SIGN LANGUAGE CONVERTER USING ML TRAINED MODELS

line 1: Nagaram Kridey   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: Sanda Manoj Kumar  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: Tammireddy Jagadeesh Narasimha Sai   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract— In the past couple of years, AIML has transformed such diverse fields as natural language processing and human-computer interaction, and indeed turns out to be a major phenomenon that is changing the world around us. This paper presents an innovative approach towards converting live feed of sign language into English text using advanced techniques of AIML. The system propounds the utilization of computer vision algorithms for detecting and interpreting sign language gestures in real time so as to facilitate smooth communication between sign language users and non-sign language speakers. The system shall process input from a camera capturing live sign language, utilizing deep learning models and techniques that analyze images. Such types of models were trained on a diverse dataset of sign language gestures; hence, the models have better precision in gesture recognition and more robust performance across different dialects of sign language. Through extensive tests, the system shows high accuracy and efficiency in processing the information presented, thus reducing significantly the latency between sign production and textual output.*  
  
.

Keywords—Machine Learning, Artificial Intelligence, Image Training, Models, Sign Language, Datasets.

# Introduction

Communication is one of the most basic human rights yet millions of people are excluded because of their language. It plays the big role in the life of the deaf and hard-of-hearing minority; however, this form of communication is underrepresented in mainstream discourse. As the development of AIML gains more ground, this can fill the gap in communication because sign language can be converted to easily understandable formats such as English texts.  
  
It presents a novel system using AIML technologies to translate live sign language feeds into written text in real-time based on the methodologies of computer vision techniques, capturing the interpretation of gestures through translation into textual output. Based on this fact, the research highlights the generation of a deep learning model that is trained on an overall comprehensive dataset of sign language gestures to ensure that translation recognition between the two languages is atleast above an average level of accuracy in order for it to be effective for usage in the society as well as daily life.  
  
The proposed solution will make the access to the sign language users better as well as facilitate stronger communicative inclusivity. With technology continuously advancing, a huge question arises about the need for the discovery of new applications useful in the life of marginalized communities in general and capabilities to overcome social challenges. This paper aims to contribute to the ongoing discussion on technology and accessibility by proposing a realistic approach toward making the communication lives of people more diverse and inclusive.

# Current Existing Studies

## Sign Language Recognition using Deep Learning.

The work by Ghorbani et al. focuses on the classification of static sign language gestures with CNNs. Their system was built to identify gestures from many sign languages, like predefined alphabets and common words. However their system was able to achieve reasonable accuracy for isolated gestures but failed in its attempt for dynamic and continuous sign sequences. Real conversation settings present gestures that are fluid and highly dependent on context; thus, the system failed to perfectly capture this element. This research was highly important as it set the foundation for gesture recognition based on CNNs but was needed to improve temporal analysis in continuous sign language translation.

## ASL Recognition System

Kim et al. presented the system that combined Convolutional Neural Networks for spatial feature extraction and LSTM networks to model the temporal sequences in continuous signing. The research was on American Sign Language and provides an extension to previously designed models that could only recognize isolated signs. The architectural combination went well with continuous gestures recognition with higher accuracy. However, its high computational demands meant that powerful hardware was required to make this system extendible into real-time applications. The contribution of this study was crucial for improving the accuracy of the dynamic sign language recognition systems.

## “Sign2Text Project “ By Zhang et al, 2022

Zhang et al presented a new approach toward capturing 3D sign language gestures by combining depth sensors with machine learning. The data about the hand movement details were more detailed with the depth sensors, which improved the capability of the system to recognize both complex and dynamic gestures in real time. Compared to 2D camera-based systems, the accuracy was even higher while trying to detect such fine details in terms of hand motion and finger positions. This, however makes the solution less scalable and accessible for wide usage since it calls for special hardware such as depth cameras. Although this project has made great advancement in sensor data processing toward better gesture recognition, improvements still remain to be made.

# Limitations Of Systems

## Static Gesture Limitation

This makes Ghorbani et al's system limited to the recognition of predetermined static gestures. Actually, this system will not capture the fluidity inherent in sign language in any way since gestures do not hold still between signs but keep on changing continuously. Therefore, the system cannot use it in the scenario of continuous and context-dependent situation of conversational real-life contexts of signing interpretation.

