

محمد القثمي

AI Sign Language Recognition for the Deaf and Mute.pdf

 Midocean University

Document Details

Submission ID

trn:oid::2945:379068485

Submission Date

Apr 22, 2026, 6:30 PM GMT+2

Download Date

Apr 22, 2026, 7:06 PM GMT+2

File Name

AI Sign Language Recognition for the Deaf and Mute.pdf

File Size

1.8 MB

22 Pages

4,597 Words

30,042 Characters

67% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups



16 AI-generated only 60%

Likely AI-generated text from a large-language model.



6 AI-generated text that was AI-paraphrased 7%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





AI Sign Language Recognition for the Deaf and Mute

Prepared By

Name: Anas Abdulkader Alzain

ID No.: 453001724

Name: Mohammed Omar Alqathmi

ID No.: 253001437

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

Supervisor Name: DR. Hagar Salah

College of Informatics, Midocean University, Comoros

Email:.....

Acknowledgements

We would also like to express our appreciation to our great supervisor, Dr. Hajar Saleh, whose guidance, support, and sound advice have been invaluable in the preparation and implementation of this Master's thesis in Artificial Intelligence.

She is highly experienced, has a vision informed by constructive criticism, and this has played a significant role in ensuring that this research is of high quality and scientific rigor. Her grades and her enthusiasm for improving our intellectual side have been motivating and inspiring. As it is now, this thesis could not have been composed without the grace of God, her wise management, and her invaluable directions.

There is much that we owe her for her patience, support, and constant encouragement throughout this academic life.

We would like to thank her a thousand times for her time, her efforts, and her trust in us.

Abstract

Artificial intelligence (AI) is a major part of technology today. One of its most useful uses is computer vision. It helps computers understand and analyze visual information, like hand gestures. Convolutional Neural Networks (CNNs), a type of deep learning model, have made gesture recognition systems much more accurate.

Sign language is the main way Deaf and Hard-of-Hearing people communicate, but hearing people don't always understand it, which makes it hard for them to communicate with each other. Older systems used big, heavy hardware that was hard to use. Newer systems use regular cameras, which make things easier, but there are still some problems.

This study aims to develop a real-time American Sign Language (ASL) recognition system that converts hand gestures into readable text. The proposed system is expected.

Table of Contents

List of Abbreviations	5
List of Figures	6
List of Tables	7
Chapter 1: Introduction and Overview	8
1.1 Introduction	8
1.2 Problem Definition.....	8
1.3 Thesis Objectives.....	9
1.4 Thesis Contribution.....	9
1.5 Thesis Organization.....	9
Chapter 2: Background and Related Work	10
2.1 Introduction	10
2.1.1 Artificial Intelligence (AI).....	10
2.1.2 Computer Vision (CV).....	10
2.1.3 Deep Learning (DL).....	11
2.1.4 Convolutional Neural Networks (CNN)	11
2.1.5 Natural Language Processing (NLP)	13
2.1.6 Explainable AI (XAI)	13
2.1.7 Image Preprocessing Techniques.....	13
2.1.8 Linguistic Post-Processing	13
2.1.9 Evaluation Criteria.....	13
2.1.10 Accuracy.....	14
2.2 Related Work	15
Review of Related Work	15
References	19

List of Abbreviations

- **Acc.:** Accuracy
- **AI:** Artificial Intelligence
- **ASL:** American Sign Language
- **CNN:** Convolutional Neural Network
- **CV:** Computer Vision
- **D&M:** Deaf and Mute
- **DL:** Deep Learning
- **GPU:** Graphics Processing Unit
- **ML:** Machine Learning
- **NLP:** Natural Language Processing
- **RGB:** Red, Green, Blue
- **ROC:** Receiver Operating Characteristic
- **SLR:** Sign Language Recognition
- **XAI:** Explainable Artificial Intelligence

List of Figures

Fig. 1. Artificial Intelligence (AI)	10
Fig. 2. Computer Vision (CV)	10
Fig. 3. Deep Learning (DL)	11
Fig. 4. How CNN Works	11
Fig. 5. Convolutional Layers	12
Fig. 6. Pooling Layers	12
Fig. 7. Fully Connected Layers	12
Fig. 8. Natural Language Processing (NLP)	13
Fig. 9. Explainable AI (XAI)	13
Fig. 10. Experimental Framework	22

List of Tables

Table 1: Comparison of Sign Language Recognition Studies.....**18**

Chapter 1: Introduction and Overview

1.1 Introduction

Artificial intelligence (AI) is a big part of modern technology. Computer vision is one of its most useful applications. It lets machines see and understand how people move their hands.

