



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: <https://doi.org/10.22214/ijraset.2024.59239>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

ASL Spelling Detection Using Deep Learning

T. Kameswara Rao¹, Tanubuddi Jyothirmayee², Reniguntla Bharadwaj³, Prathipati Gnana Venkat⁴, Sri Ruthuhasa Komma⁵

Department of Computer Science & Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh

Abstract: *Understanding American Sign Language is a difficult task for anyone who wish to communicate with the deaf and dumb. It creates a gap between people. This paper introduces a finger spelling recognition system designed to help bridge that gap. Our system uses a camera and specialized CNN architecture to understand hand gestures that represent letters. This approach offers a few key benefits: it uses the familiar alphabet, can be learned quickly, and works for people with different levels of hearing loss or sign language experience. This technology has the potential to make communication much easier for many people, especially in situations where spoken words or traditional signs might not be the best option. It detects the ASL alphabets and constructs words and sentences, also known as Finger-Spelling.*

Keywords: *CNN, American Sign Language (ASL), Frame Extraction, DeepLearning, Mediapipe, OpenCV*

I. INTRODUCTION

The ability to communicate effectively is a fundamental human right. For individuals who are deaf or dumb, barriers to communication in aural environments can lead to social isolation and limited opportunities. American Sign Language (ASL) provides a robust solution, yet the time and effort required for its acquisition can present challenges. This paper presents a finger spelling recognition system designed to supplement existing communication strategies for dumb and hearing impaired individuals. By focusing on the recognition of individual letters through finger spelling, this system offers several potential benefits as below.

- 1) *Familiarity:* Finger spelling utilizes the commonly understood alphabet, reducing the initial learning curve compared to comprehensive sign language systems.
- 2) *Inclusivity:* This approach can be used by individuals with varying degrees of hearing loss and differing levels of ASL fluency.
- 3) *Adaptability:* The system has the potential to expand its recognition capabilities to include names, locations, and newly encountered terms, enhancing its practicality for diverse communication needs. The technical foundation of this finger spelling recognition system lies in a methodology based on Convolutional Neural Networks (CNNs). The following sections detail this methodology and its efficacy in classifying hand gestures associated with finger spelling.

II. LITERATURE REVIEW

A lot of work and research is being done on sign language to make communication easier among the people hard of hearing and general public. Authors A.K. Sahoo, G.S. Mishra, and R. Kiran^[5] discuss the significance of sign language in facilitating communication within the deaf and mute community and with others. They highlight the importance of computer recognition systems, which range from capturing sign gestures to generating text or speech. The authors note that sign gestures can be categorized as static or dynamic, with both types being essential for effective communication. The survey delves into the various steps involved in sign language recognition, covering aspects such as data acquisition, pre processing, feature extraction, classification, and resultant findings. Moreover, the authors suggest potential future research directions in this field.

In the comprehensive analysis, authors Adeyanju, M.A. Adegboye^[1] delve into the intricacies of automatic sign language recognition (ASLR), recognizing its pivotal role in closing communication gaps between hearing-impaired individuals and the broader community. Their study, spanning two decades of research, identifies the complexities posed by the vast diversity of sign languages worldwide. By leveraging bibliometric analysis tools, they explore the temporal and regional distributions of 649 relevant publications, shedding light on collaborative networks among institutions and researchers. Additionally, the paper offers a thorough review of vision-based ASLR techniques, emphasizing the importance of integrating intelligent solutions to enhance recognition accuracy. While acknowledging the ongoing challenges, the authors anticipate that their work will inspire further advancements in intelligent-based sign language recognition systems, paving the way for improved communication accessibility and inclusivity.

Authors Y. Obi, K.S. Claudio, V.M. Budiman^[8] propose a desktop application to address the challenge of communication with special needs individuals who use sign language, mode of communication that many people do not understand. Their research focuses on developing a system capable of real time sign language recognition and conversion to text.

Employing American Sign Language (ASL) datasets and Convolutional Neural Networks (CNN) for classification, the process involves filtering hand images before passing them through a classifier to predict gestures. The primary emphasis lies on achieving high accuracy in recognition, with the application yielding an impressive 96.3% accuracy for the 26 letters of the alphabet.

