

محمد القثمي

AI Sign Language Recognition for the Deaf and Mute.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:379377656

Submission Date

Apr 24, 2026, 1:00 AM GMT+3

Download Date

Apr 24, 2026, 10:47 AM GMT+3

File Name

AI Sign Language Recognition for the Deaf and Mute.docx

File Size

1.3 MB

21 Pages

4,418 Words

26,269 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

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.

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 very experienced and has the vision that has been shaped through constructive criticism, and this has greatly contributed to the success of this research being of high standard and scientifically rigorous.

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.

She deserves our thanks for many things, including her patience, support, and encouragement in our academic journey.

It will take thousands of thanks from us to repay her for her time and effort in trusting us.

Abstract

AI is an integral aspect of modern technology.
Among the most practical applications of AI is Computer Vision.

It allows computers to recognize and interpret visual data, including hand gestures. CNN, an artificial neural network (ANN) algorithm that employs deep learning, has greatly enhanced the efficiency of gesture recognition systems.

Sign language is the primary mode of communication for the Deaf and Mute (D&M) communities, but their counterparts who are not deaf may fail to comprehend it.

Therefore, D&M individuals may encounter difficulties communicating with other people. Earlier systems relied on bulky equipment, which posed operational challenges.

Modern ASL translation devices use conventional cameras, simplifying the process; however, certain issues persist.

The current project seeks to design a real-time ASL Recognition System that translates hand gestures into readable text.

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)	12
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.....	13
2.2 Related Work	14
Review of Related Work	14
References	18

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)	12
Fig. 9. Explainable AI (XAI).....	13
Fig. 10. Experimental Framework	21

List of Tables

Table 1: Comparison of Sign Language Recognition Studies.....**17**

Chapter 1: Introduction and Overview

1.1 Introduction

Artificial intelligence is an important element of contemporary technological advancements. One of its practical implementations includes computer vision that allows machines to perceive and analyze people's gestures and actions.

Computer vision systems have become more precise at recognizing hand signs due to convolutional neural networks. Consequently, this advancement has helped Deaf and Hard of Hearing individuals to interact with others more easily.

Most Deaf and Mute individuals use sign language as a means of verbal communication but since most hearing people do not understand it, there arise difficulties communicating with each other. Previous solutions were based on using bulky hardware, making it inefficient and impractical. There still are some challenges associated with such projects but new developments use nothing but a simple camera.

Thus, the goal of this research project is to design an innovative application for recognition of American Sign Language and translating hand signs into written messages. The aim is to create an opportunity to integrate the Deaf community with other people more closely.

1.2 Problem Definition

The sign language is of great importance in the lives of people within the Deaf and Mute community.

Sign language, as seen in the case of American Sign Language, is unique in its arrangement, which means that it poses communication barriers for hearing people hence influencing various aspects such as health, education and many others of the D&M community.

American Sign Language has fingerspelling, a form whereby words are spelled through fingers, with finger automatic recognition being used in accomplishing the feat.

Hardware sensors have been the core components of early motion recognition devices. Invasiveness posed a major problem in their utilization.

The development of non-invasive methods that employed the use of cameras to perform the task was another milestone, although other difficulties, such as lack of sensitivity to environment change, came up.

Deep learning and Convolutional Neural Networks (CNN) have helped in Improving feature extraction from raw images; however, there are two fundamental problems with this approach: predictions are not invariant to time, and there are no post-processing techniques of linguistic nature.

Most approaches employing CNNs classify each frame into a particular category, and since these categories always vary due to change in environments, so do classes.

Also, these approaches have no post-processing steps to ensure proper recognition of gestures.

This study focuses on the development of ASL fingerspelling recognition method deployable on a website.

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.

2.1.1 Artificial Intelligence (AI)

Artificial intelligence is an area of study involving machines and software able to engage in cognitive activities such as learning, perceiving, thinking, and decision making. In addition to learning from experience and being adaptive to new environments, artificial intelligence is distinct from the conventional programmed approach in its reliance on statistical learning models for mining information from unstructured data. This research will leverage AI technology to classify visual communication data through sign language recognition [1].