These systems are much better at accurately recognizing hand signs thanks to convolutional neural networks (CNN). This has made it much easier for Deaf and Hard-of-Hearing people to communicate with others.

The main way Deaf and Mute (D&M) people communicate is through sign language, but most hearing people don't know how to read it, which makes communication difficult. Old systems used big hardware that was hard to use and uncomfortable. Some problems still exist, but newer systems only use a regular camera, which is a much better solution.

The purpose of this study is to develop a system that recognizes American Sign Language and converts hand signs into readable text. This will help bring the Deaf community and the rest of society closer together.

1.2 Problem Definition

Sign language is imperative to the Deaf and Mute (D&M) community.

Sign languages, including American Sign Language (ASL), are organized linguistic systems but pose obstacles to communication for the hearing population, which affects D&M people across education, health, and social life.

ASL is based on fingerspelling as a way of spelling words, and here it involves relevant automated recognition of the shapes of fingers to bridge the communication gap.

The hardware sensors were the only basis for building the first-generation motion recognition systems, and their invasiveness limited their usefulness.

This drawback led to a shift towards contactless, vision-based methods that used ordinary cameras, but these also faced new challenges, including limited sensitivity to varying conditions and backgrounds.

Deep learning and Convolutional Neural Networks (CNNs) have assisted in enhancing the feature extraction of raw images, but there are two significant gaps in the research: real-time predictions cannot be stable with time, and no linguistic post-processing exists.

New CNN-based systems tend to assign labels to every frame, with classes that are constantly changing due to the changing environment.

Moreover, they typically lack a post-processing step to correct recognition errors, thus yielding raw results.

The research aims to develop a web-deployable ASL fingerspelling recognition.

A system that incorporates high accuracy and linguistic error correction for effective communication within the D&M community.

1.3 Thesis Objectives

The overall objective of this thesis is to develop and test a real-time fingerspelling recognition system for the ASL alphabet using convolutional neural networks. The objectives of this thesis to develop an efficient preprocessing system for automatically locating and isolating the hand region from images captured using a webcam, and to remove the impact of environmental changes (lighting, background, etc.), to create and train a CNN classifier that can accurately recognize all 26 ASL letter gestures, to make sure that recognition from static images is quick and accurate, with as little processing time as possible, to incorporate post-processing techniques (spell checking, etc.) For improving the letter sequences into a readable text output and developing a web-based application for processing static images.

1.4 Thesis Contribution

Next Semester.

1.5 Thesis Organization

Next Semester.

Chapter 2: Background and Related Work

2.1 Introduction

The proposed American Sign Language (ASL) recognition system requires combining several fields, including computer vision, deep learning, and artificial intelligence.

This chapter goes into great detail on these ideas, starting with the basics of AI and then delving into computer vision and deep learning.

Like convolutional neural networks, it discusses the difficulties of recognizing sign language by showing how it can be used in application, preprocessing, and post-processing activities as necessary components of the system architecture. The chapter attempts to demonstrate how these ideas can be applied to the needs of system performance assessment. It even seeks to justify architectural and methodological choices of the next chapter or to place them in perspective within the existing literature on the subject.

2.1.1 Artificial Intelligence (AI)

Artificial Intelligence (AI) is the study of creating systems that can perform cognitive tasks such as learning, perceiving, reasoning, and decision-making. AI learns from experience and can adapt to new situations, unlike traditional rule-based systems. It uses statistical learning methods to discover patterns in unstructured data, such as images, audio, and text. This study utilizes AI to recognize sign language, focusing on accurately identifying and classifying visual communication data using advanced machine learning algorithms and computer vision techniques [1].

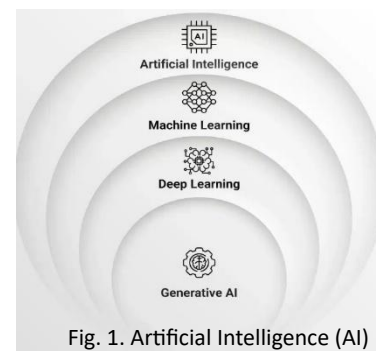


Fig. 1. Artificial Intelligence (AI)

2.1.2 Computer Vision (CV)

Computer vision is a branch of artificial intelligence that lets machines analyze and understand visual data from photos and videos. It turns raw pixel data into useful representations that help with tasks such as pattern recognition and object recognition [2].

