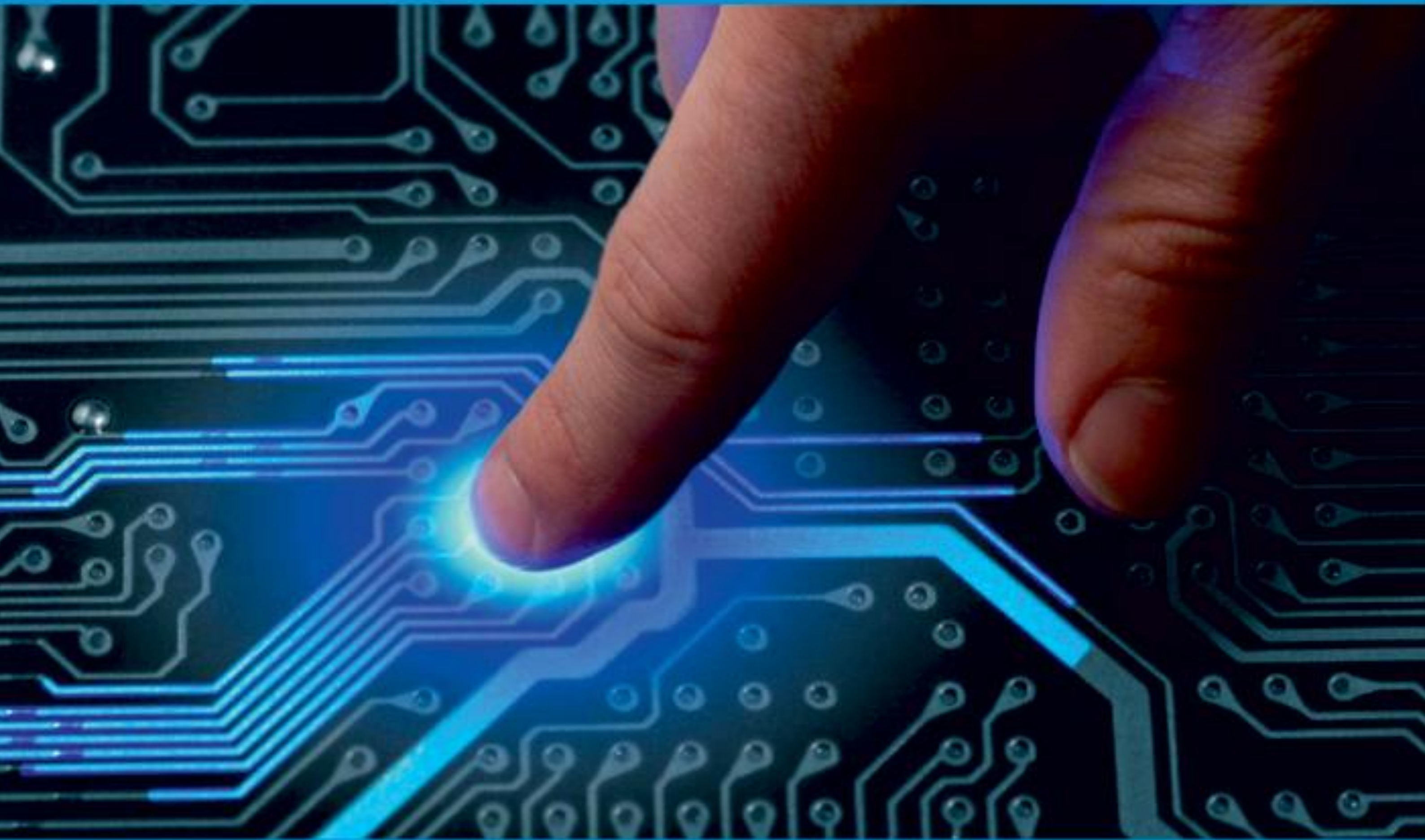
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204257 |

Recognition and Transfer of Sign Language into Text

Nirmal Bhivsane[1], Sayali Jadhav[2], Shreyas Chinchole[3], Nishant Maheshwari[4]

UG Student, Department of Computer Engineering, SSBT's College of Engineering & Technology Jalgaon, Maharashtra, India[1,2,3,4]

ABSTRACT: Sign language is the only tool of communication for the person who is not able to speak and hear anything. Therefore, Hand gestures are one of the strategies used in Sign Language for non-verbal communication. It is commonly used by the deaf and feeble-minded with hearing or speech impairments to communicate with each other or with ordinary people. A real time Sign language recognition system is used to recognize hand gestures. Deaf and dumb people rely on Sign language interpreters for communication. To recognize the sign language, the features will be compared with testing database. Sign Language Recognition System will be implemented by using Machine Learning and Natural Language Processing to enable and efficient sign language interpretation.

