iAssist – An Intelligent Reading Assistant for Visually Impaired

Dr. S. Perumal Sankar
Professor, Dept. of ECE
Toc H Institute of Science and Technology
spsankar2004@gmail.com

Renjith P K
Assistant Professor, Dept. of ECE
Toc H Institute of Science and Technology
renjithpk@tistcochin.edu.in

Ahammed Suhail P.I
Student, Dept. of ECE
Toc H Institute of Science and Technology
Ahammedsuhail2002@gmail.com

Aswathy P S
Student, Dept. of ECE
Toc H Institute of Science and Technology
aswathyps2808@gmail.com

Nithya Mary K J
Student, Dept. of ECE
Toc H Institute of Science and Technology
nithyamarykj@gmail.com

Sharon K J
Student, Dept. of ECE
Toc H Institute of Science and Technology
sharonkj457@gmail.com

***Abstract*— Vision is a crucial human sense, and visually impaired persons encounter challenges in reading and comprehending text. While various devices and assistive technologies, including advanced mobile applications, have been developed to aid visually impaired individuals in reading, these solutions are often expensive and not universally accessible. Moreover, relying on mobile phones for text reading may pose discomfort, especially for elderly individuals with visual impairments. This proposed project aims to design a system that could assist the visually impaired individuals in reading the text using TTS (Text-to-Speech) and OCR (Optical Character Recognition) technologies. The integrated camera module captures the text, and image processing is carried out using OpenCV coded in python. OCR is employed to extract text from images, which is then articulated by the TTS conversion unit. Additionally, the system incorporates language translation functionality and aids users in identifying the objects in their immediate environment using Yolov9. The system comprises a camera, Raspberry Pi, power bank and earphones. This assistive device empowers blind and partially sighted individuals to gain autonomy in reading printed texts without relying on external assistance.**

Keywords—TTS, OCR, Raspberry Pi, OpenCV, camera

I. INTRODUCTION

Visual impairment is the condition in which a person faces partial or complete inability in visual perception. Visually impaired undergoes many challenges in doing their daily activities. Impaired vision affects all activities of an individual's life including reading where the individual need to depend on others to even read a simple text. Braille the writing system consisting of raised dots which involves touch reading helps blind in reading text. But learning Braille requires lots of effort and dedication and also it poses challenges to individuals who lost their vision in later life. The emerging technologies address the difficulties faced by the visually impaired in being independent in reading the text. But many of these technologies might have a complex interface making them less user-friendly for those who are unfamiliar with the technology. Also, many reading aids are expensive and are not affordable for all. Thus, needs a reading aid which is cost effective, simple in setup and easy to use. Therefore, the main problems identified in this project are: (1) Braille system consumes more time and not always reliable (2) Blind faces difficulties in identification of objects (3) Some people face difficulty in understanding English text.

The proposed system is a reading assistant for visually impaired using raspberry pi and camera. The main objectives of the proposed system are: (1) To design a system for visually impaired individuals to recognize various text in their surroundings (2) To provide an audio version of extracted text (3) To implement a system for identifying different objects in the immediate environment (4) To develop a system that could

translate text to Malayalam language. The system named 'iAssist' could help the users in reading text in reading text by providing the audio version of the text. Optical Character Recognition (OCR) combined with Text-to-Speech (TTS) synthesis is used in this proposed project. The text captured by camera is extracted with the help of OCR and is provided as audio output through the earphones connected to raspberry pi. Also, this involves translation of the text to Malayalam language. The proposed project helps the visually impaired in being independent by reducing the reliability on others in reading text.

II. LITERATURE SURVEY

The proposed system involves image processing, grey scale conversion and binary conversion of the captured image. In [1], author developed a mobile app that assists visually impaired or blind (VIB) people in reading texts. A standalone smartphone application extracts text and reads it aloud to users. The interface is built and modified based on user feedback, and on every interaction the text to speech model generates voice instructions for user convenience. The accuracy and cost-effectiveness of the solution are its strongest points.

A complete Image2Speech system that can be trained without textual data is suggested [2]. In this, an image captioning model uses discrete labels, and in an unsupervised way a discrete representation of a speech is made by vector-quantized variation autoencoder(VQ-VAE) model. Moreover, least amount of paired image-speech data is required to train this model. The text-to-speech which performs the synthesis in small units by preserving the naturalness of the speech is provided in [3] by the author. The suggested incremental TTS approach accounts for future contextual information without adding to latency by using a pseudo lookahead produced by a language model.

Under the supervision of perceptual loss, a novel method is introduced in [4], by training a TTS model that could improve the quality of speech. The distance between the predicted one and the maximum possible speech quality score is calculated in this. Also, a pre-trained mean opinion score (MOS) prediction model is used. In [5], author proposed a design with advanced image processing and obstacle avoidance algorithms for object detection by incorporating a camera and sensors. The faces and eyes are detected and distance measurement is implemented by using the object detection application interface (API) TensorFlow, libraries and its frameworks such as OpenCV and Haar cascade classifier. A Mel spectrogram is generated from a given phoneme sequence by developing a paragraph TTS model that consists of a modified Tacotron2 [6]. The encoders are used to capture information in a paragraph and multi-head attention mechanisms learns the information from inter-sentence in a paragraph.