Authors R. S. Rani, R Rumana, and R. Prema^[2] address the significance of hand gestures in sign language as a vital form of non-verbal communication, particularly for individuals with hearing or speech impairments. They emphasize the limitations of existing sign language systems in terms of flexibility and cost effectiveness for end users. To address these challenges, the authors present a software prototype capable of automatically recognizing sign language, aiming to enhance communication effectiveness for deaf and mute individuals with both their community and the general population. They highlight the societal barriers faced by mute individuals, who often rely on interpreters or visual aids for communication, which may not always be readily available or easily understandable. Sign language emerges as the primary mode of communication within the deaf and mute community, emphasizing the importance of initiatives to facilitate better communication accessibility for all individuals.

Authors Mathavan, N.K. Kumar, and Angappan^[3] introduce a Hand Gesture Recognition System (HGRS) as an intuitive and efficient tool for human computer interaction. With applications ranging from virtual prototyping to medical training, the system focuses on recognizing Indian Sign Language (ISL) gestures to aid deaf and mute individuals. The proposed ISLR system employs a pattern recognition approach comprising feature extraction and classification modules. By leveraging Discrete Wavelet Transform (DWT)-based feature extraction and nearest neighbour classification, the system achieves exceptional accuracy, with experimental results demonstrating a maximum classification accuracy using the cosine distance classifier.

III. PROPOSED METHODOLOGY

A. Data Collection and Preprocessing

This research used a publicly available dataset^[11] of American Sign Language (ASL) alphabet images found on kaggle. The dataset had 87,000 images, each showing a single hand sign. The signs were organized into different categories: letters A-Z, plus 'SPACE', 'DELETE', and 'NOTHING'. Each image is 200*200 pixels in size and there are 29 classes each containing 3000 images. The dataset is approximately 1.03GB in size and is available in kaggle.

1) *Image Preprocessing* : Before using the images, we made them clearer for the computer to understand. We slightly blurred the images to get rid of tiny distractions. By Focusing on the hand, We made the hand shapes stand out from the background using a gaussian blur technique that automatically figured out the best way to make the image black and white as shown in below fig 1.

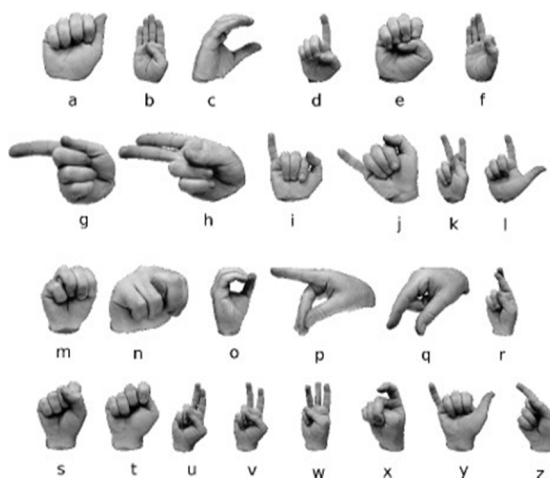


Fig. 1 Images of the ASL alphabet after focusing on the hand

B. Convolutional Neural Network (CNN) Architecture

After preprocessing the images, we developed a custom CNN model that correctly identifies the sign. The layers used in the CNN are explained below.

Network Structure: The CNN model has multiple layers:

- *Convolutional Layers (Feature Finders)*: These layers use tiny filters to search for the important parts of hand signs, like lines and curves.
- *Max Pooling Layers (Summarizers)*: These layers shrink the information down, making things simpler for the next step.
- *Dropout Layers (Prevention of Overthinking)*: These layers randomly stops parts of the model sometimes, which helps it learn more generally and not get stuck on tiny details.
- *Flatten Layers (Organizing the Info)*: This layer gets everything ready for the final decision-making part of the model.
- *Dense Layers (The Decision Makers)*: These layers take all the sign features, put everything together, and decide which letter the sign most likely is.
- *Rationale*: We chose a function powerful enough to recognize signs accurately, but still fast enough to work in real-time.

C. Real-time Hand Detection

- *Library*: We used a special tool called 'mediapipe' to find hands within videos. It's good at locating hand landmarks and works fast.
- *Detection Parameters*: We experimented with parameters that regulate how certain 'mediapipe' has to be before claiming it detected a hand. It's a balancing act between identifying all the hands and not becoming distracted by other things.

D. Gesture Recognition and User Interface

We applied Gesture Isolation by drawing a bounding box around the hand in the video, so our system only focuses there. The Prediction Module takes the picture of the hand inside the box that gets fed into our brain (the CNN) for classification.

