IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Academic Writing

Title: Hand Gesture Recognition using Machine Learning

[1]Nidhi Ternikar, [2]Palash Kalghatgi, [3]Mrutyunjaya Emmi
[1]MCA student, [2] MCA student, [3]Associate Professor
[1]Department of MCA,
[1]K. L. S. Gogte Institute of Technology, Belagavi, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

ABSTRACT

Sign language is a visual language that uses hand motions, changes in hand shape, and track information to convey meaning. It is the primary mode of communication for those with hearing and language impairments. The use of sign language for communication is limited, despite the fact that sign language recognition can help a large number of such persons deal with regular people. As a result, there is a need to create a more comfortable approach for people with hearing and language impairments to learn and work in order to improve their lives. Therefore, the basic idea behind this article is to make the communication between normal human beings and deaf people much easier. In order to recognize static gestures associated with sign language alphabet and a few commonly used words, we conducted a comprehensive research study employing the hand tracking technique Mediapipe and a gesture classification model based on Support Vector Machine (SVM). The results of the experiments are validated using Recall, F1 Score and Precision. Based on the validated results, we recommend the application of the discussed techniques for such communication. The suggested methods have high generalization qualities and deliver a classification accuracy of around 99 percent on 26 alphabet letters, numerical digits, and some regularly used words[4].

INTRODUCTION

Humans communicate with one another using natural language channels such as words and writing, or by body language (gestures) such as hand motions, head gestures, facial expression, lip motion, and so forth. Comprehending sign language is equally as vital as understanding natural language. People with hearing impairment use sign language as their preferred mode of communication. Without a translation, people with hearing impairments have difficulty speaking with other hearing people. As a result, implementing a system that understands sign language would have a substantial positive impact on the social lives of deaf people. According to the World Health Organization, 466 million individuals worldwide (more than 5 percent of the population) have impaired hearing, with 34 million of them being teens (WHO). According to studies, by 2050[1], these numbers will have surpassed 900 million. Furthermore, the majority of cases of profound hearing loss, which afflict millions of individuals, occur in low and middle-income nations. There are more than 135 distinct sign languages spoken worldwide, including American Sign Language (ASL), British Sign Language (BSL), and Indian Sign Language (ISL). Machine learning enables the development of systems that accurately interpret sign language, which can greatly improve communication and social lives of deaf people. These technologies are particularly important for those living in low and middle-income nations where the majority of hearing impairments occur. The growing prevalence of hearing loss worldwide highlights the urgent need for technological solutions to help bridge the communication gap between hearing-impaired individuals and the rest of society. Machine learning is a branch of artificial intelligence that deals with the methods that let computers extract meaning from data and create AI applications. In the meanwhile, deep learning is a subset of machine learning that enables computers to resolve increasingly challenging issues . As deep learning develops transferable answers, it is more powerful than traditional machine learning. Through neural networks, or layers of neurons/units, deep learning algorithms are able to produce transferable solutions [2].

MOTIVATION

The ability to interact with technology with ease is becoming more important than ever in a society that is growing more and more digital. Machine learning-powered gesture-based communication systems allow machines to recognise and react to natural human gestures, revolutionising human-computer connection. This invention allows for more inclusive and user-friendly interfaces, which is especially helpful in improving accessibility for people with impairments. Furthermore, by offering simple and smooth control mechanisms, it improves user experience across a range of industries, including gaming, virtual reality, and smart homes. Machine

learning in gesture recognition holds the potential to enhance precision and flexibility, rendering these systems more effective and accommodating to diverse user requirements and surroundings. With this initiative, we hope to expand the possibilities for Hand Gesture Recognition using ML

DATASET

In this work, we have utilized the ASL dataset consisting of 51 classes, with approximately 4000 images per class. The classes comprise the alphabet, numbers, and commonly used words such as 'Hello', 'Help', and 'Stop'. The alphabet class enables the formation of new words through fingerspelling, where individual letters are used to represent words without a designated sign symbol. A Python script was employed to efficiently convert the image class folders into a .csv file, which stores the (x, y, z) coordinates of all landmark points of each sign with their respective outputs. An 80:20 train-test split was implemented to improve the model's feature extraction process[3].

Fig. 0 Various Sign Symbols.

