IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 1, March 2025

Sign Language Recognition with Deep Learning

Prof. Neelam Jadhav[1], Sanskruti Daundkar[2], Anjali Gaikwad[3], Ayusha Pawar[4], Sakshi Wadkar[5]

Professor, Computer Engineering Department[1]

Students, Computer Engineering Department[2,3,4,5]

Genba Sopanrao Moze College of Engineering, Balewadi, India[2,3,4,5]

Abstract*: A Sign Language is one of the ways to communicate with deaf people. Inability to speak is true disability. Speech impairment is a disability that affects an individual's ability to communicate using speech and hearing. Mode of communication such as sign language is used by people affected by this impairment. There exists a challenge for non-signers to communicate with signers although the sign language is ubiquitous in recent times. There has been a strong progress in the fields of motion and recognition of gestures with the recent advancements in computer vision and deep learning techniques. The major focus of this work is to create a deep learning-based application that offers sign language translation to text thereby aiding communication between signers and non-signers. We use a custom CNN (Convolutional Neural Network) for recognizing the sign from a video frame. MNIST dataset is used.*

Keywords: Convolution Neural Network (CNN), Gesture Recognition, Pattern Recognition, Sign Language Recognition, NLP

I. INTRODUCTION

Deafness and voice impairment have been persistent disabilities throughout history, hindering individuals from engaging in verbal communication and leading to their isolation from the predominantly vocally communicating society. Sign language has emerged as the primary mode of communication for people with these disabilities. However, it presents a language barrier as it is not commonly understood by those who can hear. To address this issue, various methods for recognizing sign language have been proposed. This paperaims to develop a machine learning-based system that can recognize sign language in real-time.

Loss of hearing may impede the capacity of individuals to express them verbally. This restriction makes it difficult for healthy and deaf individuals to communicate with one another. Due to this, deaf and mute people often find it difficult to have meaningful conversations with people they encounter on a regular basis. The primary method of communication for persons with trouble speaking or hearing impairments is sign language [11]. For this mode of communication to work, the regular individual must be able to understand sign language. Acquiring adequately trained translators in any of the numerous signed languages often requires much time and resources and may also hinder the privacy of the hearing- impaired person [12]. For this reason, many research papers have been carried out to develop sign recognition systems for various languages.

To address this, we use a custom CNN model to recognize gestures in sign language. Convolutional neural network of 11 layers is constructed, four Convolution layers, three Max-Pooling Layers, two dense layers, one flattening layer and one dropout layer. We use the American Sign Language Dataset from MNIST to train the model to identify the gesture. The dataset contains the features of different augmented gestures. Introduced a custom CNN (Convolutional Neural Network) model to identify of the sign from a video frame using Open-CV

Aim & Objective

- "Sign Language Recognition with Deep Learning" is to develop an intelligent system capable of accurately recognizing and translating sign language gestures into text or speech using deep learning techniques.
- The system will leverage advanced computer vision and neural network algorithms, particularly Convolutional Neural Networks (CNNs), to process images or video streams of hand gestures and classify them into corresponding sign language symbols or words.

- To study and analysis various sign recognition systems and gesture recognition using machine learning techniques.
- To design and develop a deep learning-based approach for activity recognition from real-time streaming data.
- To design and develop an algorithm for the detection of gesture signs and predict the results using a conventional neural network.
- To explore and validate the proposed system results with various existing systems and show the effectiveness of proposed approach

II. LITERATURE SURVEY

Numerous research studies have been conducted on sign language recognition, utilizing various sensor technologies and machine learning methods. Below is a quick rundown of some of the recent research that has been carried out on the detection and recognition of sign language.

In their study, Thakur et al. [1] proposed employing a Convolutional Neural Network (CNN) to achieve real-time detection of sign language and generate corresponding speech. The dataset used in the study mostly consisted of American Sign Language alphabet, and the preprocessed gesture dataset were trained using the CNN VGG-16 model with Python libraries and tools such as OpenCV and skimage. The system detected input and generated speech accordingly. The results depicted that the training loss and accuracy were 0.0259 and 99.65%, respectively, and 99.62% of the tests were accurate. By demonstrating the feasibility of using CNN, the study showed how real-time detection of sign language and subsequent speech generation could be achieved, potentially offering a more efficient communication method for people who are deaf or hearing impaired. Further research can be done to broaden the application of the system to include more sign languages and gestures.

In the study conducted by Shrenika and MadhuBala [2], the authors explored the use of a template matching algorithm for sign language recognition. The study involved recording different hand gestures using a camera and processing the images using various algorithms. The first step in the process was pre-processing the image, followed by edge detection to identify the edges of the sign. Once the sign was recognized using the template matching algorithm, the corresponding text was displayed. The system was able to successfully detect basic static hand signs, indicating that template matching can be an efficient technique for sign language recognition.

