language recognition. For instance, a study on Chinese Sign Language letter identification revealed that the random forest method outperformed artificial neural networks (ANN) with an average recognition rate of 95.48%. Another study suggested a pipeline that uses the random forest classifier to identify gestures in American Sign Language. The study built a sign language recognition model using three distinct algorithms, and the random forest classifier produced promising results. Overall, the Random Forest algorithm has demonstrated success in recognizing sign language, producing reliable and accurate results.

Support Vector Machine
Support Vector Machine can be used to classify hand gestures into corresponding words or letters in sign language. The algorithm works by finding the optimal boundary (or hyperplane) between the classes in a high-dimensional feature space, and then classifying new data points based on which side of the boundary they fall on. The SVM method has many advantages for sign language identification, including excellent accuracy and the capacity to categorize input signs. One study, for instance, suggested classifying input signs into various categories using SVM classification to create a sign language recognition system. Another study recognized Indian Sign Language using the SVM algorithm with a good degree of accuracy using the photos. Similar experiments were conducted on the classification of sign language using Support Vector Machines and Hidden Conditional Random Fields. A hand gesture detection algorithm utilizing SVM and HOG was also presented in order to recognize gestures quickly and in real-time. Overall, the SVM algorithm has demonstrated effectiveness in the recognition of sign language, with excellent accuracy and the capability to categorize input signs into various classes.

**K-Nearest Neighbor**

K-Nearest Neighbor (KNN) is a non-parametric, instance-based machine learning algorithm. In sign language recognition, KNN can be used to classify hand gestures into corresponding words or letters in sign language. The algorithm works by computing the similarity between a given test sample and the training samples, and then selecting the K training samples with the highest similarity. The KNN algorithm's capacity to categorize objects based on feature space and employ distance measurements as its classification criteria are two advantages of using it for sign language recognition. An Indian Sign Language identification system, for instance, was developed in a study employing the K-nearest neighbor (KNN) classifier, which categorizes objects based on feature space. KNN is one of the machine learning algorithms that uses distance measures as its classification criteria, according to another study. Similar to this, the KNN classifier was used to identify the signs in a sign language recognition system that was intended to help deaf-dumb persons. The KNN technique is also used in a study that developed a K-nearest correlated neighbor classification for Indian sign language recognition, which groups signs according to their features. Overall, the KNN algorithm has demonstrated success in classifying signs accurately based on feature space and distance measurements.

### 3.3 Model Training and Testing

Using the dataset, the model is trained. There will be two datasets namely training and testing. Usually training and testing datasets are separated. A training dataset is a starting set of data that is used to instruct machine learning models on how to recognize specific patterns or carry out a specific activity. The dataset that is utilized to train a machine learning model is rather vast. The machine learning model is trained or fitted using the training dataset, which is a crucial stage in creating machine learning models. Prediction models are taught to see patterns and generate predictions based on the data using the training dataset. Making sure that the training dataset is indicative of the issue being addressed is essential for the success of machine learning models. Once the model is trained it goes to the testing phase. This is done using a testing dataset. the testing dataset used to gauge its effectiveness. A testing dataset is a subset of data used to evaluate the performance and advancement of machine learning
algorithms and to adjust or improve their training for improved results. Unseen data is used in the testing dataset
to provide a fair assessment of how well a model fits the training dataset while changing model hyperparameters. Every machine learning model requires the testing dataset, which helps to confirm the model's accuracy and dependability. The model's performance is assessed using the testing dataset, and it may be modified or optimized for better outcomes. By doing this training process we are creating numerous models which are used for the evaluation purposes. For this we are selecting the best model by comparing the model using their evaluation metrics and the time they have taken for computing the output. After selecting the best model, we can proceed with the step of testing which occurs when the input is provided by the means of a web camera. This input is compared with the provided dataset inside a model and the output is given to us in the form of alphabets. For this we are selecting the best model by comparing the model using their evaluation metrics and the time they have taken for computing the output. After selecting the best model, we can proceed with the step of testing which occurs when the input is provided by the means of a web camera. This input is compared with the provided dataset inside a model and the output is given to us in the form of alphabets.