We created a GUI Framework that is a kind of control panel using a tool called Tkinter, so it's easy to interact with our system.

Our control panel shows the video with the box around the hand. The letter our system thinks you're signing, and how confident it is. The word being formed, letter by letter. A spell-checked version of the word to help catch mistakes.

Custom Word Construction: Individual characters predicted by the CNN were combined sequentially using a custom word construction algorithm. This algorithm handled special characters such as 'space' to delineate words and 'del' to enable backspace functionality. Sentence formation was achieved by concatenating completed words with spaces. To improve the communicative quality of the constructed text, an optional spellchecking module Py spellchecker was integrated to correct potential errors in real-time.

IV. IMPLEMENTATION

A. Data Loading and Preprocessing

Loaded the ASL-Alphabet dataset using *pandas*. Implemented pre-processing techniques like hand segmentation using techniques like color thresholding and background subtraction libraries. Normalize the images to a consistent size and pixel values. Converted the images and labels into a suitable format to being fed to the deep learning framework

B. Training the Model

Data Splitting: The pre-processed data will be split into training, testing and validation sets. The training set is used to train the model that detects ASL signs, the validation set helps monitor the learning process and prevent over fitting, and the testing set evaluates the model's performance of identifying ASL signs on new or unseen data. Common splitting ratios used are 80% for training, 10% for validation, and 10% for testing.

We defined the training process using the chosen deep learning library's functionalities. This includes specifying the optimizer i.e., Adam, the loss function i.e., categorical cross-entropy for multi-class classification, and the number of training epochs (iterations). We will implement a training loop that iterates through the training data in batches.

Within each iteration, the model processes a batch of images, computes the loss based on the predicted and actual labels, and updates its internal parameters through the chosen optimizer to minimize the loss. This process continues for 10 epochs. The model summary can be seen in the below fig 2.

Model: "sequential_3"

Layer (type)	Output Shape	Param #
conv2d_15 (Conv2D)	(None, 62, 62, 128)	1280
conv2d_16 (Conv2D)	(None, 60, 60, 256)	295168
max_pooling2d_12 (MaxPooling2D)	(None, 30, 30, 256)	0
conv2d_17 (Conv2D)	(None, 28, 28, 256)	590080
max_pooling2d_13 (MaxPooling2D)	(None, 14, 14, 256)	0
conv2d_18 (Conv2D)	(None, 12, 12, 512)	1180160
max_pooling2d_14 (MaxPooling2D)	(None, 6, 6, 512)	0
dropout_9 (Dropout)	(None, 6, 6, 512)	0
conv2d_19 (Conv2D)	(None, 4, 4, 512)	2359808
max_pooling2d_15 (MaxPooling2D)	(None, 2, 2, 512)	0
dropout_10 (Dropout)	(None, 2, 2, 512)	0
flatten_3 (Flatten)	(None, 2048)	0
dropout_11 (Dropout)	(None, 2048)	0
dense_6 (Dense)	(None, 1024)	2098176
dense_7 (Dense)	(None, 29)	29725

=====
 Total params: 6554397 (25.00 MB)
 Trainable params: 6554397 (25.00 MB)
 Non-trainable params: 0 (0.00 Byte)

Fig 2 Model summary

C. Results and Discussions

Our proposed Convolutional Neural Network (CNN) model achieved a remarkable accuracy of 98% in recognizing signs from the ASL-Alphabet dataset. This exceptional performance highlights the model's strength to effectively extract the crucial visual features associated with each sign and translate them into corresponding text characters with a high degree of precision.

The CNN architecture successfully learned to identify the distinctive visual patterns that define individual signs in the dataset. The convolutional layers effectively captured the spatial relationships between pixels within the images, allowing the model to differentiate between similar-looking signs.

1) *Efficient Learning*: The training process equipped the model to learn well from the training data. The model can not only recognize the signs it was trained on but also perform well on unseen examples with a high degree of accuracy. This is particularly important for real-world applications where users might employ various signing styles or encounter variations in lighting conditions.

2) *Potential for Real-World:* An accuracy of 98% suggests that the model has the capability to be an useful tool for sign language translation. It can potentially bridge the communication gap between deaf and mute communities by facilitating accurate and efficient sign-to-text translation.

The output of our application can be seen from the below fig 3 and fig 4. The additional incorporation of SPACE, DELETE, NOTHING has improved the sentence construction which is easier for the people.