Tolentino et al. [3] conducted research on the application of deep learning techniques to recognize static sign language. The authors employed a skin-color modeling technique to isolate the pixels of the hand from the background. Afterwards, the images were subjected to classification using a CNN model, which had been trained using Keras. When adequate illumination and a consistent background were maintained, the average testing accuracy reached an impressive 93.67%. The system's accuracy in recognizing American Sign Language (ASL) letters was 90.04%, number recognition was 93.44%, and static word identification was 97.52%. The study showed that deep learning techniques can be useful for static sign language recognition and can outperform previous research in this f ield. The authors suggest that further research could focus on improving the system's accuracy under varying lighting conditions and complex backgrounds to make the system more robust for real-world applications.

Purl et al. [4] proposed a method for Indian Sign Language (ISL) recognition through the use of Python programming language. The program's code was written using Python, and several modules, including Numpy, Os, Tensorflow, Keras, OpenCv (Cv2), and various pre-processors were employed to train the system. To enhance accuracy, the training process utilized both a locally created library of ISL symbols like numbers 0-9, and an online database obtained from GitHub. The proposed algorithm demonstrated an accuracy range of 79 to 100%. The study demonstrated the potential of using Python-based deep learning techniques to recognize Indian Sign Language gestures accurately. The authors suggest further research in t his direction could be fruitful in enhancing the system's efficiency and generalizability to various signing styles.

The study by Sharma and Kumar [5] proposed a technique called ASL-3DCNN for recognizing American Sign Language (ASL) using 3-D Convolutional Neural Networks (CNNs). To prepare the video sequences for processing, the frames were separated and then pre-processed by converting them to grayscale, filtering out noise and spots, and removing illumination variations through histogram equalization. 25 frames are then condensed and normalized before being trained on 3-D CNNs. The suggested technique outperformed current cutting-edge models in metrics including accuracy, recall, and f-measure, with a 0.19 second computation duration per frame, which makes it well suited for real-time applications.

Quin et al. [30] developed a British sign language recognition system with multi-class support vector machines (SVM) and histogram of oriented gradients (HOG) computer vision techniques. A real-world dataset of 13,066 cases was used to successfully test the system, which had an accuracy rate of approximately 99% and a 170 ms mean processing time, making it suitable for real-time visual signature.

III. DESIGN DETAILS

A. Requirement Analysis

Complexity of Gestures: Sign language gestures are complex and may involve dynamic hand movements, body postures, and facial expressions. Variation Across Languages: Different regions have their own versions of sign languages, such as American Sign Language (ASL) and British Sign Language (BSL), making the development of a universal system challenging.Feature Extraction: Extracting relevant features such as finger positions, movement, and orientation from images or videos in real time can be computationally intensive.Noise and Occlusion: Variations in background, lighting conditions, and the occlusion of hands or faces during gestures can affect the accuracy of recognition.

B. Approach & Methodology

Image Acquisition for dataset creation

The Download the dataset from open source websites like Kaggle, MNIST database etc. or create your own database. Sign Language MNIST: A simpler dataset for sign language recognition, focusing on the American Sign Language (ASL) alphabet (A-Z), used for basic classification tasks. The ASL alphabet dataset is a collection of labelled hand gesture images representing the letters A-Z in ASL.

Pre-processing

To prepare the data for further processing at this stage, we did some pre-processing on the images of our proposed dataset. We first converted the gray-scale image from our dataset and then resized it to res 64x64 pixels by keeping the aspect ratio locked. Image Pre-processing and background removal: This is most important phase, as it involves the quality assurance of the data.

Feature Extraction

Images are usually interpreted as color, texture, and shape features.Color is commonly defined as moments and histograms. Properties like, contrast, homogeneity, variance, and entropy, can be attached to texture. Similarly, for shape, roundness, area, eccentricity and concavity characteristics are identified.

Classifier

CNN's architecture is somewhat different from the model of a traditional neural network. Input values are transformed in the conventional neural network into a sequence of hidden layers by traversing. Each layer consists of a series of neurons, where each layer is entirely linked to all the neurons in the previous layer.

C. CNN Algorithm

Convolution: Convolution is a mathematical technique that derives characteristics from input data, often pictures or video frames. In the context of CNNs, a convolutional layer employs a filter (or kernel) on the input to generate feature maps that emphasize significant attributes, such as edges, textures, and patterns.

Pooling

Pooling is a down-sampling process that reduces the spatial dimensions (width and height) of feature maps. The predominant forms of pooling are max-pooling and average-pooling.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, March 2025

Max Pooling: Selects the maximum value from each patch of the feature map, which helps retain the most important features while reducing the size of the data.