Text-to-speech conversion is accomplished using eSpeak software [7], further by providing a gadget with input the required product is found out by using a natural language processing algorithm. A self-supervised pre-training model Robustly Optimised Bidirectional Encoder Representations from Transformers (RoBERTa) based on deep neural networks can be used for texts that cannot be fully transcribed to fill in the blanks by predicting the masked sections of text [8]. The author recommended a setup in [9], in which a wooden plank is used to held the Raspberry Pi in place and a wooden stand is mounted on it. This equipment on the printer requires more manpower. For scanned documents, a brand-new multilingual optical character recognition (OCR)A new multi-lingual Optical Character Recognition (OCR) system is being created [10]. The proposed multilingual OCR system incorporated three neural blocks: a segmenter, a switcher, and several recognizers for different languages-as well as the reinforcement learning of the segmenter were combined into suggested multilingual OCR system.

For image processing, a smartphone app called Image Assistant is suggested; it primarily does tasks including word extraction, text reading, face restoration, and object discovery [11]. With support for Arabic, a specific image that is saved and which includes both text and an image can be searched for and located by the user with the help of this program. Utilization of Google Play Services Library facilitates facial recognition in conjunction with image detection technology. Using Raspberry Pi 3b and python a text reader prototype is created in [12] in text extraction using CRNN and OCR. The Synth 90k word dataset is used in conducting the experiment and training. The inventor of [13] introduced a brand-new Android-based text reader app that helps elderly and partially sighted individuals comprehend the content of objects, especially texts. The system recognizes and detect text using Google Mobile Vision Text API and uses the Android library with text-to-speech engine.

The author of [14] envisioned a wearable gadget that could act as a sort of Virtual assistant system for visually impaired person. The core element of the system is a voice-over Chat-bot that facilitates comprehension of surroundings, seeking for an object, perceiving the expression in an individual's face, aiding with reading, etc. AIML creates the assistive chatbot. An embedded architecture made up of different hardware and

software components is used to construct a personal assistive robot that performs assistant functions [15]. Coral USB Hardware Accelerator is used as a coprocessor in conjunction with Raspberry Pi. The robot was created during the height of the COVID-19 epidemic and is programmed to alert users when it is necessary to alert users when it is necessary to wear masks.

TABLE I: Comparison of various systems

Sl.No	Methodology	Advantages	Drawbacks
[1]	Optical character recognition Text-to-Speech	Mobile app to assist blind in reading text	Not convenient for all Text translation is absent
[2]	Vector quantized variational autoencoder (VQ-VAE) model Automatic speech recognition (ASR) system	Self-supervised speech representation Text free training Requires less amount of paired image speech data	Limited to speech-only data Computational complexity
[3]	TTS framework based on Tacotron-style model Kaldi-based waveform extraction	Synthesize speech in small linguistic units Synthetic speech quality equivalent to sentence-level TTS	Requires a large pretrained language model Require additional computational resources
[4]	MOS prediction Neural TTS Speech synthesis	Improve speech quality Efficient and does not increase the inference time or model complexity	Pre-training requires additional data and computational resources Effectiveness on other TTS models and datasets is not fully explored
[5]	Tensor flow object detection Haar cascade classifier Distance measurement	Helps in navigating blind Wearable device	Text translation is absent Not affordable
[6]	Hidden Markov Model Toolkit (HTK) TTS corpus Multi-band Wave RNN vocoder	Improved quality of synthesized speech Better performance compared to baseline models	Relatively small size of the training data used Not optimized for real time speech synthesis
[7]	Optical Character Recognition (OCR) Text-to-Speech synthesizer Natural Language Processing Algorithm	The device can extract texts from documents and convert them into speech Process multilingual documents and convert them into voice format	Efficiency is low Object detection is not possible
[8]	Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa) Natural Language Processing (NLP) OCR	Improved Efficiency Time and Error Reduction Improved Accuracy	Require a large amount of training data Difficult to interpret
[9]	Optical Character Recognition (OCR) TTS Tessaract platform	Helps partially visionless people to convert text to voice output Voice output translation is possible	System is not portable Not user friendly
[10]	REINFORCE algorithm Multilingual OCR	Improves the performance for multi-lingual scripts	Training requires a large amount of annotation efforts Accuracy is not high

III. PROPOSED METHODOLOGY

The proposed system is about designing a reading assistant for visually impaired. The components used are raspberry pi 4 model B, camera, push switch, earphones and battery. Raspberry pi 4 model B of 8GB RAM, used as processor due to its processing speed. Push switch is connected to control the starting of the project as the process is executed only when the user needs it. Camera helps in better capturing of images and the output audio could be heard through the earphones. Fig.1 depicts the system's functioning.