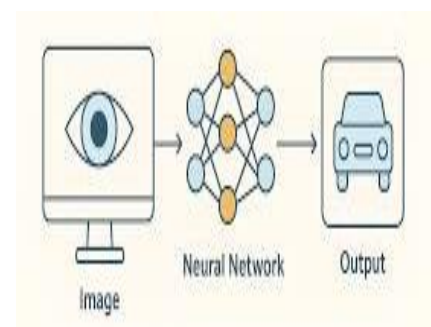


Fig. 2. Computer Vision (CV)

2.1.3 Deep Learning (DL)

Deep learning is a branch of machine learning that uses multi-layered computational networks to perform hierarchical computation, i.e., transforming input data into more abstract representations and adapting them to specific tasks. The transformation stages through which the data is changed before the final prediction is obtained are referred to as deep.

Deep learning is used to address key limitations of previous computer vision methods for sign language recognition [3].

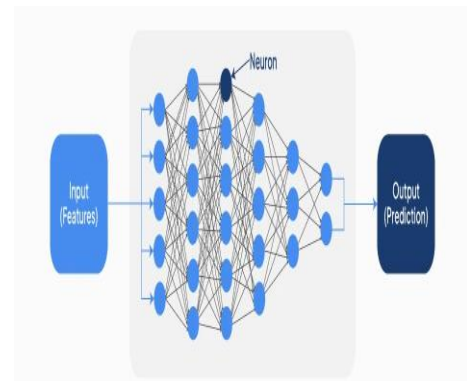


Fig. 3. Deep Learning (DL)

2.1.4 Convolutional Neural Networks (CNN)

Convolutional neural networks (CNNs) are high-level deep learning systems designed to handle data with a discrete structure, in particular, two-dimensional images.

The key feature of them is that they are based on convolutional operations that exploit the spatial cohesiveness and regularity of the transition of visual information. This also enables more efficient learning of hierarchical feature representations than traditional fully connected architectures.

A standard convolutional neural network processes input images in several steps to extract features and group them.

The CNN model performs both feature extraction and classification through fully connected and softmax layers [4].

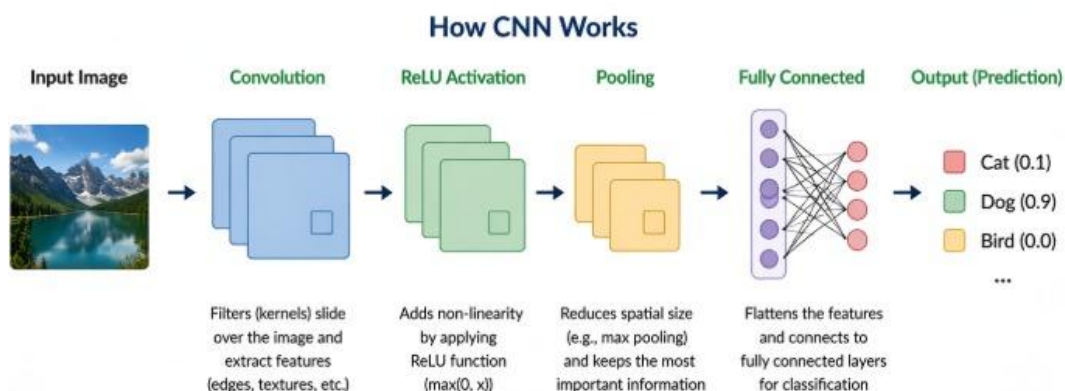


Fig. 4. How CNN Works

Convolutional Layers

The most important part of a CNN is its convolutional layers. They use filters that can learn to scan input feature maps and make activation maps that show important visual patterns [5].

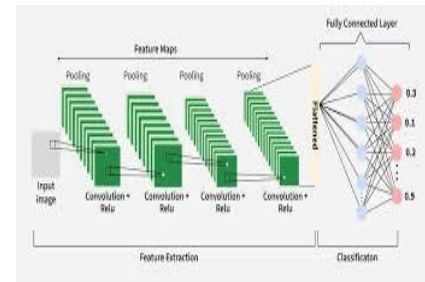


Fig. 5. Convolutional Layers

Activation Function (ReLU)

The Rectified Linear Unit (ReLU) is an activation function that makes CNNs more complex. It outputs zero for negative inputs and the input value for positive inputs. This helps the network learn complex patterns quickly and address the vanishing-gradient problem [6].