Average Pooling: Takes the average of the values in each patch, generally used when smooth features are more important

IV. SYSTEM ARCHITECTURE

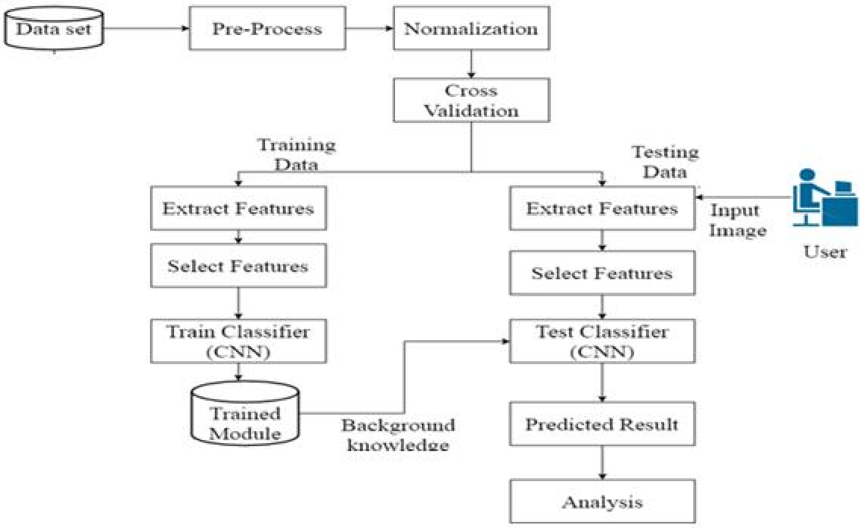

Fig. 1.System Architecture

DFD0

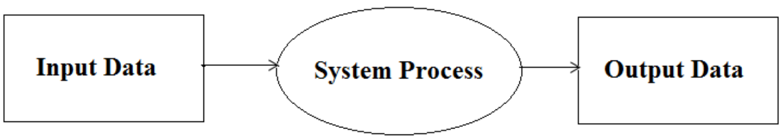

Fig. 2.DFD0

DFD1

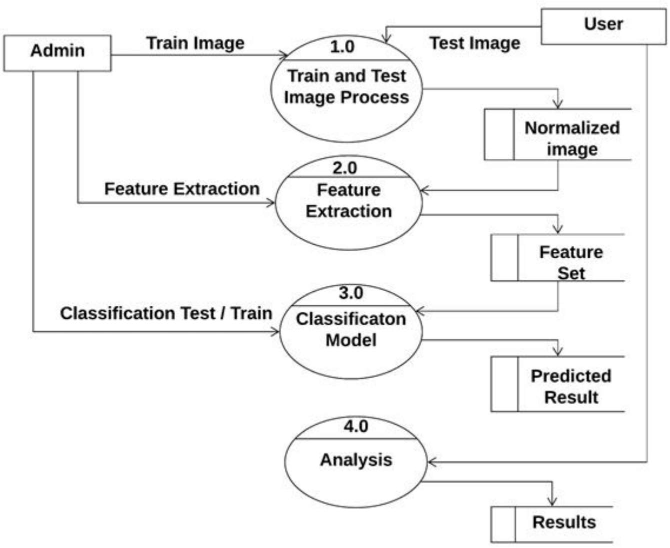

Fig. 3.DFD1

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, March 2025

Activity Diagram

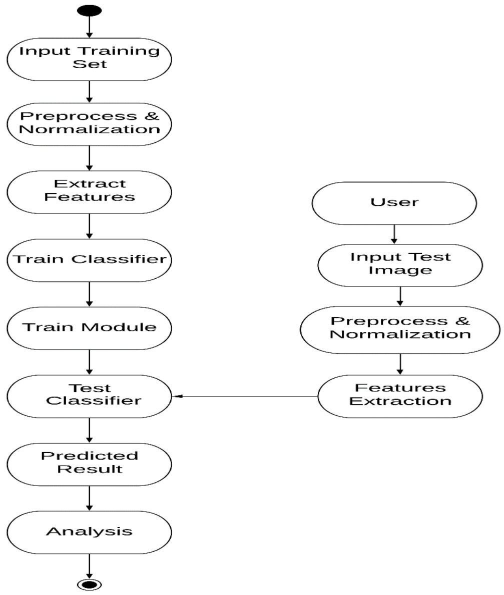

Fig. 4.Activity Diagram

Deployment Diagram

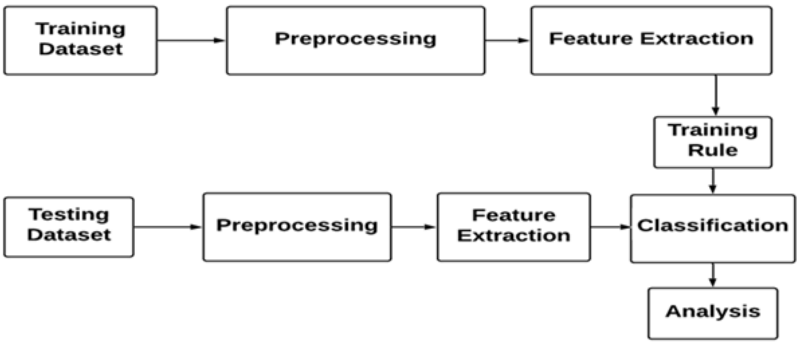

Fig. 5.Deployment Diagram

Sequence Diagram

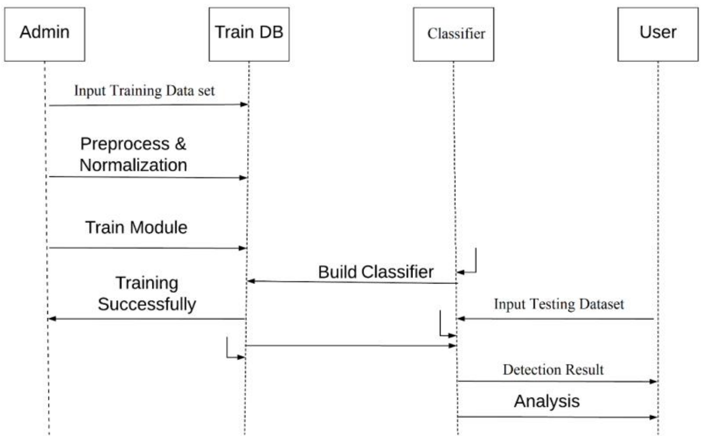

Fig. 6.Sequence Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, March 2025

Use Case Diagram

Fig. 7.Use Case Diagram

V. CONCLUSION

The prototype The proposed system will successfully predict the signs of sign and some common words under different lighting conditions and different speeds. Accurate masking of the images is being done by giving a range of values that could detect human hand dynamically. The proposed system uses CNN for the training and classification of images. For classification and training, more informative features from the images are finely extracted and being used. The proposed system can deal with various deep learning frameworks for predicting the activity of sign. By using a convolutional neural network system can able to provide higher accuracy than traditional machine learning algorithms.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the paper on “Sign Language Recognition with Deep Learning”. We would like to take this opportunity to thank our guide, Prof. Neelam Jadhav, Professor, Department of Computer Engineering Department, Genba Sopanrao Moze College of Engineering, Balewadi, Pune for giving us all the help and guidance we needed. We are grateful to her for her kind support, and valuable suggestions were very helpful.

