Indian Sign Language to text/speech translation

Ms. P. Yasmin Bhanu
Assistant Professor of Dept of Computer Applications
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College (Autonomous))
Tirupati, Andhra Pradesh, India

Pinjari Dastagiri
UG Scholar, Dept of BCA
Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College(Autonomous))
Tirupati, Andhra Pradesh, India

Syamtan Sai Yadav Bontha
UG Scholar, Dept of BCA
Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College(Autonomous))
Tirupati, Andhra Pradesh, India

Paturu Sandeep
UG Scholar, Dept of BCA
Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College(Autonomous))
Tirupati, Andhra Pradesh, India

Tharugu Soma Shekar Reddy
UG Scholar, Dept of BCA
Mohan Babu University
(Erstwhile Sree Vidyanikethan Engineering College(Autonomous))
Tirupati, Andhra Pradesh, India

Abstract—The Indian Sign Language to Text or Speech Translation project represents a groundbreaking integration of technology in the field of communication and accessibility. Through this system, ISL communication connects to verbal or written expressions for continuous dialogue. The system functions as an assistive tool for deaf and hard-of-hearing people because it generates real-time text or speech output from sign language inputs. The system facilitates diverse interaction between users of sign language and individuals who do not speak that language. Renewed Indian Sign Language (ISL) translation requires advanced technological functions such as natural language processing (NLP) and computer vision along with machine learning operations for precise translation. Deep learning models, including CNNs and transformers, enhance gesture recognition and speech synthesis. By incorporating real-time processing, the project ensures a smooth and natural user experience. It also contributes to language accessibility, making education and daily interactions more inclusive. The use of AI-driven models enables continuous learning and improvement for higher accuracy. This creative approach shows a notable development that helps better communication and supports a more linked society.

Keywords—Accessibility Technology, Sign Language Interpretation, Audio Processing, Speech-to-Text Conversion, Natural Language Processing (NLP), Computer Vision Deep Learning, Artificial Intelligence.

INTRODUCTION

Among people or groups, communication is a vital tool for expressing ideas and feelings. Good communication promotes good thinking and helps to move things forward. Communication depends on language; it includes not just words but also deeds. Deaf or mute people use sign language to express themselves using physical movement and hand gestures. Many people lack sign language knowledge, which makes it difficult for those who are deaf, hearing impaired, or speech unable to communicate and convey ideas. This issue creates a barrier between those who are deaf or mute and others.

The challenge requires a strong tool called a sign language recognition system which has prompted numerous ongoing research projects about this field. Good for society. Technological developments are always advancing in this competitive climate, hence this interpreter is vital in offering equal chances for everyone, regardless of their handicap.

This world is home to many languages, each with unique regional variations including sign language, which also varies depending on the local language. While the discussion takes place in English, The research adopts American Sign Language (ASL) as its main communication method. The recognition of sign language exists in two different types: static and dynamic. The research adopts static sign language for its methodologies. This means the data is shown as pictures, and it monitors hand motions using a good hand-tracking system. [1]. This system detects hand movements in real time from camera capture.

With data collected and supplied to a deep learning model, this system uses computer vision and natural language processing to provide real-time predictions.

LITERATURE SURVEY

[1] Proposed hand gesture detection utilizing the Karhunen-Loeve (K-L) transform, in conjunction with convolutional neural networks (CNN). The system used skin filtering then separated the palm area through palm cropping and finalised the process with edge detection for hand detection purposes. Subsequently, feature extraction of the hand was performed utilizing the K-L transform approach, followed by picture classification employing Euclidean distance. They evaluated 10 distinct hand motions with an accuracy of 96%.

[2] Proposed Contour tracing characteristics help to identify single-handed sign language actions. This paper shows that The system extracts hand outlines by separating them from background elements through skin colour identification in RGB along with YCbCr colour domains, together with threshold intensities of grey levels. Gesture outlines detected by segmentation were traced using the contour tracing descriptor. To evaluate the accuracy, they used picture categorization using KNN and SVM-supervised machine learning techniques.

[3] Identification of hand gestures based on Principal Component Analysis (PCA). The system applies hand vein recognition through colour models with thresholding

methods and fast template matching that enables efficient hand verification identification. Portable skin colour modelling in the YCbCr colour space defines recognition. The foreground and background are separated using Otsu thresholding. Gesture recognition uses template matching based on Principal Component Analysis (PCA). For low-brightness pictures, the algorithm achieved 91.43% accuracy.

