JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GESTURE SPEAK

Jia Ann James
Dept. of Computer Applications
Saintgits College of Engineering
(Autonomous)
Kottayam, Kerala

Rehan Nelson Thomas
Dept. of Computer Applications
Saintgits College of Engineering
(Autonomous)
Kottayam, Kerala

Akshay P Kumar
Dept. of Computer Applications
Saintgits College of Engineering
(Autonomous)
Kottayam, Kerala

Sreelakshmi P S
Dept. of Computer Applications
Saintgits College of Engineering
(Autonomous)
Kottayam, Kerala

Ms Rani Saritha R
Assistant Professor
Dept. of Computer Applications
Saintgits College of Engineering
(Autonomous)
Kottayam, Kerala

***Abstract*— Sign language, as a fundamental means of communication, holds immense significance for individuals who face challenges in speaking or hearing, providing them with a powerful tool to convey their thoughts and sentiments effectively. Its role is indispensable in bridging communication gaps within the physically challenged community, fostering connections and understanding. This project's primary aim is to develop an advanced neural network capable of accurately interpreting and categorizing the intricate hand gestures inherent in sign language. Through comprehensive analysis of images depicting signing gestures, the neural network will discern and classify the corresponding alphabet representations, enabling seamless translation into written English and further into Malayalam. The creation of such a sophisticated translator has the potential to revolutionize communication dynamics for individuals with hearing or speech impairments, empowering them to participate more actively in various social interactions and engagements. Furthermore, the successful execution of this project stands to significantly enhance inclusivity and accessibility for the deaf and mute community, thereby promoting a more equitable and inclusive society. It represents a significant stride forward in leveraging technology for societal welfare, underscoring the importance of embracing innovation to address pressing social challenges. By prioritizing the development of inclusive technologies, this project advocates for equal opportunities and rights for all individuals, regardless of their physical abilities or disabilities. In summary, the implementation of this project not only addresses a critical need within the community but also embodies a broader commitment to social responsibility and technological innovation for the greater good.**

***Keywords*— sign language, gesture, Convolutional Neural Networks, deep learning, OpenCV, media pipe**

I. INTRODUCTION

Sign language serves as a visual-gestural mode of communication employed by individuals who are deaf or hard of hearing. It utilizes three-dimensional spaces and hand movements (as well as other body parts) to convey meanings, possessing its distinct vocabulary and syntax that markedly differs from spoken or written languages. In spoken languages, oratory faculties generate sounds corresponding to specific words and grammatical combinations to convey information, with the auditory faculties then receiving and processing these oral elements. In contrast, sign language employs visual faculties, presenting a divergence from spoken language modalities. Just like spoken language, sign language adheres to intricate grammar rules for generating comprehensive messages.

A sign language recognition system involves an efficient and accurate mechanism for converting language into text or speech. This process utilizes computerized digital image processing and various classification methods to recognize the flow of alphabets and interpret sign language words and phrases. Sign language information can be conveyed through gestures involving hands, head positioning, and other body parts. The key components in a gesture recognition system encompass gesture modelling, gesture analysis, gesture recognition, and applications based on gestures.

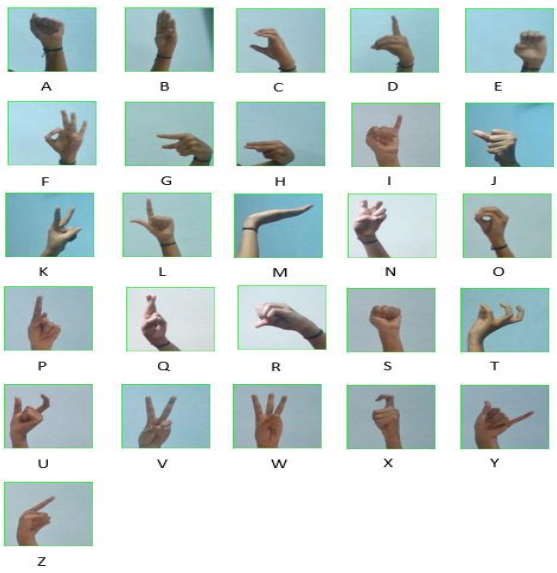
II PREVIOUS STUDY

The need for communication has been intrinsic throughout human history, dating back to early ages. Language serves a crucial role in facilitating communication, allowing the expression of emotions, feelings, thoughts, and ideas. This expression can take the form of verbal or non-verbal communication, encompassing speech, symbols, and gestures. While language has greatly enhanced communication, it comes with its challenges. Effective communication often requires a common language to ensure mutual understanding and appropriate responses.