KEYWORDS: Sign language, Hand gestures, Real-time Machine Learning, Non-verbal communication, Efficient.

I. INTRODUCTION

Sign language is the most natural and expressive method for those who are dumb and deaf. People who are not dumb or deaf never try to learn sign language for communicating with them. This leads to isolation for those dumb and deaf people. But if the computer can be programmed built such a system in such a way that convert sign language into text format, because of which differences between the normal and deaf communities can be reduced. Each alphabet and gesture are represented by both hands in Indian sign language (ISL).

ISL research is becoming increasingly popular among researchers. Deaf and Dumb people rely on sign language interpreters for communications. A real time Sign Language Recognition system was designed and implemented to recognize 26 gestures from the Indian Sign Language by hand gesture recognition system for text generation. The signs are captured by using web cam. These signs are processed for feature extraction using some colour model. The extracted features are compared by using pattern matching algorithm. To calculate the sign recognition, the features are compared with testing database.

Finally, recognized gesture is converted into text. This system provides an opportunity for a deaf-dumb people to communicate with normal people without the need of an interpreter. It is difficult to finding a sign language translator for converting sign language every time and everywhere, but electronic devices interaction system for this can be installed anywhere is possible. Computer vision is one of the emerging frameworks in object detection and is widely used in various aspects of research in artificial intelligence. Sign language is categorized in accordance with regions like Indian, Chinese, American and Arabic. This system introduces efficient and fast techniques for identifying the hand gestures representing sign language meaning. In this system we will extract the Media Pipe Holistic Key points, then build a sign language model using an Action detection powered by LSTM layers. Then Predict Indian sign language and convert them into text.

II. RELATED WORK

Aman et al., in [1], the sign language recognition project employs a webcam to capture intricate hand gestures, crucial for facilitating communication among the deaf community. By meticulously labelling these gestures and leveraging a pre-trained SSD Mobile Net V2 model, the system navigates challenges posed by environmental factors and accurately detects sign boundaries, ensuring robust recognition and interpretation of sign language in real-time. This technological endeavour represents a significant stride towards inclusivity, aiming to empower individuals with hearing impairments by providing them with a reliable tool for effective communication.

Shubham et al., in [2], the text explores an advanced approach to translating American Sign Language (ASL) gestures into text, leveraging the power of deep learning to overcome the constraints of traditional wearable sensors and hand

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204257 |

detection techniques. By adapting a Convolutional Neural Network (CNN) model, specifically a modified version of the renowned VGG16 architecture, this solution promises enhanced accuracy and efficiency in recognizing ASL gestures. This innovative method not only aims to provide a seamless real-time translation of sign language but also significantly improves the interaction experience for the deaf and hard-of-hearing community by making technology more accessible and user-friendly.

Yulius et al., in [3], this research underscores the imperative for leveraging advanced Human-Computer Interaction (HCI) methodologies, particularly emphasizing the integration of robust gesture and speech recognition systems. By prioritizing these technologies, the aim is to establish more natural modes of communication between individuals and computers, fostering intuitive interactions that mimic human-to-human communication patterns. Through the seamless integration of gesture and speech recognition capabilities, HCI endeavours to enhance user experiences across various domains, ranging from accessibility solutions to immersive virtual environments.

Kaustubh et al., in [4], the research endeavours to develop a gesture-to-text translation system through computer vision methodologies, emphasizing key techniques such as skin masking for isolating hand regions, edge detection to delineate gestures, and employing the SURF model for robust feature extraction. By integrating these advanced techniques, the system aims to accurately interpret hand gestures in real-time and translate them into text, facilitating seamless communication for diverse user groups, including the hearing-impaired and non-signing individuals. This innovative approach represents a significant stride towards breaking down communication barriers and fostering greater inclusivity in social interactions.