[4] Suggested deep convolutional neural networks for letters and numbers proposed American Sign Language gesture identification. This paper addresses the preparation of images, whereby a background removal technique was used to remove the backdrop. Using CNN for image categorization, The dataset was split into two parts: one for training and the other for testing. Concerning the motions of the alphabet, the system was 82.5% accurate.

[5] Proposed Suggested examination of plans for language and hand gesture identification. Executing data collecting and pre-processing, The system used median and Gaussian filters to reduce noise whereas it applied histogram equalization to clear unnecessary data in morphological approaches. It then segmented, tracking for hand detection and skin colour segmentation included. Culminating in picture categorization, the following stage was feature extraction using several approaches. This talk provides a comprehensive overview of the subject of automatic gesture and language recognition.

[6] Proposed dynamic sign language recognition tool. For image categorization, forecasting, and identification, they used a supervised learning technique called SVM. This system triggers sign gesturing detection from live video stream capture. The author utilizes darkened images and isolated white hand borders to achieve their goal. Pulls hand outlines from video frames. This boundary defines the hand's contours.

[7] Proposed Deep learning methods help to identify static signs in sign language. Hand identification using this approach is based on skin colour modelling; it uses a defined skin colour range to isolate hand pixels (foreground) from non-pixels (background). Using images with a consistent background, the approach classified images using a CNN. In American Sign Language alphabet recognition, the system achieved 90.04% accuracy; in testing accuracy, 93.67%.

[8] Suggesting hand gesture identification for static images using convolutional neural networks (CNN). The system contains photo pre-processing methods which include Morphological operations, contour extraction, polygon approximation and segmentation. The research team employed CNN architectures for extracting photos' characteristics before performing classification to obtain results from multiple CNN designs.

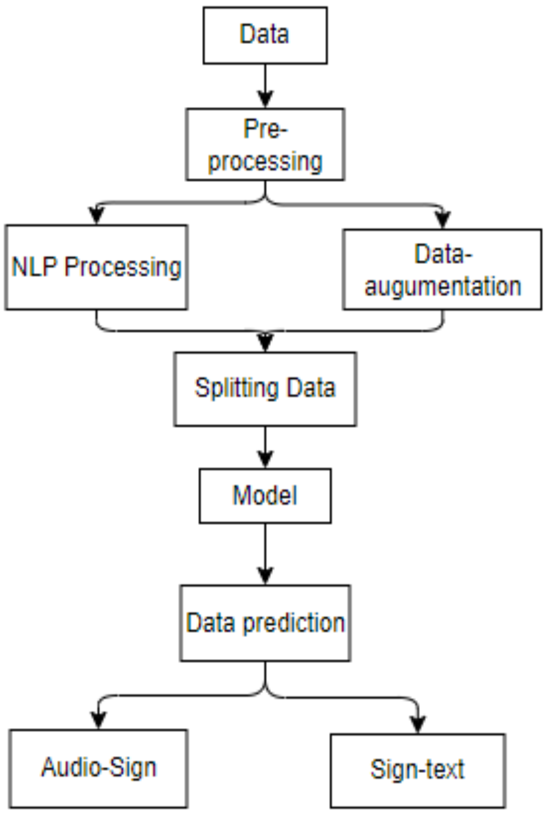
[9] Proposed The authors designed a sign language recognition system combining CNN with computer vision technology. [9] HSV colour technology functions as the detection algorithm for hand gestures while the researchers designated the background to show black colour. The image pre-processing involves conversion to grayscale then dilation and mask operation followed by gesture segmentation. The CNN architecture was used for picture categorization, and then feature extraction in the first layer. The system correctly identified 10 letters with a 90% accuracy.

[10] Stokoe's 2005 study highlights the linguistic structure of American Sign Language (ASL). He identifies key phonological components—handshape, location, and movement—comparable to phonemes in spoken languages. His work proves ASL is a fully structured language with its grammar. The study emphasizes ASL's visual-spatial nature, differing in modality but not complexity. It was instrumental in changing perceptions of sign languages. Stokoe's research remains a cornerstone in sign linguistics and deaf studies.

PROPOSED SYSTEM

The proposed system aims to develop a real-time translation tool that converts spoken language to sign language and vice versa using advanced NLP, AI, and computer vision technologies. Unlike traditional approaches that rely on pre-recorded videos, this system dynamically interprets and generates sign language using MediaPipe for gesture recognition and NLP techniques for accurate text translation. This enhances promoting communication for the deaf and hard-of-hearing community.