Pooling Layers

Pooling layers in CNNs reduce the spatial dimensions of feature maps, lowering computational cost and making features more stable. Max pooling, the most common type, retains the highest value in local regions, preserving important features while removing unnecessary information [7].

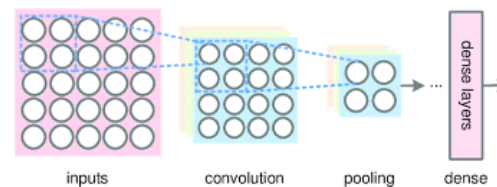


Fig. 6. Pooling Layers

Fully Connected Layers

Fully Connected Layers

Fully connected layers are used at the final stage of a CNN to combine extracted features and perform classification. These layers connect all neurons from the previous layer to generate output probabilities for each class [8].

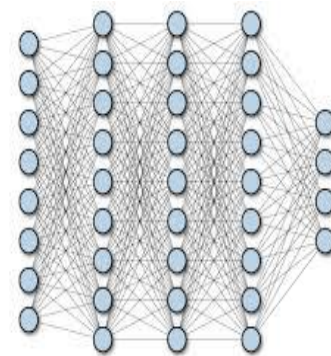


Fig. 7. Fully Connected Layers

Dropout Regularization

Dropout is a technique to prevent CNNs from overfitting by randomly deactivating some neurons during training. This makes the model stronger and more general [9].

2.1.5 Natural Language Processing (NLP)

Natural Language Processing (NLP) is the computational processing of human language and may encompass tasks such as tokenization and sentiment analysis. It is not the topic of the current work, but it gives rise to the subsequent linguistic process of system outputs. This process enhances the accuracy of recognition products, which is important in gesture recognition applications such as text-to-speech and machine translation [10].

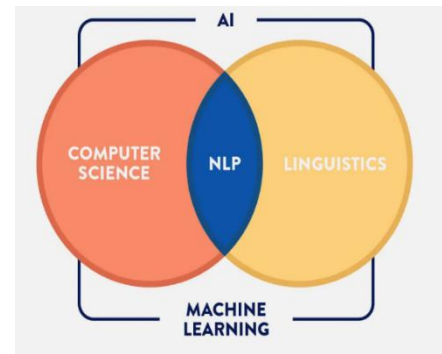


Fig. 8. Natural Language Processing (NLP)

2.1.6 Explainable AI (XAI)

Explainable AI (XAI)

Explainable Artificial Intelligence (XAI) seeks to elucidate AI decision-making for humans, mitigating the opacity of intricate models.

Techniques such as class activation mapping and model-agnostic approaches enhance the clarity of model predictions by highlighting salient features [11].

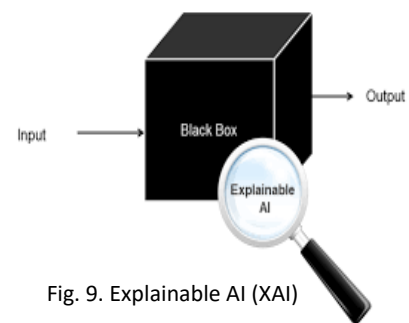


Fig. 9. Explainable AI (XAI)

2.1.7 Image Preprocessing Techniques

The quality of input to a convolutional neural network (CNN) classifier is crucial for hand gesture recognition. To get good results, the image-preprocessing pipeline needs to be well-organized. This process enhances features and reduces unnecessary variation, thereby boosting the audience's confidence in the method's ability to improve model performance [12].

2.1.8 Linguistic Post-Processing

Linguistic Post-Processing is an important step in improving the accuracy and consistency of gesture recognition systems in real-time applications. It addresses issues such as recognition failures, differences between human sign language and the American Sign Language alphabet, and similar characters. This step helps find and fix mistakes by putting the recognition engine in context. This makes the system's language output more reliable, even when things go wrong [13].

2.1.9 Evaluation Criteria

The proposed recognition system will use a complex assessment model to measure performance in multiple ways, particularly for the 36-category American Sign Language (ASL) gesture classification task [14].

2.1.10 Accuracy

Accuracy is the most important measure; it is the percentage of correctly predicted cases in the test set. Despite its usefulness, relying solely on accuracy can be misleading in cases of class imbalance, as the model may be highly accurate while overrepresenting the more common classes at the expense of the less common ones. Thus, criteria that assess individual classes are necessary to have a complete assessment [15].

2.2 Related Work

Review of Related Work

The author [16] proposed a hybrid model comprising a dual-path Convolutional Neural Network architecture integrated with a Vision Transformer, which recognized 29 American Sign Language signs from images. They applied this approach to extract gesture features and hand details, which were merged using element-wise multiplication to enhance recognition. They used the ASL dataset for training and testing, comprising 87,000 RGB images with a wide range of hand poses and lighting conditions.

The hybrid model achieved the highest accuracy, reaching 99.97% when tested on a set of 29 distinct signs.

The author [17] proposed a vision-based recognition system. This system used Transfer Learning to identify thirty-two signs from the Arabic Sign Language dataset.

They used a pretrained Deep Learning (DL) model, specifically EfficientNetB4, to handle different signing styles and extract features from hand gestures.

The ArSL2018 dataset, comprising 54,049 RGB images from 40 subjects, was used for training and testing to ensure a variety of lighting and camera positions.

The model achieved the highest accuracy, reaching 95% across the 32 signs.

The authors [18] proposed a real-time fingerspelling recognition system that utilized a 2D Deep CNN with two-stream feature extraction. They applied a method to capture spatial and temporal information from dynamic gestures to improve communication for people who are deaf or hard of hearing. They used a dataset for training and validation, which included a variety of gestures recorded in different environments.

The model significantly outperformed machine learning techniques and achieved high accuracy as data volume increased.

The authors [19] proposed a real-time system to convert American Sign Language fingerspelling into text and speech using a Convolutional Neural Network. They filtered the hand images before sending them to the model to improve class predictions. The researchers used an ASL dataset for training and testing.

The results showed that the model achieved high accuracy in predicting gestures and provided an effective real-time solution.

The authors [20] proposed a novel system for image-based semantic communications, specifically designed for American Sign Language transmission. They applied an encoder-decoder method to transform ASL images into semantic feature vectors, thereby reducing data size. The researchers utilized the dataset to validate the system's ability to reconstruct original images even under noisy conditions.

The results showed that their proposed semantic approach outperformed communication methods, maintaining high structural similarity and recognition accuracy at much lower

transmission rates.

The author [21] presented a comprehensive review of vision-based sign language recognition systems and discussed existing techniques. The researcher analyzed the images by considering preprocessing, feature extraction, and classification techniques to bridge the communication gap. The researcher evaluated the methodologies and the datasets that were already in place. The datasets included sign languages such as ASL and ISL. The results were presented, and it was clear that the deep learning techniques attained the performance.

The authors [22] proposed a novel American Sign Language recognition model using complex data representation. The authors proposed the Complex Zernike Moments technique for effective feature extraction. The authors proposed a Complex-Valued Deep Neural Network to effectively process extracted features. The authors applied the Complex-Valued Deep Neural Network technique to the three benchmark datasets. The datasets included the ASL Dataset, which contains thousands of images. The results system attained 98.67% accuracy and computational efficiency.

The authors [23] proposed a comprehensive system for recognizing and generating American Sign Language (ASL) using a Convolutional Neural Network (CNN). They used a deep learning technique for the recognition and generation of American Sign Language. This technique processes images using several layers of the Convolutional Neural Network. This approach guarantees reliable feature extraction from images, regardless of lighting conditions and variations in skin tone. The researcher used a dataset comprising 25,000 images depicting hand gestures corresponding to the 26 letters of the alphabet and frequently used words. The researcher's model demonstrated a noteworthy training accuracy of 98.4%, thereby facilitating improved communication between deaf and hearing populations.

The author [24] has proposed a system for recognizing Arabic Sign Language using a Convolutional Neural Network to generate speech. He has used a specific architecture for speech recognition and generation, employing several layers of a Convolutional Neural Network. This has enabled automatic image feature extraction. The researcher has used a large image dataset to generate speech. This dataset comprises thousands of images to support data generalization.

The results show that the researcher's proposed model for speech recognition and generation using a Convolutional Neural Network achieves 90% accuracy.

The authors [25] proposed a Convolutional Neural Network-based system for classifying American Sign Language finger spelling. The authors compared two deep learning models, ResNet50 and DenseNet121, for American Sign Language classification using Convolutional Neural Networks. The authors have used different data-splitting techniques, i.e., 70:30, 80:20, and 90:10, to evaluate the model's performance in classifying hand gestures. They used a publicly available American Sign Language dataset from Kaggle, which includes 24 static hand-

gesture classes (excluding 'J' and 'Z' due to their hand movements). The researchers conducted their study. The results show that the ResNet50 model with an 80:20 split outperforms DenseNet121, achieving near-perfect accuracy with higher sensitivity and precision.

