IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Adaptive Solutions Transforming Lives for the Disabled People

Dr. Shilpa V [1], Archana V [2], Akshitha N [3], Pavithra V N [4] and Sindhu K M [5]

[1]Assistant Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology (CITech), Bengaluru, India
[2,3,4,5]Student, Department of Computer Science and Engineering, CITech, Bengaluru, India

Abstract: The goal of this project is to promote inclusivity in social interactions and employment possibilities by addressing the communication issues that people who are blind, deaf, or mute confront. The suggested remedy consists of a two-way smart communication system made to make it easier for people with and without sensory impairments to communicate with each other. The initial component of the system provides an easy-to-use interface that helps people who are blind, deaf, or mute effectively communicate. To empower users with a range of sensory needs, the system makes use of cutting-edge technologies including gesture control, speech recognition, and haptic feedback. Voice instructions, tactile gestures, or both can be used by users to input messages. The system then converts these inputs into a textual and auditory format that is simple to comprehend, making the messages transmitted understandable to people with different levels of communication proficiency.

With the use of machine learning algorithms, this interface enables accurate and speedy transcription of spoken words into text through speech-to-text conversion.

***Index Terms* - Inclusivity, Social Interactions, Employment opportunities, sensory impairments, gesture control, speech recognition, haptic feedback, Voice instructions, tactile gestures, user-friendly interface, machine learning algorithms, speech-to-text conversion.**

I. INTRODUCTION

The development of technology in recent years has opened the door for innovative solutions meant to improve the lives of people with disabilities [1]. These adaptive solutions—which range from cutting-edge software programs to assistive devices—have completely changed accessibility, inclusion, and independence for individuals with impairments [2]. This report offers a thorough analysis of the several adaptive approaches that are reshaping the disability support market. [3]We investigate the significant effects that the most recent developments in wearable technology, robotics, assistive technologies, and inclusive design have on the day-to-day lives of people with disabilities [4].We also look at the potential and problems associated with creating and implementing adaptive solutions, taking sociocultural, economical, and practical aspects into account [5]. Our goal in conducting this investigation is to draw attention to the significance of continued study, cooperation, and advocacy in building a more just and inclusive society that benefits individuals with a range of abilities.

II. LITERATURE SURVEY

Recent advancements in sign language recognition have been observed across various domains. **Rajalakshmi et al.** introduced a hybrid deep neural architecture aimed at enhancing sign gesture recognition through multi-semantic discriminative feature learning [1]. **Zhou et al.** proposed a spatial-temporal multi-cue network designed to improve sign language recognition and translation capabilities [2]. **Wang et al.** presented a real-time end-to-end sign language recognition system, demonstrating its efficacy in mobile computing environments [3].In the realm of wearable technology, **Kudrinko et al.** provided a comprehensive review of sensor-based sign language recognition systems, shedding light on their potential for real-time, unobtrusive recognition solutions [4].**Al-Qurishi et al.** surveyed deep learning techniques for sign language recognition, addressing current methodologies, benchmarks, and open challenges.[5]. **Adaloglou et al.** conducted an extensive study on deep learning-based methods, exploring their applications and performance in sign language recognition tasks.

[6]. Beyond technical advancements, **Mariya Celin et al.** developed a speaker-specific communication aid for dysarthric speakers, highlighting the broader implications of sign language recognition in augmentative and alternative communication

[7].**Effendi et al.** proposed an image-to-speech generation system, showcasing innovative applications of sign language recognition beyond traditional linguistic boundaries [8]. Additionally**, Papastrat et al.** investigated cross-modal alignment techniques, facilitating continuous sign language recognition through the integration of video and text embeddings [9]. Finally, **Kamal et al.** reviewed technical approaches specific to Chinese sign language processing, offering insights into language-specific challenges and methodologies [10]. Together, these studies underscore the diverse and interdisciplinary nature of sign language recognition research, contributing to advancements in accessibility and communication for individuals with hearing impairments.

III. METHODOLOGY

3.1 Text To Speech

The provided code uses the Tkinter toolkit in Python to put up a graphical user interface (GUI) for a text-to-speech application. Using the pyttsx3 library, it initializes a text-to-speech engine and implements a function that turns text typed into a text box into speech. A title, a text field where text can be entered, and a "Play" button that initiates the text-to-speech conversion are all included in the GUI. Users can input text, select a voice, and play the transformed speech using this application. The "Text to Speech" title appears in a purple top frame, and a green "Play" button facilitates straightforward user interaction.