PROJECT FLOW



SYSTEM DESIGN

The system design is structured to handle a comprehensive pipeline from data ingestion to actionable insights. It starts with data collection, where raw data from various sources, such as text, audio, and visual inputs, is aggregated. The preprocessing phase involves cleaning, normalizing, and transforming this data to ensure consistency and quality. For textual data, NLP processing is employed to tokenize, lemmatize, and extract meaningful features, enabling the system to understand and process human language effectively.

Then, methods of data augmentation are used to strengthen the quality of the dataset., particularly useful in scenarios with limited data. This step involves generating synthetic data or modifying existing data to improve model generalization. The data is subsequently split into training, validation, and testing sets to facilitate a structured approach to model development and evaluation.

The model design phase focuses on selecting and fine-tuning appropriate algorithms, whether they are traditional machine learning models or advanced deep learning architectures, depending on the task's complexity. Once the model is trained and validated, it is deployed for data prediction, where it processes new inputs to generate accurate outputs.

For multimodal data, such as audio and sign language, the system incorporates specialized modules. Audio signals are converted into text using speech recognition techniques, while sign language gestures are interpreted into textual representations through computer vision and gesture recognition models. This integration ensures the system can handle diverse data types seamlessly.

The final design emphasizes scalability, accuracy, and efficiency, ensuring that the system can adapt to varying data volumes and types while maintaining high performance. Continuous monitoring and updates are incorporated to keep the system relevant and effective in dynamic environments.

IMPLEMENTATION

Index Page

Provides an overview of the Indian Sign Language to Text/Speech Translation System and navigation options to other pages such as Registration, Login, User Dashboard, and Information Sections.

Registration Page

Allows users to sign up for an account by providing the required details.

Login Page

Enables registered users to log in using their credentials, hence granting access to their tailored dashboard and features.

User Home Page

The main dashboard for registered users offers access to key features such as text translation, sign language recognition, and accuracy metrics.

Text Processing & NLP Page

Handles text input processing using NLP techniques like tokenization, lemmatization, and grammar correction before conversion into sign language or speech.

Speech-to-Text Page

Allows users to record or upload audio, which is converted into text using speech recognition APIs, helping to link sign language and spoken language.

Sign Language Recognition Page

Captures live video input and uses computer vision models (MediaPipe, CNN, ViT, GRU) to detect hand gestures and facial expressions for sign language translation.

Sign Language Animation Page

Displays sign language animations corresponding to the processed text, using a gesture library and real-time rendering techniques.

Accuracy Page

Displays performance metrics and visualizations of the gesture recognition and NLP models, including translation accuracy, error rates, and processing speed.

Prediction Page

Uses machine learning models to predict text, speech, and sign gestures, ensuring real-time and accurate translation results.

Logout Page

Allows users to securely log out, terminating active sessions and protecting their account data and privacy.

SYSTEM STUDY AND TESTING

The feasibility study ensures that the proposed system is practical and beneficial. It includes economic feasibility, which evaluates whether the project stays within budget, utilizing free technologies except for necessary customizations. Technical feasibility ensures that the system does not overburden existing resources, requiring minimal changes for implementation. Social feasibility assesses user acceptance, ensuring proper training and a positive experience to encourage adoption. System testing is conducted to identify and fix errors, verifying that the software meets requirements and functions as expected.

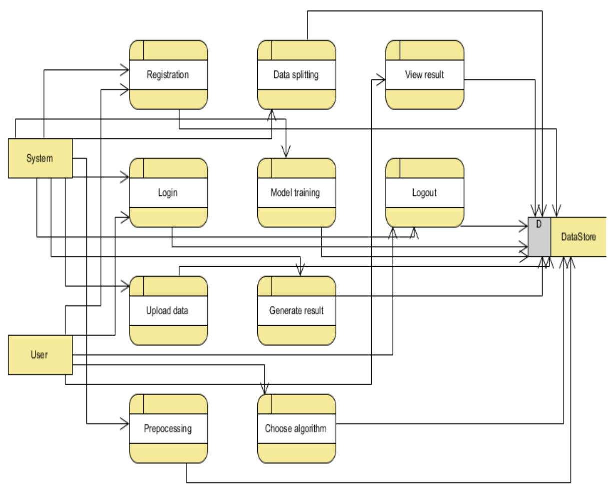
Test cases :

S.NO	Test cases	I/O	Expected O/T	Actual O/T	P/F
1	Read the dataset.	Dataset path.	Dataset need to read successfully.	Dataset fetched successfully.	P
2	Performing pre-processing on the dataset	Pre-processing part takes place	Pre-processing should be performed on dataset	Pre-processing successfully completed.	P
3	Model Building	Model Building for the clean data	Need to create model using required algorithms	Model Created Successfully.	P
4	Text-sign Sign-Text	Input image provided.	Output should be the Sign and text	Model successful	P