3.4 Sign Prediction and Self Mode Training:

The last step in the model that has been proposed is displaying the output. There are two steps in this. One is actual sign prediction and the other is self-mode training. The sign prediction is one of the models that has been created. The model to be displayed in the manner of English alphabets so that it will be an accurate description of the process. This will show us the perfect alphabet that we have been trying to predict. It also shows us the confusion matrix which will indeed result in the accuracy of the alphabet. The model proposed has a 98 percent accuracy as the least and 100 percent accuracy on most of the alphabets. Once the model has been predicted training model comes in.

The above model is connected with a web camera so that the system can take the live footage as the input for the model and the tool. By connecting the web camera, it will have the live camera footage for the model. The footage should be at least somewhat clear so that the device will provide accurate results. Taking input in the form of a live camera instead of a recorded video also provides a great benefit in the training tool as the person will know what mistake he has made and how to rectify it then and there instead of watching the video he made again and again. This makes a great difference in efficiency and it will help the person learn quickly and efficiently.

A raw footage of some letter is taken as the input of the model. This is achieved by reading the live camera feed which is acquired using a web camera. The next step in taking input is removing all the unnecessary background which is just an obstacle in the path of decoding. By removing the background, the model can achieve a more accurate shape of what the shape of the hand is. Binarization is the method of converting any entity's data characteristics into vectors of binary values to improve the performance of classifier algorithms. By doing this step, it makes it efficient for the document to achieve a more accurate shape of the hand so that the output will be precise. The individual pixels in an image
can be identified using a computer vision method called image segmentation. By doing this step we are making the image as a prefect input for the system as it can understand the language of decoding. By removing the background, the model can achieve a more accurate shape of what the shape of the hand is. Focuses more on details that are unwanted as the image is pixelated. This includes mainly the wrist but, in some cases, arms are also included if the whole arm is read by the system. The last main step in generating a perfect input for the model is centralization of the image. Once the small details are removed, there will be a more precise silhouette of the hand gesture. This is centralized for the model so that it can have a good input.

The next step in the model proposed is the comparison of the input to the trained model. Now this step occurs in the training model as shown in the above diagram. Once the input is entered this and follows the other step it will enter the chosen training model which has been selected with the help of the test and their speed. Once entered these inputs are compared with possible images from the dataset that the model finds quite alike to the input. After this the images are made sure that they are compared with every possible scenario that the model believes may match the silhouette of the image. After this the model will take the decision about the letter that the found about being quite an accurate description of the shape the hand is making to display the output and the output will be an alphabet the model is predicted. The training model will have the following steps as above and will get in once the sign is predicted.

Using the model created above we are planning to build a tool which will help people to learn ASL. This tool will make sure that the people has kept the signs of alphabets quite accurately. By creating this tool, we will make sure that everyone can learn sign language with ease and it will also help the deaf people who can't afford a school. This training tool will be integrated with the web camera available on the system. This is made sure that it captures the correct shape of hand which the person is trying to make. Once it reads the video feed by the camera, it will segment the video into frames. It will go through the above process in the selected model where the model will compare the available dataset with the video frame. It will make sure that the video frame will be compared with every possible image that is available in the dataset. There is also a time limit where the people can keep track of time when they learned their symbol. This is very crucial as both the computer and the person can have the idea of how well they are doing. If the person does it in less time it will take it as a failure and it will push the person more. For e.g. If a person is training to make an “A”, this training tool will make sure the person is making the sign quite accurate to the original and if the person comes for the second time it will take the previous time as example and will make sure the person does it in less time for it considering the learning as success. After this only the person can move forward to the next letter. By doing this we can make sure that everyone learns sign language accurately.