Additionally, individuals with disabilities, whether present from birth or acquired, may face limitations in expressing and sharing their thoughts and ideas. Those who are hearing impaired or unable to speak often resort to gestures and sign language for communication. Efforts to bridge the communication gap between those who use sign language and those who do not typically involve text or translators. However, both approaches have drawbacks; the former can be time-consuming, while the latter compromises privacy. There exists a compelling need to eliminate barriers between speech-impaired individuals and other members of society. In general, individuals with speech or hearing disabilities utilize gestures for communication. However, these gestures may not be universally understood, and conveying even a simple message can be time-consuming. An AI vision-based system capable of converting sign language into text or speech could effectively address the communication challenges faced by differently-abled individuals when interacting with the general population.

III. METHODOLOGY

3.1 Dataset Collection



The hand gesture dataset collected using the MediaPipe library and OpenCV. The process involves capturing gestures for each alphabet from A to Z. It involves utilizing a webcam to record hand movements in real-time. For each alphabet, 150 samples are captured. The dataset is saved in a directory named 'hand_dataset', with separate pickle files for each alphabet containing the corresponding hand gesture data.

The code initializes the MediaPipe Hands module and uses it to detect and track hand landmarks in each frame. The webcam feed is processed frame by frame, and for frames where hand landmarks are detected, the coordinates of these landmarks are extracted. The extracted data, consisting of landmarks and the associated alphabet label, is then appended to a list. This process is repeated for the specified number of samples.

The captured gesture data is serialized and stored using the pickle library. Each alphabet's data is saved in a separate pickle file within the 'hand_dataset' directory. The data collection process is repeated for each alphabet from A to Z.

3.2 Data Preprocessing

The data pre-processing step involves organizing and structuring the collected gesture data for further use. The stored pickle files, each containing a list of tuples representing landmarks and associated alphabet labels, can be loaded for subsequent analysis or model training. The landmarks are extracted from the dataset for each alphabet, and these datasets can be used for training machine learning models, particularly for tasks related to hand gesture recognition. Pre-processing may include normalization, scaling, or any other necessary transformations to prepare the data for model input. The pre-processed data is then ready for training a deep learning model.

3.3 Model Training

The model is trained using Convolutional Neural Network (CNN). The dataset which is collected is loaded from separate pickle files, one for each alphabet, and combined into a single dataset. After shuffling the data, it is split into features (X) and labels (y), where X represents the hand landmarks, and y is the corresponding alphabet labels converted to integer format.

The neural network model is constructed using the TensorFlow and Keras libraries. The model architecture consists of a Flatten layer, assuming 21 hand landmarks with (x, y) coordinates, followed by two Dense layers with rectified linear unit (ReLU) activation functions. The final Dense layer has 26 units with a softmax activation function, representing the 26 classes (alphabets A to Z). The model is then compiled using the Adam optimizer and sparse categorical crossentropy as the loss function. Training is performed using the fit method, where the model is trained on the training set (X_train, y_train) for 15 epochs, with validation on the

testing set (X_test, y_test). After training, the model's performance is evaluated using the testing set, and the test accuracy is printed. Finally, the trained model is saved as 'hand_gesture_model.h5'.

In summary, CNN model is trained for hand gesture recognition, with a focus on recognizing alphabets A to Z based on hand landmarks. The model is constructed, trained, evaluated, and saved for potential future use in recognizing hand gestures in real-time scenarios.

3.4 Model Evaluation

The neural network is trained on the training set and validated on the testing set for 15 epochs. After training, the model predicts on the test set, and the predicted probabilities are converted to class labels. The classification report, including metrics such as precision, recall, F1 score, and support, is then displayed, providing an assessment of the model's ability to correctly classify sign language gestures for the alphabets A to Z.