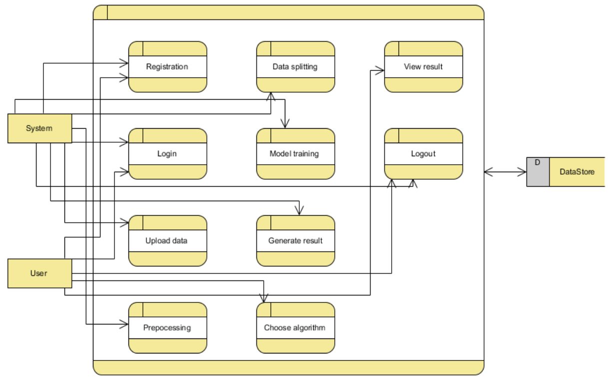
DATAFLOW DIAGRAM

Information flow within a system is typically depicted through the use of a data flow diagram. DFD a well-structured and Visually organised DFD can effectively illustrate a substantial portion of the system requirements. The process can be either manually automated or a combination of both methods. This illustrates how data flows in and out of the system, what modifies the data and where data is kept the primary objective of ADFD is to visually represent the overall scope Unlimited of a system. The tool facilitates communication between system analysts and any individual involved in the system, serving as a launchpad for system reconfiguration.

Level 1 diagram :



Level 2 diagram :



CONCLUSION

The development and implementation of the audio-sign and sign-audio translation project represent a significant development in communication access and assistive technologies a significant advancement in the field of assistive technology and communication accessibility. This project successfully integrates several complex technologies, including natural language processing (NLP), speech-to-text conversion, gesture recognition using Media Pipe, and dynamic sign language animation. The result is a versatile and user-friendly platform that facilitates real-time interaction between non-users and sign language users bridging a critical gap in interpersonal interactions.

FUTURE ENHANCEMENT

Future advancements in the Indian Sign Language to Text/Speech Translation System will focus on improving its intelligence, speed, and inclusivity. AI-driven enhancements will refine speech recognition to better handle background noise, diverse accents, and variations in tone. Gesture recognition models will be trained on larger datasets, allowing the system to identify complex sign language dialects and non-manual signals like facial expressions, body posture, and movement speed, which are crucial for accurate interpretation. Real-time processing improvements will minimize latency, ensuring instant translation between sign language, text, and speech for a natural and fluid user experience. The sign animation module will be upgraded with 3D avatars that replicate human signing with lifelike movements and expressions, making digital communication more engaging. Expanding language and dialect support, the system will accommodate regional variations of sign language and provide translations in multiple spoken languages, bridging communication gaps worldwide. To increase accessibility, mobile and wearable device integration will be explored, allowing users to interact through smartphones, AR glasses, or IoT devices. Further, cloud-based processing will be introduced to enable scalability and real-time global access. Lastly, user customization options will be developed, allowing individuals to train the system with personalized gestures and speech patterns for more adaptive and user-friendly interactions. These enhancements will push the boundaries of assistive technology, fostering greater inclusivity, independence and social inclusion for those who are hard of hearing or deaf.

REFERENCES

1. Liddell, S. K., & Johnson, R. E. (1989). "American Sign Language: The phonological base." Sign Language Studies.
2. McNeil, M. R. (Ed.). (2016). "Clinical Management of Sensorimotor Speech Disorders." Thieme.
3. Sutton-Spence, R., & Woll, B. (1999). "The Linguistics of British Sign Language: An Introduction." Cambridge University Press.
4. Goldin-Meadow, S., & Mylander, C. (1998). "Spontaneous sign systems created by deaf children in two cultures." Nature.
5. Stokoe, W. C. (2005). "Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf." Journal of Deaf Studies and Deaf Education.
6. Valli, C., & Lucas, C. (2000). "Linguistics of American Sign Language: An Introduction." Gallaudet University Press.

7. Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). "Language and Deafness." Jones & Bartlett Learning.

8. Vogler, C., & Metaxas, D. (2001). "A Framework for Recognizing the Simultaneous Aspects of American Sign Language." Computer Vision and Image Understanding.

9. Ladefoged, P., & Maddieson, I. (1996). "The Sounds of the World's Languages." Blackwell.

10. Corina, D. P., & Knapp, H. (2006). "Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence." Journal of Deaf Studies and Deaf Education.