4. RESULTS

There are n numbers of hearing-impaired people around the world and most of them can’t afford school. There are also people who are interested in learning the ASL to communicate with all the persons in the
world and there are also people who wants to learn what they are trying to say. By the help of suggested method, we achieve all those things coincidently or simultaneously. The proposed idea not only focuses on just translating the letters the person is trying to make, but it also will help the person who wants to learn the letter and by doing this it will help all the person who are willing to learn but connect afford for any schools are institutions. Even though our model makes it easy for the people to learn the alphabets of the ASL it lacks behind when it comes to the topic of the complicated symbol which has an entire word as the meaning. This training tool as we have said above will mainly help the hearing impaired or deaf and dumb people. But it doesn’t need to stop there as a lot of people can use this tool to gain the knowledge in sign language and will be able to communicate freely. A confusion matrix is a table that lists how well a machine learning model performed using a sample of test data. An algorithm’s performance can be seen using a particular table arrangement, usually a supervised learning approach. The number of true positives, false negatives, false positives, and true negatives are reported in the matrix’s two rows and two columns. A confusion matrix has rows for each actual class and columns for each expected class. The confusion matrix aids in comprehending the classes that the model misinterprets as being different classes. It is a gauge of machine learning classification performance. The below figures are the outcome of the confusion matrix which has been taken in consideration for the comparison of different algorithms.

Figure 4.4 Comparison between algorithms

Comparing the above confusion matrices, the outcomes clearly show that the results found in random forest algorithm have more accuracy compared to that of the remaining algorithms. For that very reason this proposed system has taken Random Forest as its base algorithm.

5. CONCLUSION

In conclusion, the development of sign language recognition technology has significant implications for the non-verbal and hearing-impaired community, enabling greater communication accessibility and inclusion. Implemented a key point detection-based sign language recognition system for American Sign Language (ASL) using machine learning algorithms. Graphical user interface (GUI) training tool that captures user input and performs sign letter prediction and this project gives the accuracy of Random Forest 97%, KNN 96%, SVM 95%. This project’s success highlights the potential for future research to continue to improve the accuracy and efficiency of sign language recognition technology and ultimately make a significant impact on the lives of those who rely on sign language for communication.

Abbreviations

Random Forest algorithm (RF), Support Vector machine (SVM), K-Nearest Neighbor (KNN), American sign language (ASL), Sign language recognition (SLR)

Declarations
**Authors Declaration:**

- Ethics Approval: The manuscript has not been published elsewhere no is it under consideration in any other journal.
- Consent to Participate: The authors have given their consent for participation
- Consent for publication: All authors of this paper consent for publishing , manuscript, tables and figure in this journal.
- Availability of data and materials: The authors declare that data and related material not been published elsewhere no is it under consideration in any other journal.
- Competing interests: The authors declare that they have no competing of interest.
- Funding: Not Applicable
- Authors’ contributions: V. Anjana Devi - Wrote the main manuscript. Charulatha and Dharshinie - prepared figures and implementations. All authors Reviewed the manuscript

**References**

8. Steven J. Simske “Introduction, overview, and applications” Meta analytics (2019)
10. Abbas Muhammad Zakariya, Rajni Jindal Published 1 July 2019 Computer Science 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)


21. P. Subha Rajam, Dr G. Balakrishnan, —Real Time Indian Sign Language Recognition System to aid Deaf – Dumb People, 13th IEEE International Conference on Communication Technology (ICCT), Sep 2011, Secretary General School of Information Science and Engineering, Shandong University, Jinan, Shandong, China.


(Springer CCIS).

Figures

![Diagram](image-url)

Figure 1
Fig 1.1 – Working of Sign Language Recognition

![Diagram showing the working of Sign Language Recognition]

Figure 2

Fig 1.2 – Working model of SLR

![Diagram showing the working model of Sign Language Recognition]

Figure 3

![Diagram showing the working model of Sign Language Recognition]
Fig 3.1 Proposed system Architecture

**Figure 4**

Fig 3.2 Training tool design
Figure 5

Fig 4.1 Confusion matrix of Random Forest algorithm
Figure 6

Fig 4.2 Confusion matrix of Support vector machine
Figure 7

Fig 4.3 Confusion Matrix of K-Nearest Neighbour
Figure 8

Fig 4.4 Comparison between algorithms