Fig 3 output of the application

We maintained a threshold of 0.95 as seen from the images fig 3, to make sure that the confidence of the predicted sign is greater than the threshold for it to be considered into words or sentence construction.

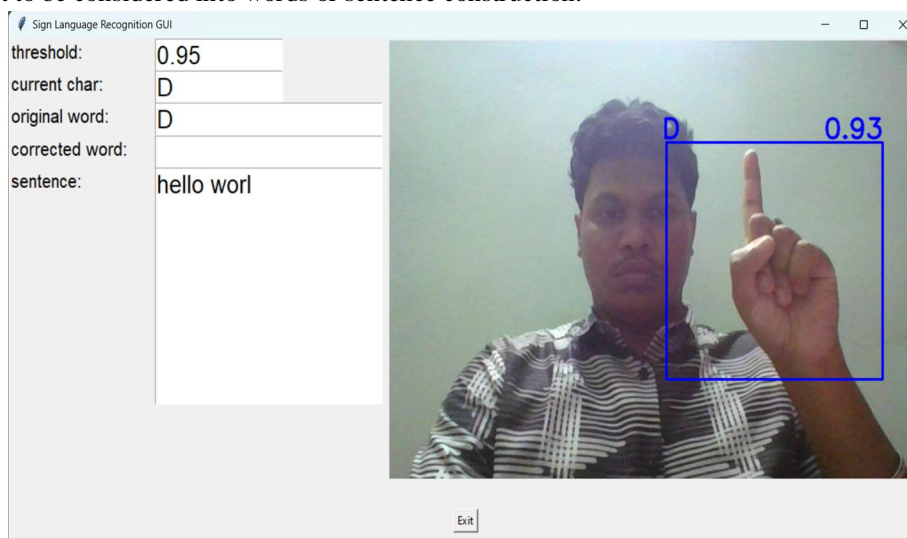


Fig 4 sentence construction

The ASL-Alphabet dataset focuses on alphabets, and real-world scenarios involve a broader range of signs encompassing numerals, punctuation, and grammatical elements. Expanding the training data with more comprehensive sign language datasets can potentially improve accuracy for a wider range of communication.

Variations in Signing Styles: Signing styles can vary between individuals and regions. While achieving high accuracy on a specific dataset is promising, the model's performance might need further evaluation on datasets encompassing diverse signing styles to ensure generalizability.

Overall, the 98% accuracy as shown in the fig 5 achieved by our CNN model demonstrates its remarkable effectiveness in sign language recognition. This accomplishment paves the way for further development and refinement, leading to a robust and practical sign language translation system for real-world applications.

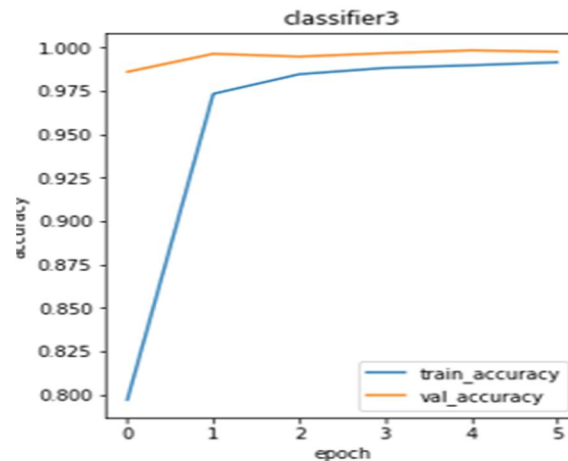


Fig 5 Model's Accuracy

V. CONCLUSIONS

This work demonstrates promising steps towards real-time automated recognition of American Sign Language (ASL) fingerspelling from images. A modular convolutional-transformer neural network architecture is shown to effectively classify isolated fingerspelled letters from spatial landmark coordinates. The model achieves 98% accuracy on real-world test data, improving on prior published results. The architecture combines convolutional feature extraction, transformer self-attention, and connectionist temporal classification loss to handle the subtle visual cues differentiating hand shapes. This could enable new assistive technologies for deaf communication and accessibility. While results are encouraging, there remains many opportunities for future work. Confusions between highly similar hand shapes need to be addressed, potentially via ensembles or multi-task training. Explicit sequence modelling would also help capture co-articulation effects in fluent signing. Training data diversity remains a challenge, requiring expanded corpora across signers, environments, and vocabulary. End-to-end integration from raw video rather than pre-processed landmarks would also enhance applicability. Nonetheless, this work helps advance sign language recognition toward real-world utility. The techniques presented help push computer vision and sequence modelling capabilities for this important application domain. By bridging communication gaps, assistive recognition technologies can help expand accessibility and equality for the deaf community. This work aims to provide a valuable baseline as research progresses in better understanding the complexities of fluid and natural sign language communication.