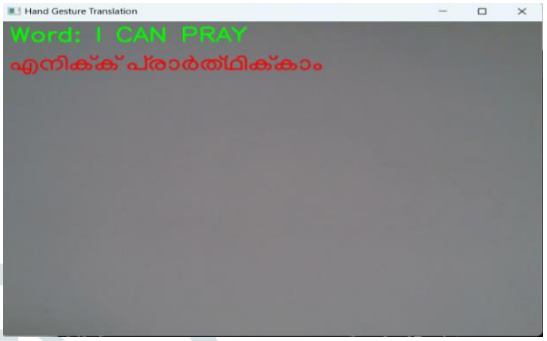
	precision	recall	f1-score	support
A	0.9027	0.9126	0.9076	183
B	0.8498	0.9427	0.8938	192
C	0.9946	0.9786	0.9865	187
D	0.9714	0.9392	0.9551	181
E	0.9227	0.9176	0.9201	182
F	0.9548	0.9694	0.9620	196
G	0.9319	0.9570	0.9443	186
H	0.9843	0.9641	0.9741	195
I	0.9368	0.8717	0.9030	187
J	0.9565	0.9724	0.9644	181
K	0.9119	0.9312	0.9215	189
L	0.9752	0.9336	0.9540	211
M	0.9462	0.9362	0.9412	188
N	0.9096	0.9500	0.9293	180
O	0.9484	0.9528	0.9506	212
P	0.9673	0.9628	0.9650	215
Q	0.9590	0.9639	0.9614	194
R	0.8808	0.8673	0.8740	196
S	0.8667	0.8366	0.8514	202
T	0.9375	0.9116	0.9244	181
U	0.7879	0.8168	0.8021	191
V	0.7849	0.7941	0.7895	170
W	0.9176	0.8520	0.8836	196
X	0.8571	0.8867	0.8717	203
Y	0.9215	0.9215	0.9215	191
Z	0.9667	0.9613	0.9640	181
accuracy			0.9243	5547
macro avg	0.9245	0.9240	0.9240	5547
weighted avg	0.9250	0.9243	0.9244	5547

IV. RESULTS AND ANALYSIS

The system captures video frames from the webcam using OpenCV. MediaPipe is employed to detect and visualize hand landmarks on each frame, drawing connections between them. These hand landmarks are then extracted for further processing. The trained model is used to predict the alphabet corresponding to the hand gesture. The predicted alphabet is displayed on the video feed in real-time.

The system also includes functionality for building words through a sequence of detected gestures. Users can form words by pressing the 's' key to append the predicted alphabet, the spacebar for adding a space, and the 'b' key for backspacing to correct the input. The formed word is displayed on the video feed.

Translation to Malayalam is integrated into the system, allowing users to press the 't' key to translate the formed word. The translated text is displayed on the video feed below the word. The system provides options to clear the translated text ('c' key) and exit the application ('q' key).



V. CONCLUSION

The presented system is a real-time hand gesture translation tool that utilizes computer vision and machine learning techniques. It employs a webcam to capture video frames, employing the MediaPipe library to detect and recognize hand gestures. A pre-trained neural network model is loaded to predict the corresponding alphabet of the detected hand gesture, providing real-time feedback on the video feed. Users can dynamically build words by appending predicted alphabets, inserting spaces, and correcting mistakes. The system also integrates translation functionality, allowing users to translate the formed word into Malayalam with a keypress. The translated text is displayed beneath the formed word on the video feed. The system enhances communication accessibility, particularly for individuals with hearing impairments, by providing an interactive and visually intuitive means of forming words through hand gestures and translating them in real-time.

VII.REFERENCES

[1]. Anna Deza, Danial Hasan (2018). MIE324 Final Report: Sign Language Recognition

[2]. K. Mahimanvitha, Dr. M. Arathi2. (2023). Real Time Sign Language Recognition Using Deep Learning

[3]. Ashok Kumar Sahoo, Gouri Sankar Mishra, Kiran Kumar Ravulakollu (2014). Sign Language Recognition: State of the art

[4]. I.A. Adeyanju, O.O. Bello, M.A. Adegboye (2021). Machine learning methods for sign language recognition: A critical review and analysis

[5] Danielle Bragg, Tessa Verhoef, Christian Vogler, Meredith Ringel Morris (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective