Journal of Engineering Technology and Sciences
Volume 01, Issue 02, October 2024
www.jets.org.in

Machine Learning-Based Classification and Recognition of Indian Sign Language

T.Hima Bindu[1] | K.Swati Priya[2]
[1,2]KL University, Department of CSE, Vijayawada, Andhra Pradesh, India
Corresponding Author: bindubtech99@gmail.com

To Cite this Article
T.Hima Bindu and K.Swati Priya, "Machine Learning-Based Classification and Recognition of Indian Sign Language", Journal of Engineering Technology and Sciences, Vol. 01, Issue 02, October 2024, pp:10-19

Abstract *The discourse surrounding communication barriers for deaf individuals is increasingly recognized as a significant issue. People with hearing impairments often utilize various strategies to interact with others, necessitating specific resources to facilitate engagement. Developing a sign language application would greatly benefit deaf users and enhance communication with those unfamiliar with sign language. Our initiative aims to bridge the gap between hearing individuals and those who communicate through signs. The primary objective of this study is to create a perception-based paradigm to distinguish gestures visually. Vision-based systems are particularly effective as they provide a more intuitive means of interaction between humans and computers. This research focuses on 46 different gestures and incorporates both temporal and spatial dimensions from video sequences to classify sign language movements.*

Keywords*: CNN Classification, KNN Classification, Attribute Extraction, and Indian Sign Language*

I. Introduction

Gesture recognition encompasses movements from various parts of the body, including the hands and even facial expressions like those from the ears. In this context, we employ image detection and computer vision techniques for effective gesture recognition. This technology enables computers to interpret human behaviour intuitively, allowing for natural interactions without the need for mechanical interfaces. In a society where traditional communication may falter, sign language becomes a vital means of expression. Individuals who are deaf or hard of hearing often rely on sign language not only as a form of communication but also as a bridge to a world of sound they cannot access. While sign language facilitates interaction, it is also a cultural expression, rich in meaning and context. Sign language varies globally, with forms such as Indian Sign Language (ISL) and American Sign Language (ASL) representing localized adaptations. These languages can be categorized into discrete signs—where each sign corresponds to a single word—and continuous signs, where sequences of movements convey more complex ideas. In this research, we focus on techniques for detecting ASL motions, utilizing various methods to accurately recognize gestures and improve communication for those who use sign language.

Sign Language: Around the world, the deaf and hard of hearing communities use a visual language that relies on hand, facial, and body expressions to communicate, rather than spoken words. While the phrasing of gestures is not universally standardized, distinct sign languages exist, reflecting the diversity of speakers across different nations. For instance, countries like Belgium, Britain, the USA, and India each have their own variations of sign language, and there are hundreds of sign languages globally, including Japanese, British, and Spanish Sign Languages.

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.

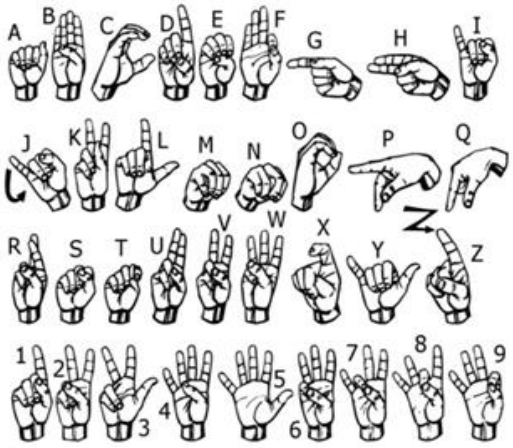


Figure 1: American Sign Language Finger Spelling

II. Literature Survey

In recent years, extensive research has focused on interpreting hand sign language, leading to significant advancements in gesture recognition technology. The following technologies are commonly employed for effective gesture recognition:

Computer Vision: Utilizing cameras and image processing techniques, computer vision systems can analyze hand movements and recognize gestures in real time.

Machine Learning: Algorithms, particularly those based on deep learning, are trained on large datasets of sign language gestures. This enables the models to learn patterns and improve accuracy in gesture classification.

Convolutional Neural Networks (CNNs): CNNs are particularly effective for image-based data, automatically extracting features from video frames to identify gestures.