3.2 Speech To Text

Using Python's Tkinter toolkit, the provided code creates a graphical user interface (GUI) for a speech-to-text application. The speechToText() function manages the interface's configuration and operation. In order to convert text, users must first set up the speech recognition system by speaking into the microphone. The user interface asks users to speak, waits for their response, and then shows the text that has been identified in a text field inside the GUI. With an obvious headline ("Speech to Text") and a button marked "Click here to Speak" to start the speech recognition process, the design is user-friendly. This program makes it easier to turn spoken words into text, making it a useful tool for many different applications that need speech-to-text functionality.

3.3 Object Detection

The supplied code uses a pre-trained SSD MobileNet V3 model to achieve real-time object detection. To start the detecting procedure, it loads the frozen model and the required configuration files. It also assigns suitable names to observed objects by reading class labels from a text file. The model is then set up for input size, scale, mean, and color swapping by the code. After starting video capture from the default camera, it goes into a loop that keeps taking pictures, uses the loaded model to identify things, and then overlays labels and bounding boxes on the objects it finds. A window labeled "Object Detection" displays the processed frames that have object detections. When the user hits the 'q' key, OpenCV closes all of its windows and releases the video capture resources. All things considered, this code offers a complete solution for object detection in real-time in video streams, with possible applications in robotics, automation, and surveillance, among other fields.

3.4 Image To Speech

A series of procedures must be followed in order to translate a tesseract image into speech. To extract the text included in the image, processing must first be performed on the image. Optical Character Recognition (OCR) techniques are frequently used in this procedure to extract characters from the image. OpenCV, Pytesseract, and Tesseract are a few of the OCR packages available for Python; Tesseract is a recommended option because of its stability. Installing the Tesseract OCR engine and its Python wrapper, pytesseract, on the system is necessary to put this into practice. After everything is set up, the image can be loaded using Python's PIL package, and the text can be extracted from the image using OCR using Pytesseract. The captured text is then transformed into speech using a text-to-speech (TTS) library, like pyttsx3, after that. By initializing a TTS engine, this module makes speech parameters like speed customizable. Ultimately, the text is fed into the engine, which produces speech out of it. By combining the OCR and TTS components, a script is produced that can read in a tesseract image, extract text from it, and output the text as speech. This makes the tesseract's information more accessible and useful for other purposes.

.

3.5 Sign Detection

As part of our effort to develop adaptable solutions that would aid people with disabilities, especially in the area of sign gesture identification, we have carefully employed preprocessing techniques to increase the performance of our model. We have simplified the data format while maintaining important visual clues by converting our sign gesture photographs to grayscale, which will help our system learn more quickly. Furthermore, we have expanded the dataset with variants covering various lighting situations, orientations, and backdrops using thorough data augmentation techniques, which improves the model's capacity to adapt to real-world circumstances. Our attempts to create a reliable and strong sign gesture recognition system, which has the potential to transform communication accessibility and enable people with disabilities to express themselves more effectively in their day lives.

We have employed rigorous preprocessing methods and comprehensive data augmentation procedures as we can see in the Fig 1 in our quest to create a robust and flexible sign gesture detection system. Grayscale conversion of sign gesture photos allows us to streamline the data format while maintaining important visual clues, which speeds up our model's learning process. In addition, we have added variations of various lighting situations, backdrops, and orientations to the dataset through our augmentation efforts, which guarantees the model's resilience and suitability for a variety of real-world circumstances. This all-encompassing strategy not only improves the functionality of the system but also has the potential to greatly increase the accessibility of communication for people with disabilities. Our approach seeks to allow users with impairments to participate more completely in their daily lives by facilitating more effective sign language expression in a variety of settings.

Fig 1. Mediapipe landmarks

We have a project on adaptable solutions for people with disabilities, and we have a dataset that we obtained via Kaggle. A multitude of datasets on a variety of subjects, including accessibility and assistive technology, are available on Kaggle. This dataset will be essential to our study and creation of creative solutions meant to improve the quality of life for those with disabilities. Repaired linear unit (ReLU) activation functions are another common feature of CNN layers that give non-linearity to the model and help it identify complex patterns and correlations in the data. CNN architectures combined with maxpooling and ReLU activation functions can be used by sign language recognition systems to effectively process and extract pertinent characteristics from the sign language images displayed in Fig. 2. This improves the accuracy and robustness of the recognition process.

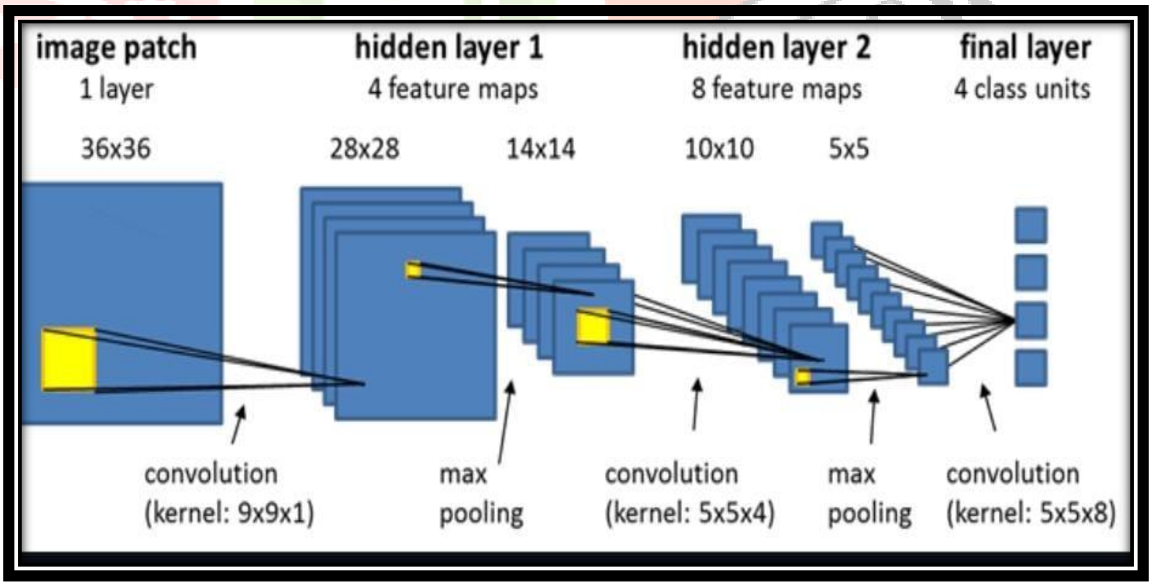
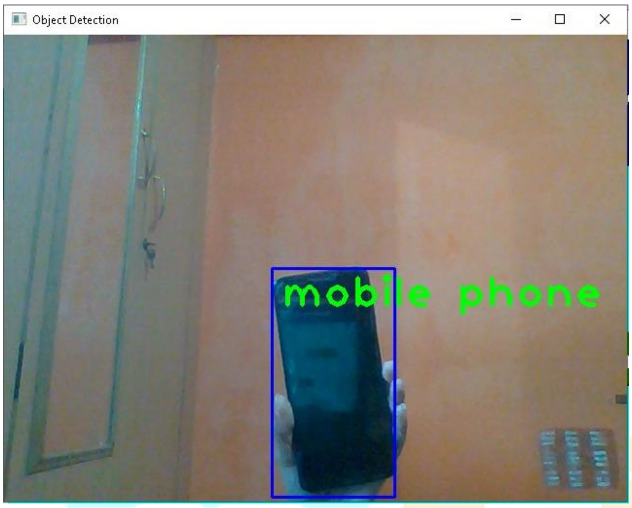


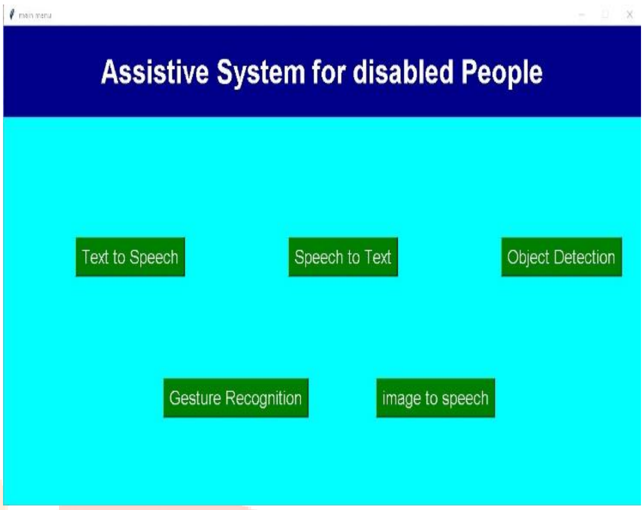
Fig. 2. CNN Architecture

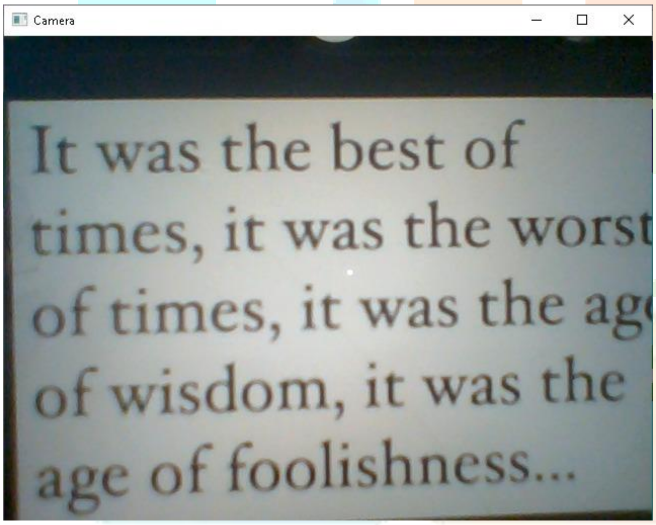
Convolutional Neural Networks (CNNs) have become a mainstay in computer vision due to their exceptional performance in image recognition applications. CNNs use pooling operations and convolutional filters to learn hierarchical representations of visual data with great success. With the help of these filters, the network is able to record spatial patterns and features at various abstraction levels, ranging from basic elements like edges and textures to more complex ideas like shapes and objects. CNNs can process large-scale picture collections because they achieve parameter efficiency by utilizing the shared weights of convolutional layers. Furthermore, the application of pooling layers facilitates translation invariance and feature generalization by lowering the spatial dimensionality of feature maps while maintaining crucial information.

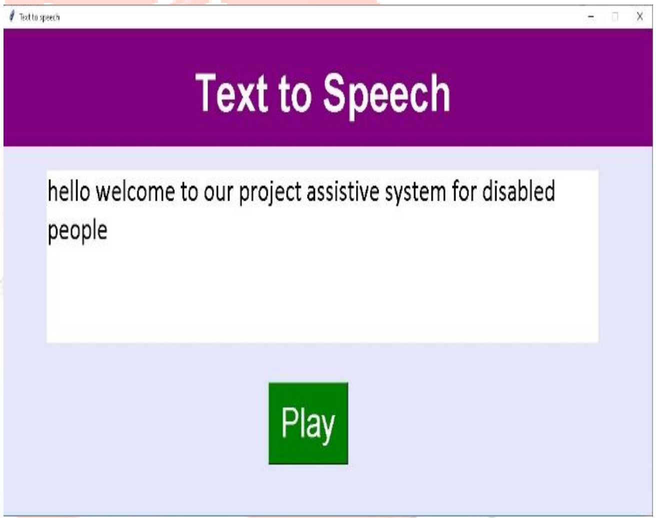
IV. RESULTS

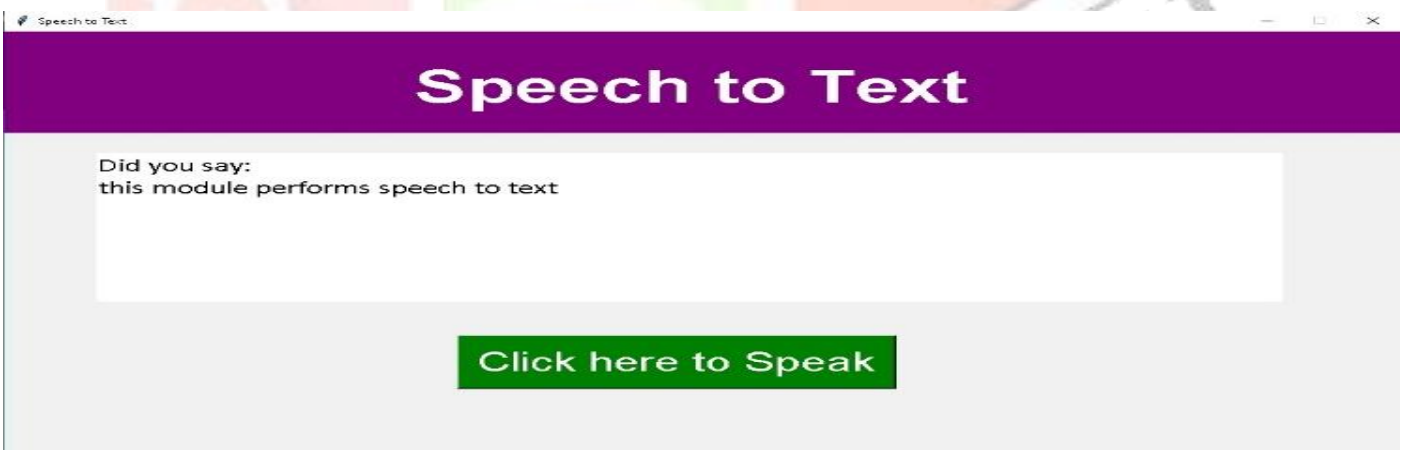
We utilized a Convolutional Neural Network (CNN) architecture for sign language recognition, while employing a Deep Neural Network architecture for object detection.











V.CONCLUSION

This project will enable the deaf and dumb community to enjoy normal lives and communicate with the rest of society more easily. The technology is used to translate hand gestures into text and voice for the dumb, text and images into voice for the blind, and speech to text for the deaf. We have designed a little device that serves as a model for individuals who are deaf, blind, or dumb. One benefit of this gadget is that it's small and light, making it convenient to carry around. The technology functions as a smart assistant to facilitate communication between people with impairments and others, regardless of their language.

VI. FUTURE ENHANCEMENT

Upcoming Improvement Can be further implemented with any other sophisticated device by utilizing a basic programming language to reduce complexity. One way to make the system more convenient is to integrate it into a mobile phone. The system might be more effective for every language. Reducing the time delay in gesture recognition.

REFERENCES

[1] E. Rajalakshmi et al., "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226- 2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[2] H. Zhou, W. Zhou, Y. Zhou and H. Li, "Spatial-Temporal Multi-Cue Network for Sign Language Recognition and Translation," in IEEE Transactions on Multimedia, vol. 24, pp. 768-779, 2022, doi: 10.1109/TMM.2021.3059098.

[3] Z. Wang et al., "Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System," in IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 2398-2410, 1 July 2022, doi: 10.1109/TMC.2020.3038303.

[4] K. Kudrinko, E. Flavin, X. Zhu and Q. Li, "Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review," in IEEE Reviews in Biomedical Engineering, vol. 14, pp. 82-97, 2021, doi: 10.1109/RBME.2020.3019769.

[5] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917- 126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

[6] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," in IEEE Transactions on Multimedia, vol. 24, pp. 1750-1762, 2022, doi: 10.1109/TMM.2021.3070438.

[7] T. A. Mariya Celin, G. Anushiya Rachel, T. Nagarajan and P. Vijayalakshmi, "A Weighted Speaker-Specific Confusion Transducer-Based Augmentative and Alternative Speech Communication Aid for Dysarthric Speakers," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 2, pp. 187-197, Feb. 2019, doi: 10.1109/TNSRE.2018.2887089.

[8] J. Effendi, S. Sakti and S. Nakamura, "End-to-End Image-to-Speech Generation for Untranscribed Unknown Languages," in IEEE Access, vol. 9, pp. 55144-55154, 2021, doi: 10.1109/ACCESS.2021.3071541.

[9] I. Papastratis, K. Dimitropoulos, D. Konstantinidis and P. Daras, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," in IEEE Access, vol. 8, pp. 91170-91180, 2020, doi: 10.1109/ACCESS.2020.2993650. Page 32

[10] S. M. Kamal, Y. Chen, S. Li, X. Shi and J. Zheng, "Technical Approaches to Chinese Sign Language Processing: A Review," in IEEE Access, vol. 7, pp. 96926-96935, 2019, doi: 10.1109/ACCESS.2019.2929174.

[11] G. Bordel, M. Penagarikano, L. J. Rodríguez-Fuentes, A. Álvarez and A. Varona, "Probabilistic Kernels for Improved Text-to-Speech Alignment in Long Audio Tracks," in IEEE Signal Processing Letters, vol. 23, no. 1, pp. 126-129, Jan. 2016, doi: 10.1109/LSP.2015.2505140.

[12] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano and F. Girardi, "Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System," in IEEE Access, vol. 5, pp. 22199-22208, 2017, doi: 10.1109/ACCESS.2017.2762475.

[13] T. Wang, J. Yi, L. Deng, R. Fu, J. Tao and Z. Wen, "Context-Aware Mask Prediction Network for End-to-End Text-Based Speech Editing," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 6082-6086, doi: 10.1109/ICASSP43922.2022.9746765.

[14] R. Jain, M. Y. Yiwere, D. Bigioi, P. Corcoran and H. Cucu, "A Text-to-Speech Pipeline, Evaluation Methodology, and Initial Fine-Tuning Results for Child Speech Synthesis," in IEEE Access, vol. 10, pp. 47628-47642, 2022, doi: 10.1109/ACCESS.2022.3170836.

[15] M. Zhang, X. Zhou, Z. Wu and H. Li, "Towards Zero-Shot Multi-Speaker MultiAccent Text-to-Speech Synthesis," in IEEE Signal Processing Letters, vol. 30, pp. 947- 951, 2023, doi: 10.1109/LSP.2023.3292740.