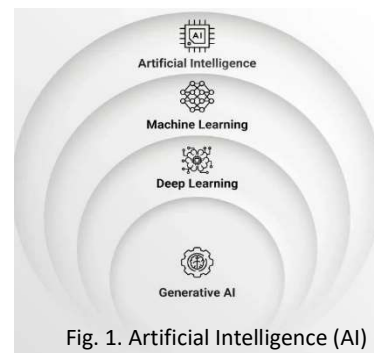


Fig. 1. Artificial Intelligence (AI)

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].

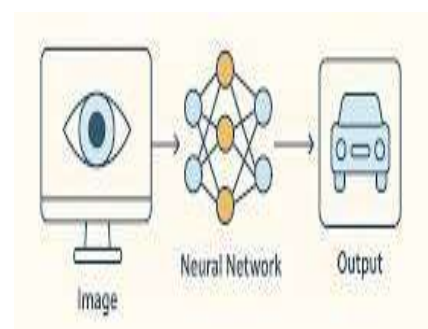


Fig. 2. Computer Vision (CV)

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].

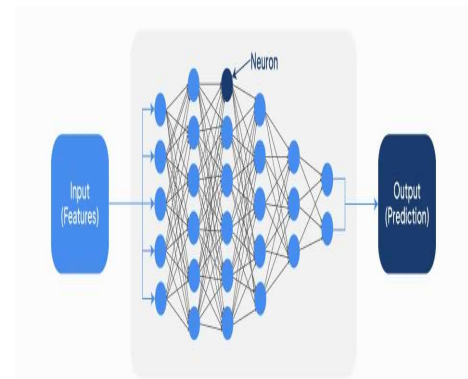


Fig. 3. Deep Learning (DL)

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].

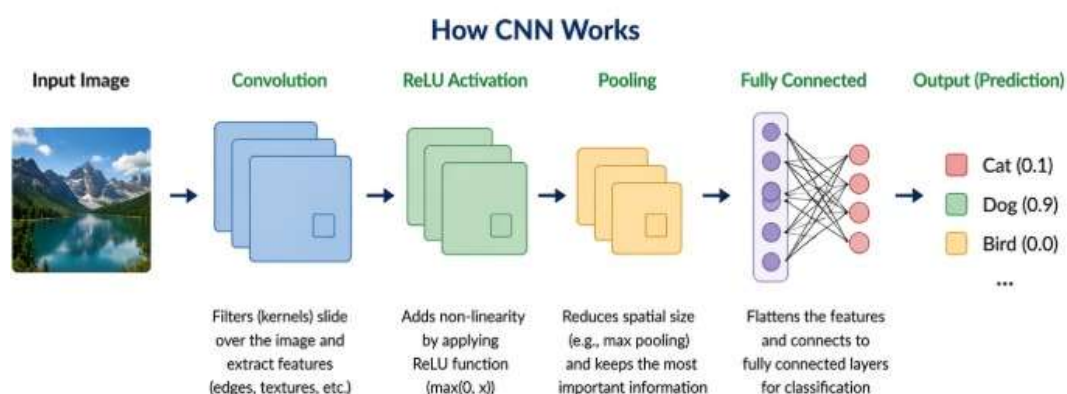


Fig. 4. How CNN Works

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].

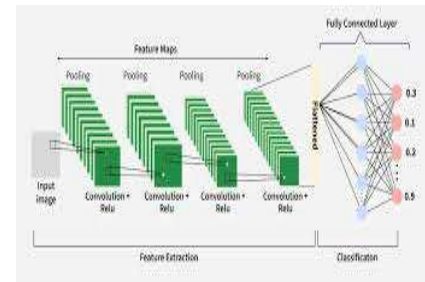


Fig. 5. Convolutional Layers

Activation Function (ReLU)

ReLU is an activation function that complicates the architecture of the CNN. ReLU gives output 0 for any negative input and outputs for positive values. It helps CNN learn fast and solve the vanishing gradient problem [6].