III. PROPOSED SYSTEM ARCHITECTURE

In the initial work on Media-pipe and holistic model feature extraction for data selection and a LSTM algorithm has been utilized to recognize the sign language. The Indian Sign Language dataset is used to recognition and transfer of Sign language into text for the understanding the language of dumb and deaf people. Proposed system architecture is depicted in Figure 1 with a detailed description of each step.

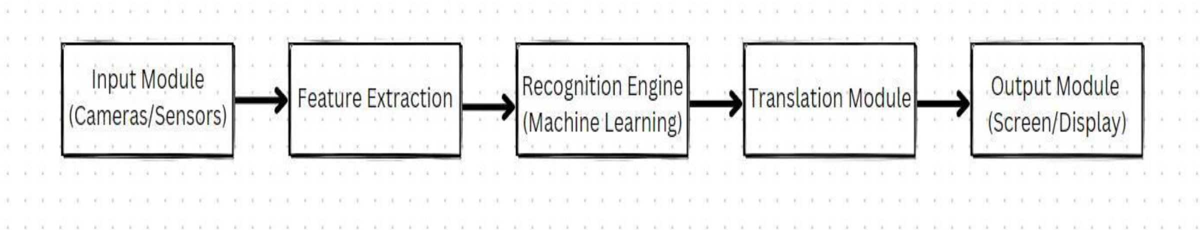


Figure 1: System architecture for Recognition and Transfer of Sign Language into Text

Input Module (Cameras/Sensors):

To develop the Recognition and transfer of sign language into text, a huge and varied dataset of hand gestures representing the Indian Sign Language is required. This dataset is collected with the help of a webcam and the Media Pipe library. The Media Pipe library provides the tools to track the hand gestures in real-time and place key points on the user's hand. The webcam captures the hand gestures and stores them as data samples for the dataset. In this input module phase in the recognition and transfer of sign language into text, with the camera or sensors the system will gather primary detail such as images, videos or sensor reading.

Feature Extraction:

In the feature extraction phase of the Recognition and transfer of sign language into text system, the key points identified by the Media pipe library are used to extract features that describe the shape and movement of the user's hand, these features are then converted into an array of numerical values that can be processed by the system. The array of features extracted from the hand movements is treated as a sequence of data points, each data point represents the position of the hand at a specific point in time. Extraction of features are demonstrated in Figure 2:

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204257 |

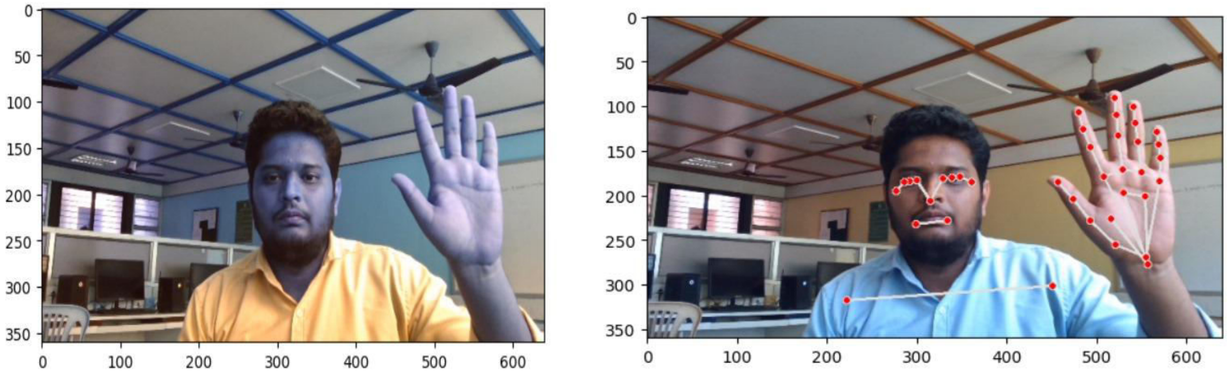


Figure 2: Feature extraction for Recognition and Transfer of Sign Language into Text

Recognition Engine (Machine Learning):
In the Recognition Engine (Machine Learning) phase, the array of features extracted from the hand movements is treated as a sequence of data points. Each data point represents the position of the hand at a specific point in time. a (LSTM) algorithm is used to process the sequence of data points and predict the sign language gesture being performed. The LSTM algorithm is trained using a dataset of labelled sign language gestures and can predict the gesture being performed with high accuracy.