The authors [26] proposed a gesture recognition system aimed at one practical problem: deaf users need tools that actually keep up with real conversation. They used YOLOv3 with DarkNet-53 for feature extraction. The dataset was collected from scratch, totaling 2,000 images, which isn't huge. Still, the model landed at 97.68% MAP. More importantly, it ran faster than SSD and Faster R-CNN while sacrificing little in accuracy. For something meant to work in live communication, speed matters as much as getting the gesture right.

The authors [27] proposed an ASL recognition system that kept the architecture relatively simple: a CNN implemented in Keras and TensorFlow, fed with webcam frames. Before classification, images were converted to grayscale. That strips out color noise and keeps the model focused on hand shape, which is really all that matters here. The training data was solid, with 25,110 images covering all 26 letters and the digits 0 through 9. The model achieved 99% training accuracy and 98% test accuracy. The gap between those two numbers is small enough that overfitting doesn't seem to be a serious concern.

The author [28] proposed a YOLOv5-based recognition system after finding that conventional models failed under messy background conditions. The single-pass detection approach combines identification and classification, reducing processing time. Data came from Roboflow, but the team added their own images on top to make the training set more representative. The accuracy comparison is worth noting: their model achieved 88.4%, while a standard CNN on the same task achieved only 52.98%. That gap is too wide to explain away as noise. It suggests detection-based framing suits this problem better than classification alone.

The authors [29] proposed ASLR-Net, a CNN designed to handle the inconsistencies that break most gesture models outside lab conditions. Lighting changes, slight angle shifts, and background clutter augmentation during preprocessing were how they addressed all of that without collecting an enormous dataset. The ASL alphabet dataset gave them enough material to make augmentation worthwhile. Final accuracy came in at 98.5%. What stands out isn't just the number, but that it held up under conditions the model hadn't been explicitly trained on, which is usually where these systems quietly fail.

The authors [30] proposed a CNN system for recognizing sign language from video rather than still images, a harder version of the problem. Static gesture recognition is one thing, but continuous motion across overlapping introduces timing, overlap, and transition issues that don't exist in single-image setups. The dataset was video-based, covering a variety of hand shapes and motions to reflect how signing looks in practice. Preprocessing standardized the frames before they hit the classifier. The system accurately tracked hand movement across live

sequences, a more meaningful result than achieving a high score on a clean-image benchmark.

Reference	Years	Strong	Limitations	Datasets	The Best models	Results	Link of Dataset
[16]	2025	Attained a near-perfect accuracy of 99.97%	Focused only on static images	ASL Alphabet	The Hybrid Transformer-CNN model emerged	Accuracy = 99.97%	https://www.kaggle.com/grassknotted/asl-alphabet
[17]	2022	High efficiency in handling complex images	Required high computational power	ArSL2018 dataset	EfficientNetB4 model	Accuracy = 95%	https://www.kaggle.com/code/zssash/arabic-sign-language-image-classification/input
[18]	2024	Model integrated advanced feature extraction	Required a large volume of training data	ASL and ISL	2D Deep CNN	High accuracy in practical	Various sign languages (ASL) [35] and Indian Sign Language [36]
[19]	2021	The model integrated a pre-processing filter that improved the accuracy	Focused mainly on static fingerspelling	ASL Alphabet	Convolutional Neural Network (CNN)	Accuracy = 95.8%	
[20]	2024	Achieved high-quality image reconstruction	Required significant computational overhead	ASL Alphabet	Semantic-based Deep Learning model	Accuracy = 99%	
[21]	2020	Provided a detailed taxonomy of recognition phases.	Complex backgrounds still posed challenges for many systems.	including ASL, ISL, and TSL datasets	Convolutional Neural Networks and Hidden Markov Models	Higher accuracy and robustness in real-world scenarios.	
[22]	2024	Attained exceptional accuracy by utilizing complex numbers	Required specialized mathematical frameworks	ASL Alphabet	Complex-Valued Deep Neural Network (CVDNN)	Accuracy = 98.67%	
[23]	2023	Attained high robustness against environmental factors	Noted the hardware's processing power	ASL letters and symbols	Convolutional Neural Network (CNN)	Accuracy = 98.84%	https://www.kaggle.com/datasets/datamunge/sign-language-mnist
[24]	2020	Attained a very high accuracy	Required computational resources during the training	Arabic Sign Language dataset	Convolutional Neural Network (CNN)	Accuracy = 90%	
[25]	2021	Attained exceptional results using the ResNet 50	Excluded the letters 'J' and 'Z.'	ASL Alphabet dataset	ResNet 50 model	Accuracy = 0.999913	https://www.kaggle.com/grassknotted/asl-alphabet?select=asl_alphabet_train
[26]	2021	Attained high speed and accuracy in detecting	Required a high-performance GPU	custom dataset of 2,000 images	YOLOv3 model	Accuracy = 97.68%	
[27]	2020	Utilized a robust CNN architecture	Restricted to static hand gestures	ASL	Convolutional Neural Network (CNN)	Accuracy = 98%	
[28]	2022	Attained much better results than traditional CNNs	Precision was recorded at 76.6%, which is lower	utilized the Roboflow labeled dataset	YOLOv5 model	Accuracy = 88.4%	

[29]	2023	High performance by utilizing an optimized CNN structure	Focused on static hand gestures	Sign Language MNIST dataset	ASLR-Net (CNN)	Accuracy = 96.9%	https://www.kaggle.com/datasets/oladayoluke/enhanced-sign-language-mnist-dataset
[30]	2023	Attained the ability to perform real-time tracking of hand gestures	The model required a large and diverse collection of video data to improve its accuracy	ASL alphabet	Convolutional Neural Network (CNN)	Attained reliable recognition & higher accuracy	

References

- [1] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- [2] Szeliski, R. (2022). *Computer Vision: Algorithms and Applications* (2nd ed.). Springer.
- [3] LeCun, Y., Bengio, Y., & Hinton, G. (2015) Deep learning. *Nature*, 521(7553), 436–444.
- [4] Li, Y., Zhang, K., & Li, K. (2021). Deep learning for image classification: A survey. *Neurocomputing*, 449, 20–36.
- [5] Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent
- [6] Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent
- [7] Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8), 5455–5516.
- [8] Li, Y., Zhang, K., & Li, K. (2021). Deep learning for image classification: A survey. *Neurocomputing*, 449, 20–36.
- [9] Srivastava, N., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- [10] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022) Natural language processing: State of the art, current trends and challenges.
- [11] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- [12] Szeliski, R. (2022). *Computer Vision: Algorithms and Applications* (2nd ed.). Springer.
- [13] Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing* (2nd ed.). Pearson.

- [14] gesture classification task (Goodfellow et al., 2021)
- [15] (Géron, 2022)
- [16] Aly, Mohammed, and Islam S. Fathi. "Recognizing American Sign Language gestures efficiently and accurately using a hybrid transformer model." *Scientific Reports* 15.1 (2025): 20253.
- [17] Zakariah, Mohammed, et al. "Sign language recognition for Arabic alphabets using transfer learning technique." *Computational Intelligence and Neuroscience* 2022 (2022): 4567989.
- [18] Zhidebayeva, Aziza, et al. "Real-time sign language fingerspelling recognition system using 2D deep CNN with two-stream feature extraction approach." *International Journal of Advanced Computer Science and Applications* 15.9 (2024).
- [19] Yadav, B. K., & Bohra, H. "Sign Language to Text and Speech Conversion." *International Journal of Advanced Research, Ideas and Innovations in Technology (IJARIIT)* 7.3 (2021): V713-1560.
- [20] Kouvakis, V., Trevlakis, S. E., & Boulogeorgos, A. A. A. "Semantic Communications for Image-Based Sign Language Transmission." *IEEE Open Journal of the Communications Society* 5 (2024): 1083-1100. [5] Zhidebayeva, Aziza, et al. "Real-time sign language fingerspelling recognition system using 2D deep CNN with two-stream feature extraction approach." *International Journal of Advanced Computer Science and Applications* 15.9 (2024).
- [21] Sharma, Sakshi. "Vision-based sign language recognition system: A Comprehensive Review." *Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT)* (2020): 978-1-7281-4685-0.
- [22] Bayrak, S., Nabiyeve, V., & Atalar, C. "American Sign Language Recognition Model Using Complex Zernike Moments and Complex-Valued Deep Neural Networks." *IEEE Access* 12 (2024): 193001-193014.
- [23] Sheth, P., Dedhia, R., Chheda, A., & Sawant, V. "American Sign Language Recognition and Generation: A CNN-based Approach." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)* 9.4 (2023): 241-250.
- [24] Kamruzzaman, M. M. "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network." *Wireless Communications and Mobile Computing* 2020 (2020): 3685614.
- [25] Marjusalinah, A. D., Samsuryadi, S., & Buchari, M. A. "Classification of Finger Spelling American Sign Language Using Convolutional Neural Network." *Computer Engineering and Applications* 10.2 (2021): 91-102.

- [26] Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. "Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model." *Applied Sciences* 11.9 (2021): 4164.
- [27] Hurroo, M., & Walizad, M. E. "Sign Language Recognition System using Convolutional Neural Network and Computer Vision." *International Journal of Engineering Research & Technology (IJERT)* 9.12 (2020): 2278-0181.
- [28] Bankar, S., Kadam, T., Korhale, V., & Kulkarni, A. A. "Real Time Sign Language Recognition Using Deep Learning." *International Research Journal of Engineering and Technology (IRJET)* 9.4 (2022): 2395-0056.
- [29] Aggarwal, J., Gupta, L., Gupta, M., Nitin, & Girdhar, P. "Real-Time Sign Language Recognition Using CNNs." *Advances and Applications in Mathematical Sciences* 22.11 (2023): 2199-2211.
- [30] Drishya. "Sign Language Recognition." *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)* 5.4 (2023): 56726/IRJMETS37594.

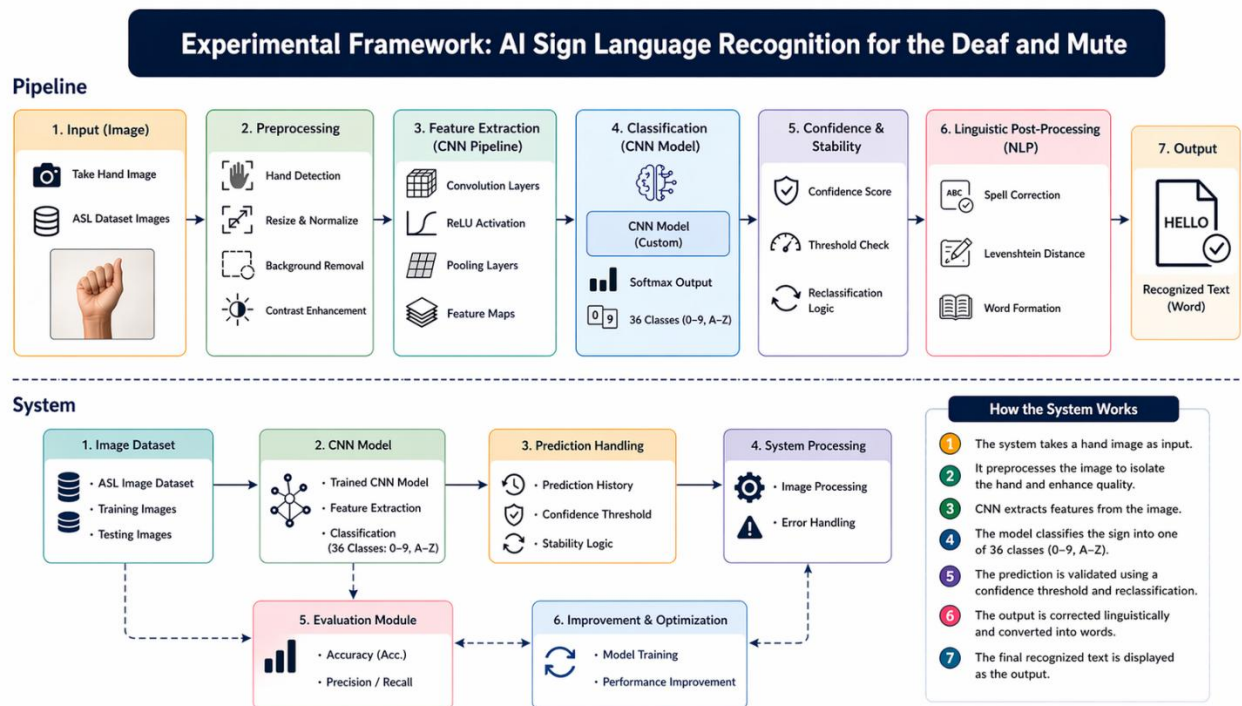


Fig. 10. Experimental Framework