AI-Powered Indian Sign Language Detection System

Ranjith Kumar G
Computer science and Engineering
Paavai Engineering College
Namakkal, Indian
ranjithkumar9860@gmail.com

Kalaiselvan B
Computer Science and Engineering
Paavai Engineering College
Namakkal, India
kalaiselvan123888@gmail.com

Nandhaakash M
Computer Science and Engineering
Paavai Engineering College
Namakkal, India
nandhaakashm@gmail.com

***Abstract*—Sign language (SL) is a vital mode of communication, bridging the gap between the hearing impaired and hearing communities. However, SL, despite its paramount importance, has received relatively limited attention from researchers. Its unique structural characteristics, distinct from those of natural languages, present novel challenges that require innovative solutions. Remarkable technological advancements in artificial intelligence (AI) and machine learning offer promising avenues for automated sign language translation systems (SLTS). This review study addresses the crucial need for a comprehensive synthesis of existing research by systematically examining and evaluating the progress made in SLTS. By analyse 58 research papers, with a particular emphasis on the most frequently cited papers from each year up to 2023, we shed light on the field's current state, identifying key advancements and challenges. This review followed a systematic approach based on clear guidelines. The methodology involved defining research questions, formulating queries, selecting studies based on clear criteria, and extracting pertinent information to address the research objectives. This review found that deep learning techniques, such as convolutional and recurrent neural networks, have shown high accuracy in sign language recognition, and their performance in recognizing the variety of signs has steadily improved over time. Additionally, integrating non-manual features has proven pivotal in enhancing recognition accuracy. Future research should refine advanced deep learning models and integrate non-manual features to improve system accuracy and applicability. These ongoing advancements hold the potential to revolutionize communication and break down barriers for individuals who rely on sign language as their primary mode of communication.**

Keywords-- Artificial Intelligence, Machine Learning Deep Learning , Convolutional Neural Networks (CNN), ResNet-50 (a popular CNN architecture used for gesture recognition

s

I. INTRODUCTION

Sign language is a vital form of communication for the deaf and hard-of-hearing communities, relying on the use of hand movements, facial expressions, and body language to convey meaning. However, the ability to communicate through sign language is not widespread among those who are not part of these communities, which can create significant communication barriers in various settings, including education, healthcare, public services, and everyday social interactions. With advancements in artificial intelligence, computer vision, and machine learning, there has been a growing interest in creating systems that can recognize and interpret sign language gestures automatically. Sign Language Recognition (SLR) systems aim to facilitate real-time translation between signers and non-signers, providing an essential tool for communication that promotes inclusivity. Problem Statement Traditional methods of sign language communication, such as interpreters or text-based alternatives, often have limitations. Interpreters may not always be available, and text-based communication may lose the nuances of the gestures and facial expressions integral to sign language. To overcome these challenges, automated systems capable of interpreting sign language offer a promising solution. The complexity of building an accurate SLR system stems from the unique features of sign languages. Unlike spoken languages that rely solely on auditory signals, sign languages use multiple visual cues, including hand gestures, body posture, facial expressions, and the spatial relationships between these elements. Moreover, different countries and regions have their own distinct sign languages, such as American Sign Language (ASL), British Sign Language (BSL), and others. This variety adds another layer of complexity to the development of a generalized recognition system. Goals of the Project The primary goal of this project is to design and implement a robust, real-time sign language recognition system capable of recognizing both static and dynamic gestures. Static gestures are often used for individual letters or numbers, while dynamic gestures represent words, phrases, or complete sentences. Our system is designed to bridge this communication gap by converting sign language gestures into spoken language or text.

II. LITERATURE REVIEW

State-of-the-art techniques are center after utilizing deep learning models to improve good accuracy and execution time. CNNs have indicated huge improvements in visual object recognition, natural language processing, scene labelling, medical image processing, and so on. Despite these accomplishments, there is little work on applying CNNs to video classification. This is halfway because of the trouble in adjusting the CNNs to join both spatial and fleeting data. A model using exceptional hardware components such as a depth camera has been used to get the data on the depth variation in the image to locate an extra component for correlation and then build up a CNN for getting the results but still has low accuracy. An innovative technique that does not need a pre-trained model for executing the system was created using a capsule network and versatile pooling.