REFERENCES

[1].M. A. Thakur, P. Budhathoki, S. Upreti, S. Shrestha, and S. Shakya, "Real Time Sign Language Recognition and Speech Generation," J. Innov. Image Process., vol. 2, no. 2, pp. 65-76, 2020, doi: 10.36548/jiip.2020.2.001.

[2]. S. Shrenika and M. MadhuBala, "Sign Language Recognition Using Template Matching Technique," 2020 Int. ConfComput. Sci. Eng. Appl. ICCSEA 2020, pp. 5-9, 2020, doi: 10.1109/ICCSEA49143.2020.9132899.

[3]. L. K. S. Tolentino, R. 0. Serfa Juan, A. C. Thio-ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. 0.Garcia, "Static sign language recognition using deep learning," Int. J. Mach. Learn.Comput., vol. 9, no. 6, pp. 821-827, 2019, doi: 10.18178/ijmlc.2019.9.6.879.

[4]. doi: 10.18178/ijmlc.2019.9.6.879. S. Puri, M. Sinha, S. Golaya, and A. K. Dubey, "Indian sign language recognition using python," in Emerging Technologies in Data Mining and Information Security, Springer, 2021, pp. 427-434.

[5]. S. Sharma and K. Kumar, "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks," Multimed. Tools Appl., vol. 80, no. 17, pp. 26319-26331, 2021, doi: 10.1007/s11042-021-10768 5.

[6]. M. Quinn and J. I. Olszewska, "British sign language recognition in the wild based on multi-class SVM," in 2019 federated conference on computer science and information systems (FedCSIS), IEEE, 2019, pp. 81-86.

[7]. K. Aashritha and V. M. Manikandan, "Assistive Technology for Blind and Deaf People: A Case Study," in Machine Vision and Augmented Intelligence: Select Proceedings ofMAI 2022, Springer, 2023, pp. 539-551

[8]. E. Sammari and A. Naceur, "The Effectiveness of an Intervention Program for the Development of Social and Emotional Capacities of Children with a Hearing impairment," Psychology, vol. 13, no. 7, pp. 1025-1062, 2022..