Recurrent Neural Networks (RNNs): RNNs, especially when combined with CNNs, are used to capture the temporal aspects of gestures, enabling the recognition of sequences of movements.

Sensor Technologies: Wearable devices equipped with sensors can capture hand and body movements with high precision, providing alternative methods for gesture recognition.

Augmented Reality (AR): AR applications can visualize sign language gestures in real time, enhancing learning and communication by overlaying digital information onto the physical world. These technologies are paving the way for more intuitive and accessible communication tools for the deaf and hard of hearing communities, fostering greater inclusivity.

Enhancement of the view: The camera serves as a crucial input device for hand and finger monitoring in vision-based techniques. These vision-focused approaches require only a monitor, enabling seamless interaction between users and devices without the need for additional equipment. The aim of these programs is to enhance biological perception by integrating artificial vision systems within software and/or hardware.
However, achieving effective performance poses several challenges. The recognition processes must be invariant and contextually relevant, as well as independent of both human characteristics and camera variations. To meet these demands, it is essential to develop systems that prioritize consistency and robustness.

Figure 2: Vision-based recognition relies on how individuals interpret visual information about their surroundings. Despite being one of the most complex methods, it offers a natural and intuitive way to interact with technology.

Gesture Recognition Techniques in Sign Language
3D Hand Model Construction: The first step involves creating a three-dimensional representation of the human hand. This model is aligned using images from one or two cameras, allowing for the measurement of joint parameters. These features are then utilized to classify gestures effectively.
Image Capture and Feature Extraction: The initial image is captured by a camera, from which specific characteristics are extracted to serve as input for the classification algorithm.
Argentine Sign Language Recognition: A notable study presents a handshake approach for learning Argentine Sign Language (LSA).
Two significant contributions arise from this work: Hand Database Creation: A comprehensive database for Indian Sign Language (LSA) was established.
ProbSom Methodology: This approach involves estimating, extracting descriptors, and classifying text through manually adapted self-organizing maps known as ProbSom. The ProbSom neural description achieves a classification precision exceeding 90%, outperforming other advanced techniques like Support Vector Machines (SVMs), Random Forests, and Neural Networks.
Automatic Recognition of Indian Sign Language
In another study focused on continuous Indian Sign Language, an architecture comprising four main modules was proposed: Data Gathering: Capturing video data. Pre-Processing: Implementing skin filters and histogram matching. Feature Extraction: Utilizing auto-vector-driven mining attributes and Euclidean-weighted auto classification techniques.
Classification: The system achieved a recognition score of 96% for 24 different alphabets.
Sentence Comprehension and Teaching in Indian Sign Language
The interpretation of continuous sign language poses significant challenges. This research employs a frame extraction approach centered on gradients to address this issue. By segmenting continuous gestures into distinct signs, it overcomes the lack of informative structures. Each gesture is treated as an individual action, with preparation functions acquired through Orientation Histogram (OH) techniques to enhance functionality. Experiments conducted at the

Robot and Artificial Intelligence Laboratory (IIIT-00A) utilized their own ISL dataset with a Canon EOS camera, employing various categorization methods for analysis.

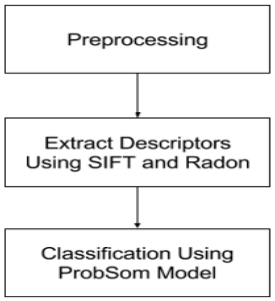


Fig 3: Manual Recognition Device Block Diagram for LSA

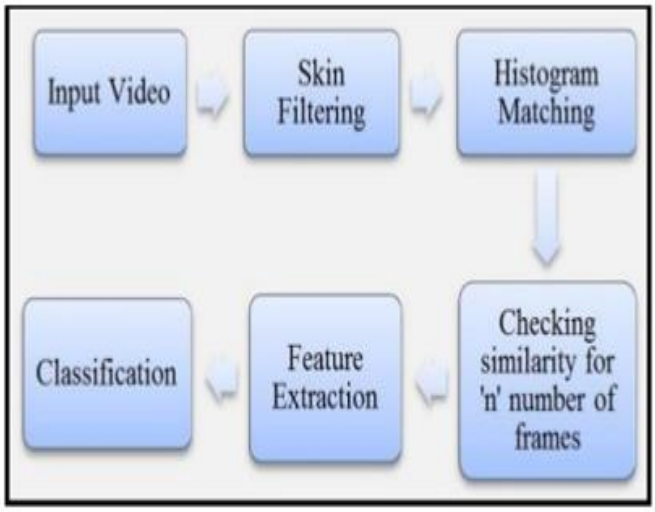


Figure 4: System Overview [2]