MEDIAPIPE

Gesture recognition has been studied extensively utilizing traditional techniques such as body component tracking, different colour glove-based tracking, Kinect depth sensor tracking, and skeleton tracking. Multiple methods have been used to solve this problem like modified CNN, image segmentation, SVM and deep learning. Many machine learning algorithms have been developed for hand gesture recognition so as to create AI-based applications. Out of them, MediaPipe can be used for hand gesture recognition. Google supported MediaPipe framework can be used for solving several problems like face-recognition, face-map, eye, hand, poseestimator, holistic, hair, object-detection, box tracking and KIFT. With the help of the MediaPipe framework, we can develop an algorithm or model for the application, then help the application by providing results that can be cloned across different platforms. The MediaPipe framework is composed of three major components: (1) performance evaluation, (2) a mechanism for collecting data from the sensor (3) an assembly of reusable parts. A graph consisting of all the parts called the calculators is known as pipeline, wherein every calculator is inter-connected by channels through which the data flows. Developers can create their required application by removing or delineating user defined calculators anywhere in the graph. This result of calculators and channels creates a data-flow diagram. Hand gesture recognition with the MediaPipe framework is a dependable and high-fidelity hand and fingertracking system. Mediapipe hands uses an integrated ML pipe of several models working together.

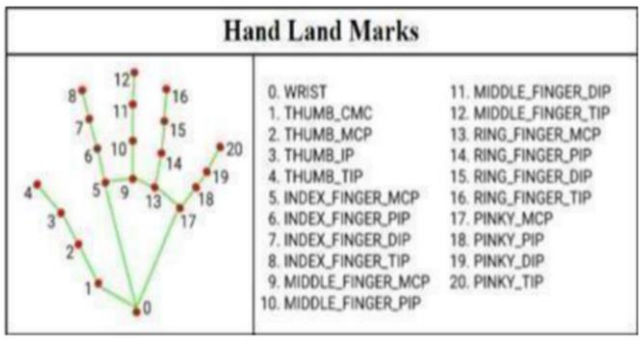


Fig. 1 Hand Landmark [Source: Ref [21]]

Hand-knuckles of the landmark have x, y, and z coordinates where x and y are normalized to [0, 1] as width and height of the image, while z represents the depth of the landmark. The closer the landmark to the camera, the value of z becomes smaller[3].

OPEN CV

Open cv computer vision library which is dominated for image processing it also called open source computer vision library. Mainly used for real time computer vision and image processing. The main library has design in c ++. Open CV open source library started in 2011[2].

EXPERIMENTATION

In order to achieve our desired objective, we have created an end-to-end web application that allows real-time communication between common people and deaf people without any use of hardware technologies like sensors, microcontrollers, etc. This website makes user interaction comfortable as it consists of combined application of Sign to Text and Text to Sign conversion, along with other essential features. To create this application, we have made use of multiple technologies and frameworks. HTML, CSS and JavaScript tools are used for Frontend and Flask (a Python web framework) is used for Backend. In Backend, the machine learning model is loaded in the form of a pickle (.pkl) file. This .pkl file allows easy serialization and deserialization of any ML model. The functionality of our website is that it takes the webcam video as the input which captures our hand image. Later, Mediapipe technique is applied to this extracted image and key points are marked accordingly which then stores the (x, y, z) coordinates of the landmarks. Last but not least, this data is sent into the Support Vector Machine classifier, a supervised machine learning classifier (SVM). Regression and classification studies both use the SVM model. Finding the most important dividing line is done using it. The primary objective of this approach is to identify the best hyperplane for dividing and separating training vectors. Using gamma as the RBF parameter, SVC is an SVM classifier (Radial basis function kernel). To determine if a model is overfitting, underfitting, or providing the optimum fit, one uses the gamma value. The pickle (.pkl) module was utilized to load the two files, X and y, which are data files used for training the SVM model. The X file contains a list of image pixels, while the Y file contains labels for the list of pixels. After loading the dataset, it is passed to the model for training purposes.

The SVM model is represented by the equation:

f(x) = sign(sum(alpha_i * y_i * exp(-gamma * ||x_i - x||^2)) + b).....(1)

where alpha_i are the Lagrange multipliers, y_i are the corresponding labels, exp(-gamma * ||x_i - x||^2) is the RBF kernel function, x represents the input feature vector, and b is the bias term. The hyperparameters C and gamma are typically determined through a grid search or cross-validation process. Once the model is trained, the webcam images are passed to the model for testing. The model recognizes the corresponding letter/word and outputs it on the screen.

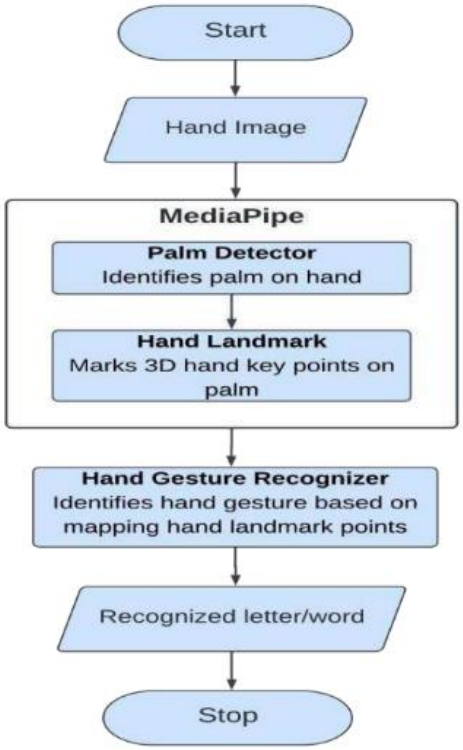


Fig. 2 Methodology

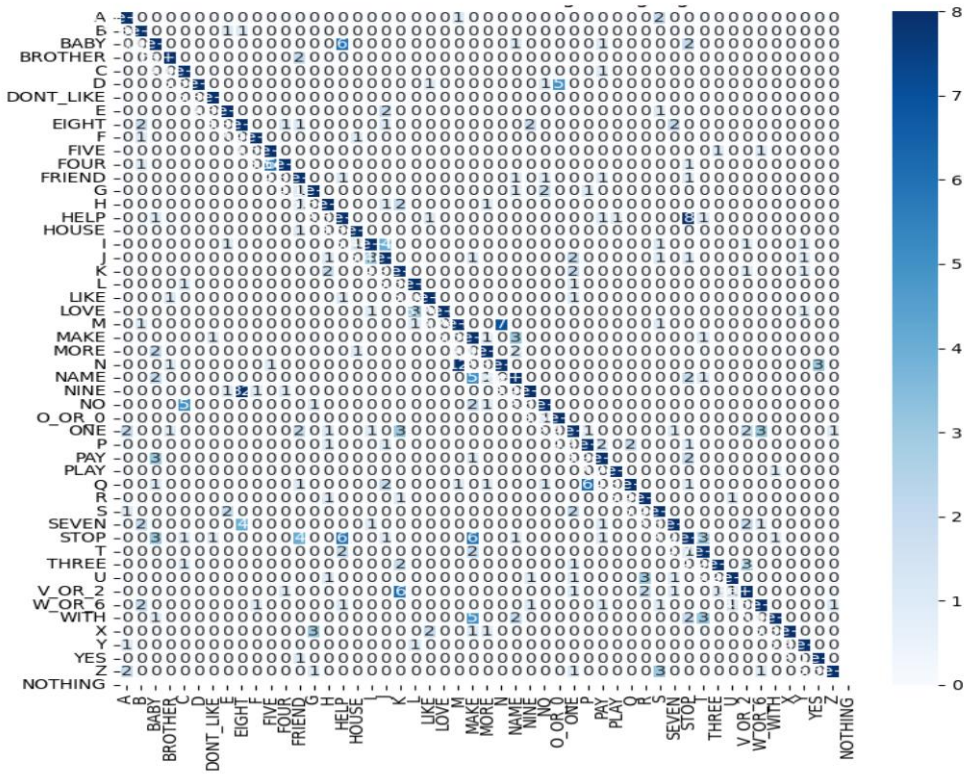
RESULTS

Various machine learning models are used for sign detection. These models are evaluated based on parameters like accuracy, recall, F1 score, etc. Among the utilized models, it is observed that SVM outperformed other machine learning techniques such as Naive Bayes, KNN, Decision Tree, etc. by achieving an accuracy of 98.65% (training) and 98.35% (testing) as shown in table 0. The reason it outperformed is because of its effectiveness in high- dimensional spaces where it draws a hyperplane boundary in order to classify the labels. It is also computationally less extensive and works well for image analysis tasks. The below table shows the values of training and testing accuracy along with Recall, F1Score and Precision for different tried models:

Table 0 Results for various ML algorithms

Model	Train (%)	Test (%)	Recall	F1 Score	Precision
SVM	**99.70**	**98.975**	0.98	0.98	0.98
Random Forest	99.89	97.50	0.97	0.97	0.97
Decision Tree	99.89	91.52	0.91	0.91	0.91
Naïve Bayes	50.63	50.84	0.50	0.50	0.50
KNN	97.62	96.39	0.96	0.96	0.96

The below confusion matrix for SVM algorithm prints the correct and incorrect values in number count which gives us a good data visualization.



Our machine learning approach is suitable for use in mobile applications since the learned model is deliberately light. Our methodology's real-time sign language identification makes it quick, reliable, and especially flexible for smart devices. Mediapipe makes feature extraction simple by deconstructing and analyzing challenging hand-tracking data. This strategy uses less computer resources and takes less time to train the model than other cutting-edge approaches.