VI. LIMITATIONS AND FUTURE SCOPE

A. Limitations

- 1) *Confusing Background Noise:* The "nothing" symbol might be misinterpreted as background noise or hand movement between signs, potentially leading to missed translations.
- 2) *Misinterpretation of Pauses:* The "space" symbol could be confused with natural pauses that occur during signing, resulting in unnecessary insertions of spaces in the translated text.
- 3) *Delete Symbol Ambiguity:* The "delete" symbol's function can be ambiguous. It could signify backspacing during text input or indicate an error in signing that needs correction. Without additional context, the model might struggle to accurately interpret its meaning.

B. Future Work

- 1) *Expanding the Dataset Scope:* Move beyond alphabets and incorporate datasets encompassing a wider range of signs, including numerals, punctuation marks, and grammatical elements used in sign language. This will broaden the model's ability to handle more complex communication scenarios.
- 2) *Incorporating Facial Expressions and Body Language:* Sign language relies not just on hand gestures but also on facial expressions and body language. Explore techniques for integrating these additional modalities into the model, potentially using multi-modal architectures that process visual data from both hands and the face. This can enhance the model's understanding of the emotional context and intent conveyed during signing.

- 3) *Transitioning from Images to Videos*: While the current model utilizes static images, we can explore incorporating video data for training. This allows the model to capture the temporal dynamics of signs, where the sequence of hand movements plays a crucial role in conveying meaning. Recurrent Neural Networks (RNNs) can be particularly valuable in this context, as they excel at analysing sequential data.
- 4) *Real-World Implementation and User Interface Design*: Developing a user-friendly interface for the sign language translation system. This could involve a mobile application or webcam integration that allows users to interact with the system in real-time, facilitating seamless communication between deaf and hearing individuals.
- 5) *Generalizability and Adaptability*: Explore techniques for improving the model's generalizability to handle variations in signing styles across regions and communities. This might involve incorporating techniques like domain adaptation or lifelong learning algorithms that allow the model to continuously learn and adapt to new signing styles.

REFERENCES

- [1] I.A. Adeyanju, O.O. Bello, and M.A. Adegboye "Machine learning methods for sign language recognition," vol:12, November 2021
- [2] Reddygari Sandhya Rani , R Rumana , R. Prema "A Review Paper on Sign Language Recognition for The Deaf and Dumb", DOI : 10.17577/IJERTV10IS100129,2021.
- [3] Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, "An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform", DOI: 10.4236/cs.2016.78162 ,vol:7, 2016.
- [4] J.-H. Sun, T. -T. Ji, S. -B. Zhang, J. - K. Yang and G. -R. Ji, "Research on the Hand Gesture Recognition Based on Deep Learning," 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Hangzhou, China, 2018, pp. 1-4, doi: 10.1109/ISAPE.2018.8634348.
- [5] Ashok Kumar Sahoo, Gouri Sankar Mishra, and Kiran Kumar Ravulakollu " Sign language recognition: State of the art," vol:9 , 2014.
- [6] T.Mohamed et al., "Intelligent Hand Gesture Recognition System Empowered With CNN," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-8, doi: 10.1109/ICCR56254.2022.9995760.
- [7] Arjun Bapusaheb Pawar, Gautam Ved, Muzammil Shaikh, Uday Kumar Singh".Deaf and Dumb Translation Using Hand Gestures," Volume 8, April 2019. Page: 13
- [8] Yulius Obi, Kent Samuel Claudio, Vetri Marvel Budiman, and Said Achmad " Vol - 216 DO - 10.1016/j.procs.2022.12.106, 2023.
- [9] H.Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language." 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.
- [10] Jayanthi P, Ponsy R K Sathia Bhamana and B Madhubalasri "Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and Prediction using LSTM," Vol. 82, DOI: 10.56042/jsir.v82i07.2375, July 2023.
- [11] Akash Nagaraj. (2018). ASL Alphabet [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/29550>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)