Pooling Layers

Pooling in the CNN helps in shrinking the feature map to reduce computational complexity as well as making the features resistant. Max pooling is the most frequently utilized type of pooling in which the maximum values in the region are preserved [7].

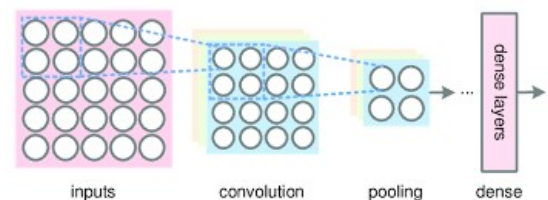


Fig. 6. Pooling Layers

Fully Connected Layers

applied to the end part of a CNN to integrate all the obtained features and classify them. This layer connects to all neurons in the previous layer to produce outputs [8].

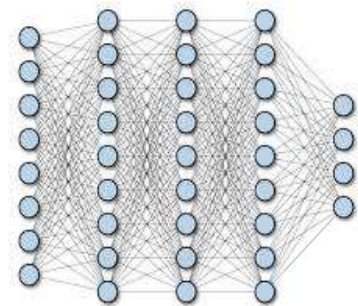


Fig. 7. Fully Connected Layers

Dropout Regularization

Dropout is a method that helps avoid overfitting for CNNs through random deactivation of some neurons in the model. The end result is enhanced learning capability [9].

2.1.5 Natural Language Processing (NLP)

Natural Language Processing means computer processing of natural language and includes methods such as tokenizing and analyzing sentiment. However, NLP does not fall within the ambit of this paper and ends up in linguistic processing of the system output. Linguistic processing helps enhance recognition performance, an important aspect of applications such as text-to-speech and translation [10].

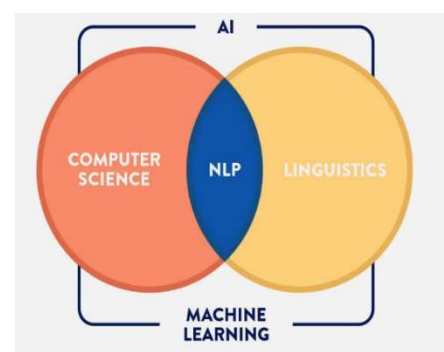


Fig. 8. Natural Language Processing (NLP)

2.1.6 Explainable AI (XAI)

XAI is an approach towards making AI decision making transparent to humans. There is a lot of complexity involved in AI systems, and XAI aims at clarifying these decision processes. Various techniques such as class activation mapping and model agnostic methods improve explainability [11].

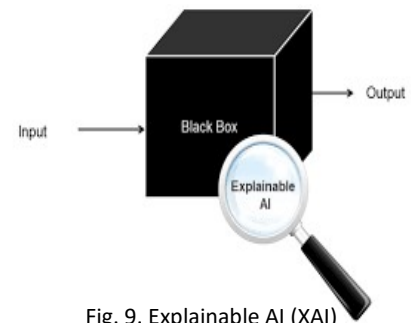


Fig. 9. Explainable AI (XAI)

2.1.7 Image Preprocessing Techniques

A CNN classifier depends on the quality of input that is provided to it for recognizing hand gestures. In order to achieve good results, there should be an organized procedure of image pre-processing. This would improve the features and minimize variability, leading to improved confidence among the audience in the effectiveness of the approach [12].

2.1.8 Linguistic Post-Processing

Post linguistic processing is a vital part of increasing the effectiveness of the process of recognition of gestures in practice.

It is used to deal with such problems as recognition errors; differences between sign language used by people and the ASL alphabet and characters that look alike. This process helps discover and correct errors through putting the process of recognition into context [13].

2.1.9 Evaluation Criteria

The recognition system that we propose to develop will employ a sophisticated evaluation criterion to evaluate its performance in various respects especially regarding ASL gesture classification for the 36 categories [14].

2.1.10 Accuracy

The criterion of accuracy is of the highest priority because it reflects the ratio of correct predictions made by the model out of all the objects in the test sample. Accuracy, though being very informative, cannot be applied alone since it gives rise to a false impression about the quality of classification in case of unequal distribution of classes because the model will be very accurate but will ignore minority classes [15].