Furthermore, it was revealed that lowering the layers of CNN, which employs a greedy way to do so, and developing a deep belief network produced superior outcomes compared to other fundamental methodologies. Feature extraction using scale-invariant feature transform (SIFT) and classification using Neural Networks were developed to obtain the ideal results. In one of the methods, the images were changed into an RGB conspire, the data was developed utilizing the movement depth channel lastly using 3D recurrent convolutional neural networks (3DRCNN) to build up a working system_where Canny edge detection oriented FAST and Rotated BRIEF (ORB) has been used. ORB feature detection technique and K-means clustering algorithm used to create the bag of feature model for all descriptors is described, but the plain background, easy to detect edges are totally dependent on edges; if the edges give wrong info, the model may fall accuracy and become the main problem to solve.

III. METHOGOLOGY

Figure A: *pre image processing sign language detection in alphabetic order.*

To accomplish this, the project employs computer vision and machine learning techniques that allow the system to recognize complex gestures and translate them effectively. Key components of the system include: 1. Data Collection and Preprocessing:

1. A dataset of sign language gestures is collected, including both static signs (e.g., alphabet or numbers) and dynamic signs (e.g., words or sentences). These datasets may include thousands of video frames or images of hand gestures performed by different individuals to ensure diversity in the training process. Preprocessing techniques are applied to standardize input data, addressing variations in lighting conditions, backgrounds, hand shapes, and signer appearances. This may include image augmentation techniques to enhance dataset quality and increase system robustness.

2. Feature Extraction: Convolutional Neural Networks (CNNs) are utilized for extracting key features from images or video frames. CNNs are well-suited for image recognition tasks due to their ability to capture spatial hierarchies and detect patterns such as edges, shapes, and motion in gesture data. ¬ For dynamic gesture recognition, Recurrent Neural Networks (RNNs) or their variants like Long Short-Term Memory (LSTM) networks are used. These networks specialize in processing sequential data, such as a sequence of hand movements in sign language, to recognize temporal dependencies and the evolution of gestures over time.

3. Gesture Classification and Recognition: The processed features are fed into machine learning classifiers that map recognized gestures to their corresponding labels, such as letters, words, or phrases. These classifiers are trained on the dataset to improve accuracy in identifying and interpreting gestures. For improved accuracy in real-world applications, techniques such as Transfer Learning can be employed, where pre-trained models are finetuned on a specific sign language dataset, reducing the need for extensive training data and computational resources.

4. Real-Time Processing and User Interface: To achieve real-time performance, the system is optimized for fast image capture, processing, and gesture recognition. This allows users to communicate through sign language without significant delays.

IV. ALGORITHM

CNN + RNN for Dynamic Gesture Recognition For dynamic gestures (i.e., gestures that involve sequences of movements over time), the system combines CNNs with **Recurrent Neural Networks** (RNNs) or their variants, such as **Long Short-Term Memory** (LSTM) networks. This architecture allows the system to model both spatial and temporal dependencies in sign language. CNN (Front-end): Extracts spatial features from each frame of a video. RNN/LSTM (Back-end): Processes the sequence of frames, capturing the temporal relationship between them.

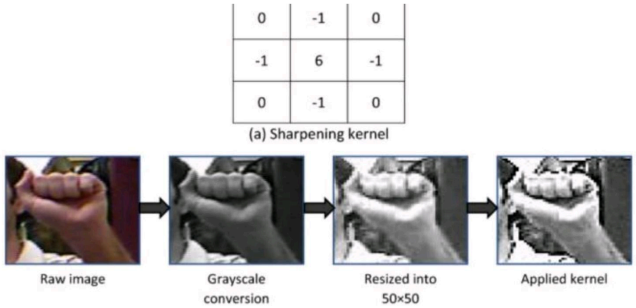

Figure B: Hand landmarks detection and extraction of 21 coordinate.

V. FLOWCHART

The whole work is divided into two main parts, one is the raw image processing, and another one is the hand landmarks extraction. After both individual processing had been completed, a custom light weight simple multi-headed CNN model was built to train both data. Before processing

through a fully connected layer for classification, we merged both channel's features so that the model could choose between the best weights.

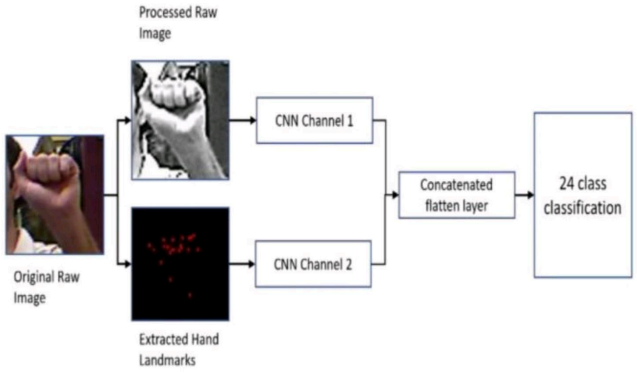

Figure C: Flow diagram of working procedure.

A.Model Building:

In this research, we have used multi-headed CNN, meaning our model has two input data channels. Before this, we trained processed images and hand landmarks with two separate models to compare. Google's model is not best for "in the wild" situations, so we needed original images to complement the low faults in Google's model. In the first head of the model, we have used the processed images as input and hand landmarks data as the second head's input. Two-dimensional Convolutional layers with filter size 50, 25, kernel (3, 3) with , strides 1; Max Pooling 2D with pool size (2, 2), batch normalization, and Dropout layer has been used in the hand landmarks training side. Besides, the 2D Convolutional layer with filter size 32, 64, 128, 512, kernel (3, 3) with Max Pooling 2D with pool size (2, 2); batch normalization and dropout layer has been used in the image training side. After both flatten layers, two heads are concatenated and go through a dense, dropout layer. Finally, the output dense layer has 24 units with Soft max activation.

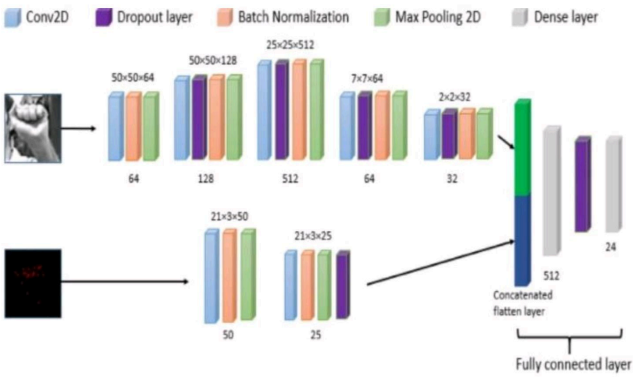

Figure D: *Proposed multi-headed CNN architecture. Bottom values are the number o require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.*

ACKNOWLEDGMENT

WE WOULD LIKE TO EXPRESS OUR SINCERE GRATITUDE TO OUR STAFF FOR THEIR CONTINUOUS GUIDANCE, TECHNICAL SUPPORT, AND VALUABLE SUGGESTIONS THROUGHOUT THE DEVELOPMENT OF THIS PROJECT. WE ALSO ACKNOWLEDGE THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, ANNA UNIVERSITY, FOR PROVIDING LABORATORY FACILITIES AND RESOURCES. FINALLY, WE THANK OUR PEERS AND FAMILY MEMBERS FOR THEIR ENCOURAGEMENT DURING THE PREPARATION OF THIS RESEARCH WORK

REFERENCES

[1] Chaudhary, A., et al., "ISL Alphabet Recognition Using Deep Learning," IJERT, 2019.

[2] Guo, L., et al., "MediaPipe and LSTM Based Hybrid Gesture Detection," Pattern Recognition Letters, 2022.

[3] Hugging Face NLP Transformers Library, 2024.

[4] OpenCV & Media Pipe Documentation, Google Research, 2023

[5] Ramesh, A., et al., "Android-Based Real-Time Sign Language Recognition," IEEE Xplore, 2021.

[6] Saini, R., & Rajesh, M., "Vision-Based ISL Gesture Recognition Using CNN," IJCSIT, 2020.

[7] TensorFlow Official Guide, Google AI, 2024.

[8] Tran, D. et al., "3D Convolutional Networks for Video-Based Recognition," IEEE CVPR, 2018.

[9] Zhou, Z., et al., "Vision Transformers for Sign Language Recognition," Elsevier Neurocomputing, 2021.