Translation Module:
In the translate module phase, the predicted gesture is verified using a gesture verification algorithm. This algorithm checks whether the predicted gesture matches the expected gesture based on the context of the conversation. It converts these recognitions into something that can be understood by other systems or shown to users easily.
Output Module (Screen/Display):

Lastly, the resulting text message is then displayed to the user or transmitted to another user in the conversation. This is the final step of the system which gives the actual output of the sign language to text, a screen or display then shows the converted information to the user. This kind of structure is found in systems that need real-time processing and understanding of data like facial recognition, vehicle detection, and self-driving cars. The effectiveness and accuracy of this system depend on data quality, feature extraction process, and machine learning algorithms used.

IV. ALGORITHM DESIGN

In our project on "Recognition and Transfer of Sign Language into Text," we have implemented a multi-layered LSTM (Long Short-Term Memory) model to effectively convert sign language gestures into textual representations. In the context of sequence prediction or sequential data analysis, the utilization of LSTM (Long Short-Term Memory) algorithms holds significant promise. The Long Short-Term Memory (LSTM) algorithm, a subtype of recurrent neural networks (RNNs).

Unlike traditional image feature extraction and clustering methods, LSTMs excel in capturing dependencies and patterns within sequences, making them a natural fit for tasks involving sequential data processing. The LSTM model consists of three LSTM layers and three Dense layers having activation function as RELU and the output layer having activation function as SoftMax, each contributing to the understanding and interpretation of the sequential nature of sign language.

In the LSTM-based proposed workflow to become the main aspect, the data sequentially, e.g. time series data or visual series data is the crucial element. The addition of multiple LSTM layers allows for the model to learn and capture increasingly complex patterns and dependencies present in sign language gestures. Each LSTM layer in the network takes in a sequence of inputs, processes them through its memory cells, and outputs a hidden state that carries information forward to the next layer.

Training our LSTM model contains forward pass and backward pass. In the forward pass stage, the LSTM layer maps input xt and previous timestep hidden state ht-1 to update hidden state ht. Then the hidden state pass though the two

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1204257 |

fully connected layers. Finally, SoftMax layer takes the outputs yt of fully connected layers as input, predict a distribution p(yt = s) at timestep t by Eq. 1, and calculate the loss by Eq. 2:

$$\rho(y_t = S) = \frac{\exp(y_{t,S})}{\sum_{S' \in S} exp(y_{t,S'})}, \quad (1)$$

$$J(x) = \sum_{t=0}^{T-1} ln \sum_{k=0}^{S-1} \delta(k-s)\rho(s_k|x_t), \quad (2)$$

where S represents category number of sign words, $\boldsymbol{\delta}$ (·) is Kronecker function, s is the ground truth label of sign word, and x is the input sequence with an arbitrary length T. The cost function of our model is to minimize the maximum likelihood loss function. In the backward pass, we use stochastic gradient descent to minimize the loss function and back-propagation through time algorithm to update all the weights. In the recognition stage, our model maps each timestep of a sign word to a distribution over S classes in SoftMax layer without calculating the loss. Then we pool all the distribution of a sign word and predict the label of this sign word.

LSTMs can dynamically deal with sequences of variable sizes. Hence, for better flexibility and adaptation purposes, LSTM was preferred. In addition, the LSTM training process is composed of it learning the representations of the sequential order patterns into the backpropagation flowback of their time steps which help the model to capture both the short-term dependencies and long-term context. Thanks to its ability to store information, LSTM under these circumstances is a preferred choice for cases that require more in-depth sequential understanding, like speech recognition, language translation or gesture recognition during sign language processing.

In summary, while the traditional approach described earlier focuses on static image feature extraction and clustering, the adoption of LSTM algorithms signifies a transition towards exploiting temporal dependencies and sequential structures within data, opening up new avenues for advanced analysis and prediction tasks in various domains, including computer vision, natural language processing, and time-series forecasting**.**

V. RESULTS AND DISCUSSIONS

The implementation has done on open-source python environment. The device runs with an INTEL Pentium 4 or higher processor and 4 GB RAM with a distributed manner on the Python V3.9, Jupyter analytic platform. Indian Sign Language dataset has used for recognition of sign language. For the validation of results. The above Figure 3 demonstrates the classification accuracy of the proposed system and comparative analysis with various state-of-art systems. In this proposed system, LSTM algorithm is used because it has more accuracy than other algorithms.