In the realm of gesture recognition and classification, various distance metrics such as Euclidean distance, Manhattan distance (also known as city block distance), and others have been employed. A comparative study of these distance classifiers has yielded interesting insights. The results of this study indicate that both the connection-based and Euclidean distance methods demonstrate superior accuracy compared to other classification techniques. This suggests that leveraging these distance metrics can enhance the effectiveness of gesture recognition systems, leading to improved classification outcomes

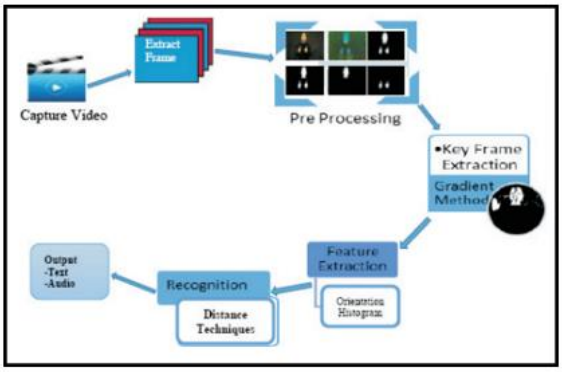


Figure 5: General Diagram of the Work [3]

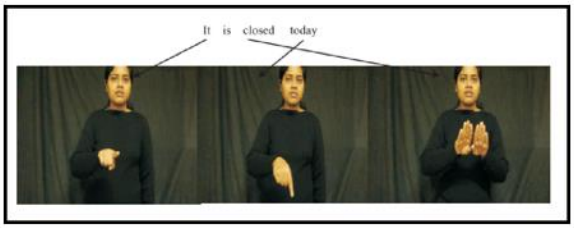


Figure 6: Gesture of Sentence It is Closed Today [3]
The isolated Indian Sign Language Manual is understood in real time [4]

This study presents statistical methods for real-time identification of Indian Sign Language (ISL) expressions, including gestures like "paws." The authors developed and utilized a multi-image video database containing various indicators to enhance the recognition process. To achieve robustness against variations in lighting and direction, the study employs the Path Histogram as the grouping function. Furthermore, the use of Euclidean distance and neighbor metrics highlights two distinct approaches for measuring similarity and classification within the dataset.

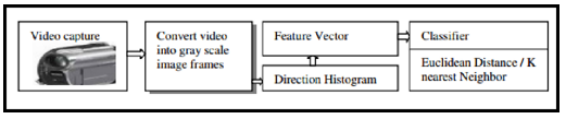


Figure 7: Methodology for real time ISL classification [4]

III. Time and Space Techniques for Gesture Recognition

The concept was developed using two distinct techniques that address both temporal and spatial characteristics. notably, the inputs for the recurrent neural network (rnn) differ from those used in other temporal analysis methods. Both techniques utilized the argentinean sign language dataset, which comprises approximately

2,300 views of 46 different gesture categories. each gesture was recorded in five repetitions, resulting in a total of 50 videos per gesture. the recordings were made by 10 non-expert participants, ensuring a diverse representation of the signs.

Id	Name	Id	Name	Id	Name	Id	Name
1	Son	13	Enemy	25	Country	37	To-Land
2	Food	14	Dance	26	Red	38	Yellow
3	Trap	15	Green	27	Call	39	Give
4	Accept	16	Coin	28	Run	40	Away
5	Opaque	17	Where	29	Bitter	41	Copy
6	Water	18	Breakfast	30	Map	42	Skimmer
7	Colors	19	Catch	31	Milk	43	Sweet-Milk
8	Perfume	20	Name	32	Uruguay	44	Chewing gum
9	Born	21	Yogurt	33	Barbeque	45	Photo
10	Help	22	Man	34	Spagheti	46	Thanks
11	None	23	Drawer	35	Patience		
12	Deaf	24	Bathe	36	Rice		

75%, i.e., 40 for planning and 25% for study is employed out of the 50 percent motions

Frames are extracted from multiple video sequences of each gesture. Background noise and irrelevant body parts are filtered out to focus solely on the gesture being performed. The extracted frames are used to train the CNN model, which captures spatial features. A deep neural network architecture is employed for this purpose. The trained model is applied to make predictions on both training and testing datasets. The predictions from the CNN are now utilized for training the RNN model, specifically using Long Short-Term Memory (LSTM) networks to capture temporal characteristics.

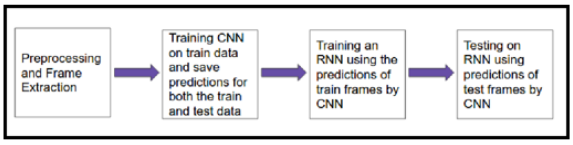


Figure 8: Estimates 23

Each video gesture is divided into multiple frames. These frames are processed to eliminate all background noise, leaving only the hands visible. The final output consists of a grayscale image of the hand, with colorful elements removed to focus solely on the gesture. This simplifies the model's learning process, allowing it to concentrate on the essential features of the hand movements.

Figure 9: One of the Extracted Frames

Figure 10: Frame after extracting hands (Background Removal)

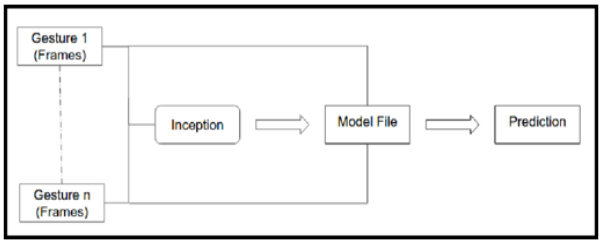


Figure 11: Train CNN (Spatial Features) and Prediction

First Row: An image depicting the "Elephant" motion.
Second Row: A collection of selected frames extracted from the video of the gesture.
Third Row: The CNN sequence of projections, representing the processed output for each frame after preparation.

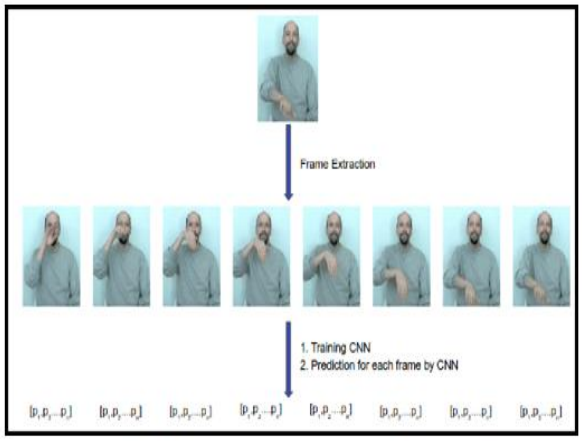


Figure 12 Training RNN (Temporal Features)

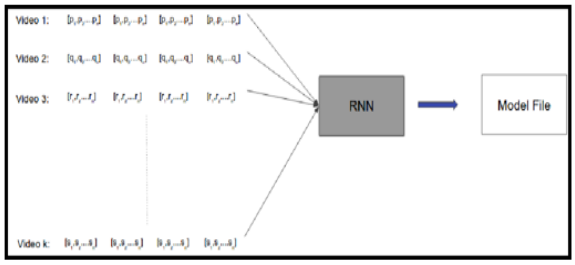
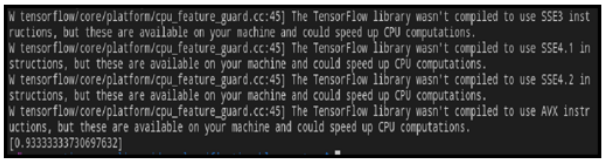


Figure 13

Limitations: The probabilistic projection period of the CNN is directly related to the number of classes identified in the frame sequences. With 46 classes, we have a corresponding 46 outputs. The length of the feature vector for each frame is determined by the number of classes, resulting in a vector length that is generally shorter than that of the class representation. In this method, we utilize the CNN approach to capture spatial features, providing the output from the pooling layer to the RNN prior to making predictions. The pooling layer generates a 2048-dimensional vector that encapsulates the spatial characteristics of the image but does not serve as a class predictor. The majority of the steps in this method are similar to the first approach, with the primary difference being the inputs to the RNN in both processes.