## Lack of Temporal Modeling

The system proposed by Ghorbani et al. cannot model temporal relations between gestures, which is what is needed to interpret continuous sign language. Sign language intrinsically has to do with the stream and sequence of signs rather than the signs themselves. In failing to model time relations between gestures, this system could only interpret isolated signs. Isolated sign interpretation does not effectively represent a user's desired realworld functionality. Thus, the system cannot interpret context-dependent gestures; it may not be effective at conversational use when only time dependencies are in consideration.

## Inflexibility Across Dialects

Ghorbani et al.'s model is inflexible when it comes to recognizing different sign language dialects. Sign languages vary across regions and cultures, with different signs being used for the same concepts. The system’s reliance on a specific, predefined set of gestures means it struggles to adapt to these variations. This lack of flexibility reduces the model’s applicability across diverse user groups. Therefore, while effective for a particular sign language dataset, it is less practical for users communicating in different regional or cultural sign language dialects.

## High Computational Demand

Even though Kim et al's system improves the accuracy of continuous gesture recognition, the hybrid CNN-LSTM architecture has a great disadvantage of utilizing enormous computational power. High-end hardware, such as GPUs, is requisite for processing spatial and temporal information complexities. Such a need for powerful computational resources confines the system's deployability in real-time applications, primarily environments that may not be equipped with such hardware. It therefore renders the system problematic to scale for daily practical use.

## Limited Real-Time Capability

Although Kim et al.'s system enhances the accuracy of continuous sign language recognition, its capability is limited to real-time performance. CNN-LSTM architecture is computationally expensive and loads with a significant processing delay between gesture input and text output, making the system less effective for situations in which real-time interaction is a pre-requisite-for instance, live conversations and public service applications. As such, though the system may be accurate, it does not come across as ideal for situations demanding instant and fluid sign language translation.

## Narrow Focus on ASL

Kim et al. based their system on American Sign Language (ASL). So, this is not much good for those who might utilize British Sign Language, regional dialects, or other sign languages. This actually restricts the applicability of the system in other parts of the world due to the use of different local sign languages being utilized across different localities and cultures. Although the system works very well in recognizing ASL, it doesn't give support in other sign languages, which decreases the accessibility of the system to a larger audience, making the system less useful for a multilingual and varied community of sign languages.

## Specialized Hardware Requirement

Zhang et al. designed a system that utilizes depth sensors: the recognition capabilities for 3D gestures have improved remarkably but use hardware so expensive and specialized that access and scalability of the system are severely hindered. Depth sensors are neither available nor affordable for most real-world settings in which one would like to deploy such a system. The system correctly detects complex gestures, but it utilizes specialized equipment and therefore is less likely to attain widespread application, especially in low-resource environments.

## High Hardware Costs

One of the main drawbacks of the Zhang et al. proposed system is that it relies on depth sensors for gesture detection, which are more expensive than any other gesture recognition systems. Although they have high accuracy since they can capture 3D gestures, they tend to be very expensive and not very accessible. This makes it harder to deploy the proposed system, especially in large-scale or low-resource settings. It is true that this system increases the accuracy of the results, but its high cost, especially of the hardware, makes it inaccessible to most, especially those who cannot afford the latest technological gadgets from more advanced places in the world.

## Limited Standard Camera Usability

Zhang et al.'s system is designed to operate based on depth sensors, which means that it cannot be applied to operating fields that merely make use of conventional 2D cameras. Many possible users lack access to the hardware, which makes the system have drastically fewer applications in real life. The majority of natural environments make use of standard cameras; thus, this system's dependency on depth sensors limits its scalability and daily applicability. Although the system is more accurate in recognizing complex gestures, it is less accessible for use in normal settings, hence for real-time application in widespread use.

| Literature Review | | |
| --- | --- | --- |
| Project Reference | Methodology | Key Features |
|  | CNN for gesture recognition | Focus on predefined gestures, lacks real-time ability |
|  | CNN + LSTM hybrid model | Improved accuracy for continuous signing |
|  | Depth sensors + machine learning | 3D gesture recognition, robust for dynamic gestures |

# Proposed Methodology

In this project, a real-time system will be created, which can take live feeds of sign language and translates it into the English text with the application of Artificial Intelligence and Machine Learning. It integrates computer vision, machine learning models, and natural language processing to seamlessly translate information with good accuracy. The methodology that will be used to achieve this would be identified in the following phases: Data Collection, Preprocessing, Model Training, Real-Time Gesture Recognition Pipeline, and Evaluation.