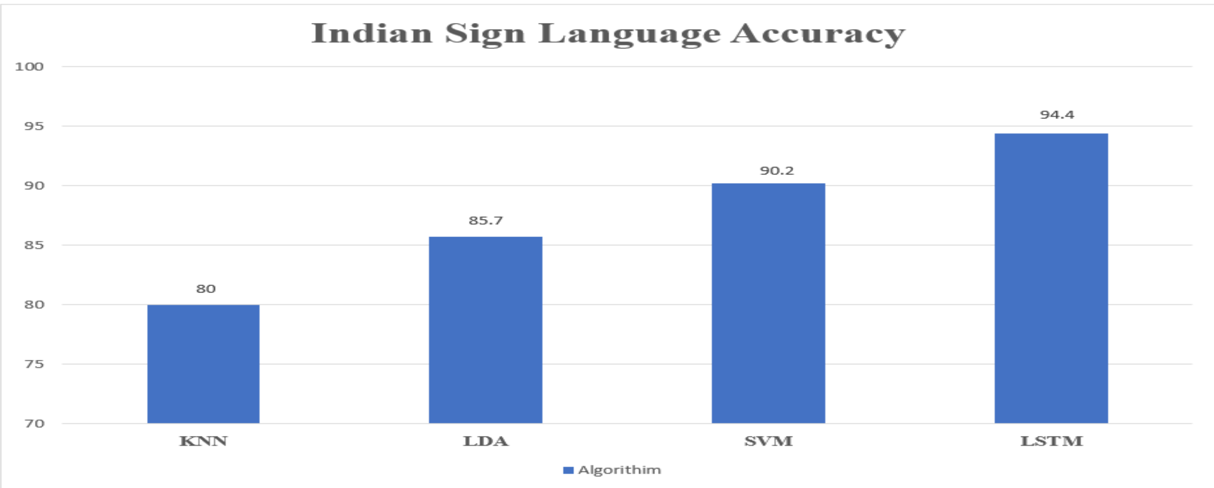


Figure 3: Comparative analysis of accuracy for recognition of sign language using proposed algorithm with other machine learning algorithms

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204257 |

To improve classification accuracy, a variety of feature extraction approaches have been used to build a solid training module. In the proposed work, we extract the numerous features such as drawing, landmarks using holistic module with left hand landmarks and right-hand landmarks, coefficient features and feed them to classify for validation. The proposed classifier has been used for the recognition of sign language and gives highest accuracy, up to 94.4%.

VI. CONCLUSION AND FUTURE WORK

The project "Recognition and Transfer of sign language into text" is focused on solving the problem of deaf and dumb people. ISL is a key for communication for deaf and dumb people in India. This system will automate the hectic task of recognizing sign language, which is difficult to understand for a normal person, thus it reduces the efforts and increases time efficiency and accuracy. Using various concepts and libraries of image processing and fundamental properties of image we trying to develop this system. This paper represented a vision-based system able to recognizes hand gestures from the sign language and convert them into text. The proposed system is tested in the real-time scenario, where it was possible to prove that obtained LSTM models were able to recognize hand gestures. As future work is to keep improving the system and make experiments with complete language datasets.

REFERENCES

[1] Kaustubh Mani Tripathi, Pooja Kamat, Shruti Patil, Ruchi Jayaswal, Swati Ahirrao, Ketan Kotecha, "Gesture-to-Text Translation Using SURF for Indian Sign Language", 2, March 2023.

[2] Yulius Obia, Kent Samuel Claudioa, Vetri Marvel Budimana, Said Achmada, Aditya Kurniawana, "Sign language recognition system for communicating to people with dis-ability", January 2023.

[3] Shubham Thakar, Samveg Shah, Bhavya Shah, Anant V. Nimka, "Sign Language to Text Conversion in Real Time using Transfer Learning", 7 Dec 2022 .

[4] Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta, Gunjan Chugh, "Real Time Sign Language Detection", 31 December 2021.

[5] Shravani K, Sree Lakshmi A, Sri Geethika M, Dr. Sapna B Kulkarni, "Indian Sign Language Character Recognition", May 2020.

[6] C. Dong, M. C. Leu, and Z. Yin., "American sign language alphabet recognition using Microsoft Kinect", June 2015.

[7] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks", 2014.