2.2 Related Work

Review of Related Work

The authors [16] suggested a hybrid model combining a dual-path Convolutional Neural Network architecture with a Vision Transformer. It can identify 29 American Sign Language gestures from images. To extract features of gestures and hands, they used this technique. Then they multiplied the two elements, thereby improving their recognition capabilities. They utilized the ASL dataset for training and evaluation, which contained 87,000 colored images of various hand postures and lighting conditions.

The researcher [17] proposed a vision-based image recognition algorithm. The researcher adopted the Transfer Learning technique to recognize the thirty-two signs available from the Arabic Sign Language Dataset. He utilized a pretrained deep learning model called EfficientNetB4 which is able to consider the variations of different signing styles and feature extraction using hand gestures. The used dataset contains 54,049 RGB images of 40 different people to capture variations in illumination and camera position. The maximum accuracy obtained was 95%.

Authors [18] have introduced the development of a real-time fingerspelling recognition method by applying a 2D deep CNN where two streams were utilized for feature extraction. They created a method for extracting spatial-temporal features from gestures, thus helping the deaf population in their day-to-day tasks. They employed a dataset for training and validating their model by including many gestures collected in different environments. Their proposed method excelled over other machine learning methods and exhibited high accuracy, especially with the expansion of the amount of data.

The Author [19] developed a method that could translate fingerspelling from American Sign Language into text and speech in real time using a CNN. The researchers preprocessed the hand images before feeding them into the CNN to improve classification accuracy. They employed an ASL dataset to train and test their model.

The authors [20] introduced a new approach for image-based semantic communications tailored to the transmission of American Sign Language. The application of the encoder-decoder algorithm helped to represent ASL images using semantic feature vectors to compress data. The dataset was used to test the system's capability of reconstructing images under noise influence. As was found, the suggested system for semantic communications performed better than communication systems in terms of maintaining structure and recognition rates with significantly smaller

The author [21] conducted an extensive analysis of sign language recognition based on vision technologies and examined various approaches. The researcher examined the images from the perspective of preprocessing, feature extraction, and classification techniques to overcome the communication barrier.

The researcher studied the methodology and datasets that are currently being used. The datasets contain sign languages like ASL and ISL. The results were then shared, and it is evident that deep learning techniques achieved the performance level.

The researchers [22] introduced a new model for recognizing ASL gestures using complex data representation. The researchers suggested the use of the Complex Zernike moments algorithm for efficient feature extraction. A new algorithm was designed by the researchers called the Complex-valued Deep Neural Network algorithm. The algorithm was used on the three benchmark datasets. These datasets include the ASL dataset containing several thousands of images.

The algorithm achieved an accuracy rate of 98.67%.

The authors [23] suggested an advanced model that would identify and create ASL through a CNN. The researchers employed deep learning methods to create and identify ASL signs. The method involves the processing of images using multiple levels of a CNN. In doing so, the researchers ensured that the system could accurately extract features from images without considering lighting conditions or skin colors. The researchers relied on a data set consisting of 25,000 images representing hand gestures associated with each of the 26 letters of the alphabet as well as commonly used words.

This created model achieved a remarkable accuracy rate of 98.4% in its training phase, thus enhancing communication between deaf and hearing individuals.

A researcher [24] has suggested a method that is based on the use of a Convolutional Neural Network for Arabic Sign Language recognition, and to produce speech. This particular researcher has designed an architecture that uses multiple layers of the Convolutional Neural Network to recognize speech and generate speech. This allows for automated extraction of features from images. A researcher has taken into consideration a set of images for speech production purposes.

The output of the researcher's research work shows that 90% accuracy is achieved in speech recognition and generation using Convolutional Neural Network.

The researchers [25] introduced a new approach to classify American Sign Language using Convolutional Neural Networks for finger spelling. In this research, the researchers compare two deep learning approaches, ResNet50 and DenseNet121, for American Sign Language classification using Convolutional Neural Networks. The researchers used various methods of data splitting, such as 70:30, 80:20, and 90:10, to test the performance of the model in hand gesture classification. The publicly available data set of American Sign Language from Kaggle was used in their research, which contains 24 classes of static hand gestures (the letters 'J' and 'Z' were excluded from the data set because of the hand movement involved in those letters). The researchers conducted their study.