## Data Collection

**It uses a vast collection of sign language gestures. The approach is reliant on this huge dataset, which it learns from. Some of the sourced datasets include:**

### RWTH-PHOENIX-Weather 2014: It's one of the very widely used German sign language recognition datasets.

### American Sign Language Lexicon Video Dataset (ASLLVD): It consists of thousands of signs for American Sign Language.

### Custom data collection: e could add more depth into it by collecting our own dataset by using a webcam and getting the concerned persons to sign different signs. This data will encompass not only hand gestures but also facial expressions and body movements that are integral parts of sign language.

## Data Preprocessing

Once the dataset is ready, preprocessing becomes the most crucial step in cleaning and optimizing data before its usage in model training. There are several key steps:

#### Frame Extraction: The video feeds of sign languages would be broken down into discrete frames to be analyzed. Since sign language is fundamentally made of continuous gestures, extracting meaningful frames with respect to the temporal context is important.

#### Keypoint Detection: The keypoint detection algorithm will be MediaPipe for hand keypoints, facial landmarks, and body postures. The algorithms capture all the important joint and hand movements which are actually required for identifying a gesture correctly.

#### Data Augmentation: The dataset will be augmented with rotations, scaling, flipping, and lighting adjustments to simulate real-world variations in images that may make the model more robust.

#### Normalization: all gesture data, such as hand movement and facial expression, is normalized so that the size of input going to each model turns out to be uniform.

#### Frame Extraction: The video feeds of sign languages would be broken down into discrete frames to be analyzed. Since sign language is fundamentally made of continuous gestures, extracting meaningful frames with respect to the temporal context is important.

## Model Training

The core of the project is the machine learning model that would process the gestures and translate them into text in English. These include:

#### Convolutional Neural Networks (CNN): The video feeds of sign languages would be broken down into discrete frames to be analyzed. Since sign language is fundamentally made of continuous gestures, extracting meaningful frames with respect to the temporal context is important.

#### Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM): Since sign language gestures are continuous in time, LSTMs turn out to be perfectly suited for learning temporal relationships between gestures. This renders a system that can interpret gestures in terms of their sequences and keeps the context alive.

#### Hybrid Model (CNN + LSTM): A hybrid CNN-LSTM model can incorporate both spatial features of each frame and the temporal dependencies between gestures to recognize in real time.

## Real-Time Gesture Recognition Pipeline

## The developed system is to be real-time thus converting live input from a webcam into English text with very minimal latency. Among the pipeline components are:

#### Video Input: The system captures live video through a standard webcam, providing real-time feeds for processing.

#### Preprocessing in Real-Time: Keypoints on hands, facial expressions, and body postures are going to be captured frame by frame by using OpenPose or MediaPipe, thus reducing heavy image processing.

#### Gesture Recognition: The processed frames will pass through the CNN-LSTM model for real-time gesture recognition.

#### Text Generation: The recognized gestures will be mapped to corresponding English words or phrases. An AI-based text processing module manages the word order and context to create grammatically correct sentences from the recognized gestures.

#### User Interface (UI): A user friendly UI will be displaying the text output with real time and how the user will interact and see the translations how they occur.

## Challenges and Future Enhancements

#### Multiple Sign Languages: The model currently supports ASL and BSL, though it may be expanded into the near future to incorporate other regional sign languages in the country.

#### Handling Ambiguities: There may be several instances of various kinds of ambiguity in the signs. Contextual meaning of the signs may vary. For the future versions, this system can be coupled with contextual AI, which will be able to perceive the flow of the conversation more clearly.

#### Incorporating Gesture Speed: Users may sign at varying speeds. The system needs to be improved for the recognition and accommodation of such variations.

# Conclusion

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
8. K. Eves and J. Valasek, “Adaptive control for singularly perturbed systems examples,” Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
9. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>
10. S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC
11. “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2