IV.Results

Result of Approach



The total accuracy for Strategy 2 is 95.217%, based on 460 actions (10 per category), with results evaluated for 438 actions. The detailed accuracy per category is summarized in the list below.

ID	Gesture	Accuracy	ID	Gesture	Accuracy
1	Name	100	24	Spaghetti	100
2	Yogurt	100	25	Patience	100
3	Accept	90	26	Deaf	90
4	Man	100	27	Eenemy	90
5	Drawer	100	28	Dance	90
6	Bathe	100	29	Rice	100
7	Opaque	90	30	To-Land	100
8	Country	100	31	Yellow	100
9	Water	90	32	Green	90
10	Red	100	33	Give	100
11	Call	100	34	Food	80
12	Colors	90	35	Away	100
13	Run	100	36	Copy	100
14	Bitter	100	37	Coin	90
15	Perfume	90	38	Where	90
16	Map	100	39	Skimmer	100
17	Born	90	40	Trap	80
18	Help	90	41	Sweet-Milk	100
19	Milk	100	42	Breakfast	90
20	None	90	43	Chewing-Gum	100
21	Urugway	100	44	Photo	100
22	Son	80	45	Thanks	100
23	Barbeque	100	46	Catch	90

The second approach demonstrated higher accuracy compared to the first. This improvement is attributed to the RNN receiving a 2048-dimensional output from the pooling layer in the second approach, as opposed to a 46-dimensional prediction sequence in the first approach. The richer feature representation in the second method enabled the RNN to better differentiate between various images, leading to enhanced performance.

V. Conclusion

Hand gestures play a crucial role in facilitating human-computer interaction across various applications. Visual methods for recognizing hand movements offer several advantages over traditional technologies. However, accurately recognizing these gestures remains a challenge. This study contributes to addressing this issue by presenting a visual system for interpreting Argentine Sign Language (LSA).

Videos cannot be classified based solely on either temporal or spatial characteristics; thus, we employed two distinct models to capture these features effectively. Spatial features are handled by Convolutional Neural Networks (CNNs), while temporal features are managed by Recurrent Neural Networks (RNNs). Our approach achieved an accuracy of 95.217%, demonstrating that CNNs and RNNs can effectively model both spatial and temporal characteristics in sign language gestures. We utilized two methods to tackle our challenges, with variations primarily in the RNN inputs discussed earlier. Moving forward, we aim to expand our research to encompass more consistent motion interpretations within sign language. This approach could also be adapted for vocabulary-level recognition. Integrating both models into a single platform presents an exciting focal point for future work.

References

[1] Srinivasa Varma and Sasidhar . "Communication in Indian Sign Languages" Science Procedia Algorithm Science 21 (1992):213-412
[2] Srikanth, Hema Bindu and Keshav Reddy "Recognition of Indian Sign Langunages using gestures" Case study in Indian Research Articles (2014):112-154
[3] Shing cham, Yuen Khan, Tole Sukio and Rehaman Abdulla "Learning and Understanding of Indian Sign Languages" Science Direct Vol 01, Issue no 3, pp 214-324.
[4] Hopper Singh, Geogry Pekk and Jeff Albert, "Long Term on Memory Loss" computation 2, no. 2 (1982):112-312
[5] Albert Robert, Shaik Abdullah, Shing Weing and Flourence "MZW12: A Sign Language Dataset for Argentina." (PMESA 2020)
[6] Kingsle, William Jeff, Vijay Rayhod and Jeff Geogry:"A Stochastic Optimisation Method." Preprint XXIV
[7] A novel study on Neural Networks for Indian Sign Languages." Case Study Article 16214(2012).