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, 2,000 images total,

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 hit 99% training accuracy and 98% on testing. 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 handles identification and classification together, which cuts 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 reached 88.4%, while a standard CNN on the same task only managed 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, background clutter augmentation during preprocessing was 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 the fact that it held up under conditions the model hadn't explicitly been trained on, which is usually where these systems quietly fail.

The authors [30] proposed a CNN system built to recognize sign language from video rather than still images, which is 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 tracked hand movement accurately across live sequences, which is 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	The model required a large and diverse collection of	ASL alphabet	Convolutional Neural Network (CNN)	Attained reliable recognition & higher accuracy	

		of hand gestures	video data to improve its accuracy				
--	--	------------------	------------------------------------	--	--	--	--

References

- [1] Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
- [2] IBM. (2023). What is Computer Vision?
- [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.

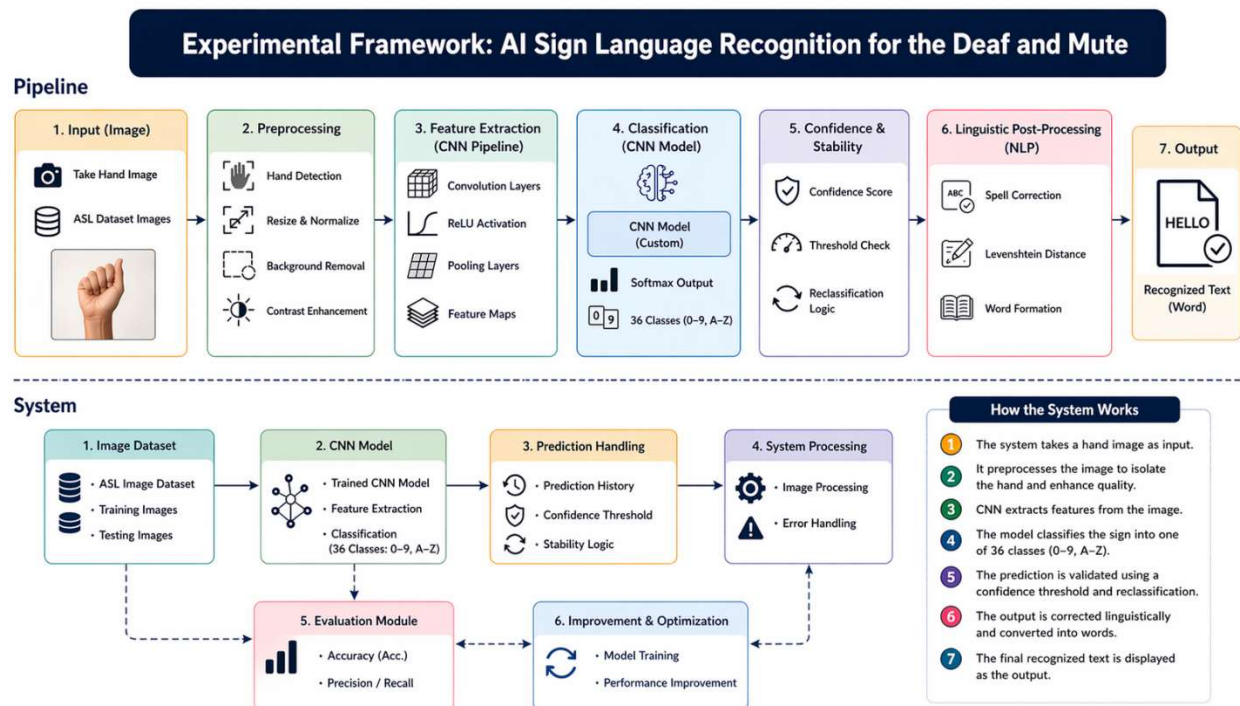


Fig. 10. Experimental Framework