Table 1 Comparative Analysis of Accuracy for Various Models and Preprocessing Techniques with Simpler Datasets

Preprocessing & Algorithm	Training Accuracy	Validation/Test Accuracy
Convex Hull + CNN	99.54%	91%
Gaussian Blur + CNN	89.8%	91%
Gaussian Blur + VGG	84.66%	84.92%
Canny Edge Detection + VGG	93.71%	93.69%
Convex Hull + ResNet	94.03%	91.98%
Convex Hull + EfficientNet	90.68%	90%

This table compares the effectiveness of different preprocessing techniques and algorithms on our dataset and simpler datasets. It includes the preprocessing technique and algorithm used, along with the training accuracy and validation/test accuracy achieved by each technique. As shown in table 1, several preprocessing techniques were tested, including convex hull, Gaussian blur, and Canny edge detection. The algorithms used included CNN, VGG, ResNet, and EfficientNet. After analyzing the results from the experiments using the techniques and algorithms presented in the above table, we found that they did not yield satisfactory performance on our dataset. Therefore, we decided to discard these techniques and algorithms and explore other approaches to achieve better results.

Table 2 Performance Comparison with Similar Techniques

Type of Dataset	Our Accuracy	Existing/Others Accuracy
Alphabets only	**99.43%**	99.15% **[7]**
Alphabets, Numbers, and Words	**98.975%**	98.62% **[7]**

Based on our analysis, we could improve the accuracy of the model by adjusting the parameters. The improvement in accuracy was found to be around 0.28% to 0.35%. Our experiments also revealed that the model tends to overfit at higher values of C, and the choice boundary's curvature weight decreases with lower values of gamma. As a result, the areas separating different classes become more generic. After tuning the parameters, we were able to identify the optimal decision boundary for our dataset at C = 52 and gamma = 0.6[3].

CONCLUSION

Individuals with hearing disabilities often face significant challenges in communicating with people who can hear. One of the most effective ways for them to communicate is through sign language. However, for people who do not know sign language, understanding what is being communicated can be a significant challenge. This communication gap can have a detrimental impact on the social and emotional well-being of individuals with hearing disabilities, making it difficult for them to engage fully in society.

The proposed Sign Language Recognition system offers an innovative solution to the communication gap between individuals with hearing disabilities and those who can hear. The proposed system successfully recognizes sign language with high accuracy, with an SVM model achieving a classification accuracy of 98.975%. Moreover, the use of Google's MediaPipe palm detector method has made the system accessible to people without any special hardware, which is a significant advantage.

Although there are still some research gaps that need to be addressed, such as improving the system's accuracy in recognizing signs for complex phrases and developing a portable and affordable device for practical use in daily life, the proposed Sign Language Recognition system offers a promising step towards creating a more inclusive society. With further development and refinement, this system can play a significant role in breaking down communication barriers and facilitating greater accessibility and understanding for individuals with hearing disabilities.

REFERENCES:

[1] https://www.ijnrd.org/papers/IJNRD2212134.pdf

[2] https://ijcrt.org/papers/IJCRT2401532.pdf

[3] https://assets-eu.researchsquare.com/files/rs-3106646/v1_covered_62b2336e-089a-467b-bf60-408fbe87b374.pdf?c=1689008377

[4] 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[5] Akoum, Alhussain, and Nour Al Mawla. "Hand gesture recognition approach for ASL language using hand extraction algorithm." Journal of Software Engineering and Applications 8.08 (2015): 419

[6] Bandgar, Bapurao. "Implementation of Image Processing Tools for Real-Time Applications." International Journal of Engineering Research & Technology (IJERT) 10.07 (2021).
[7] Bazarevsky, Valentin, and G R Fan Zhang. "On-Device MediaPipe for Real-Time Hand Tracking." (2019).
[8] Brahmankar, Vipul, et al. "Indian Sign Language Recognition Using Canny Edge Detection." International Journal 10.3 (2021).
[9] Das, P., T. Ahmed, and M. F. Ali. "Static Hand Gesture Recognition for American Sign Language Using Deep Convolutional Neural Network." 2020 IEEE Region 10 Symposium (TENSYMP) (2020): 1762-1765
[10] Devineau, G., F. Moutarde, W. Xi, and J. Yang. "Deep Learning for Hand Gesture Recognition on Skeletal Data." 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (2018): 106-113.
[11] Halder, Arpita, and Akshit Tayade. "Real-Time Vernacular Sign Language Recognition Using MediaPipe and Machine Learning." International Journal of Recent Technology and Engineering (IJRTE) 10.2 (2021): 7421.
[12] Islam, M. M., S. Siddiqua, and J. Afnan. "Real-Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language." 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (2017): 1-6.