A. Image Pre-processing

The image is taken using a camera and it is then assessed in Raspberry pi [5] with the help of python library OpenCV. OpenCV i.e. Open-Source Computer Vision Library is mainly used for computer vision applications. This Python library is a powerful tool that can perform image processing, object detection and image recognition tasks. It preprocesses the captured text to have a better extraction. OpenCV also performs various image processing tasks like cropping, resizing etc. This also involves converting the captured text to grey scale and then converting it to binary scale for better extraction by OCR [11]. In this the OpenCV helps in live capturing of the image. OpenCV initializes the camera and captures the image in front of the camera.

B. Character extraction

The text extraction is done using Tesseract OCR [7],[11]. It provides a user-friendly interface and uses sophisticated algorithms in analysing images and extracting the text even from complex and low-quality images accurately. The OCR

mainly works by two algorithms: Pattern recognition and Feature extraction. The closest match is identified by the pattern recognition algorithm, that analyses each scanned unit pixel by pixel over a database of known fonts. And in feature a more detailed comparison occurs where each curve and corners are scanned to get a more precise match. After finding out the numbers, letters and other symbols these outputs are then post processed. Which involves comparing the extracted text output with a dictionary so that only the meaningful words or those words found in the dictionary dataset are given as the final output.

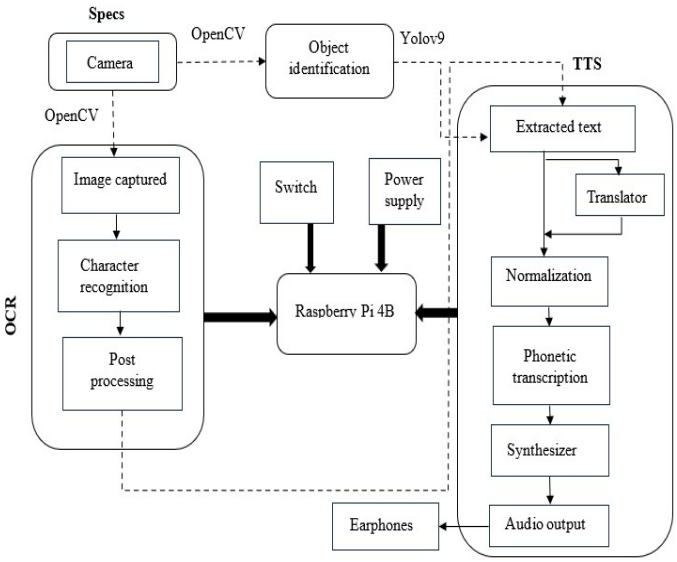

Fig. 1 Proposed system block diagram

C. Text-to -Speech synthesis

The recorded text is transformed into audio by deploying text-to-speech technology [10]. For conversion of text to speech it follows steps like normalization, phonetic transcription and synthesizing as shown in Fig. 1, after all these the audio output is obtained. In normalization each word, character and number are taken as tokens and are represented into written out words. For example, 'Mr.' is converted into 'Mister' and '12' is converted into 'twelve'. After this, the phonetic representation of each word is made. The smallest unit of sound that distinguishes a word's pronunciation and meaning from another is called a phoneme. In this way the phonetic representation of each word is generated with the help of an internal dataset. Also, pronunciation rules are applied. After pronunciation is determined prosody is generated. The prosody is which involves rhythm, pitch and stress in speech. It determines which part of the text should be emphasized. Prosodic elements such as phrasing, amplitude modelling and duration modelling (which includes the length of syllable, duration of sound, and duration of pauses) influence how natural a TTS [1] system sounds. After this the synthesizer coverts this into waveforms adjusting the amplitude, frequency etc. there by producing a more human like audio output. Through the earbuds connected to audio port of Raspberry pi, the text's audio could be heard.

D. Translate

This system also helps in translating text [13] to native language, here it is Malayalam. The google trans library in Python provides a simple way to use Google Translator's functionality within Python scripts. It's an unofficial library that acts as a Python wrapper for Google Translate. Users provide text they want to translate and specify the target language. The library constructs HTTP requests, utilizing the translation service's API, sending text and language parameters. The translation service processes the request and returns translated text or results. Here the English text is translated to Malayalam. The translation [9] feature could help anyone without the knowledge of English to read and understand the text.

E. Object detection

The project also helps visually impaired individuals in object identification [15],[2] in their immediate environment. OpenCV is used for this task. YOLOv9 which belongs to the You Look Only Once family is used for this purpose. The performance of YOLOv9 on the COCO dataset exemplifies its significant advancements in real-time object detection, setting new benchmarks across various model sizes.The YOLOv9c model, in particular, highlights the effectiveness of the architecture's optimizations.It is different from other versions of YOLO mentioned in [5],[14].

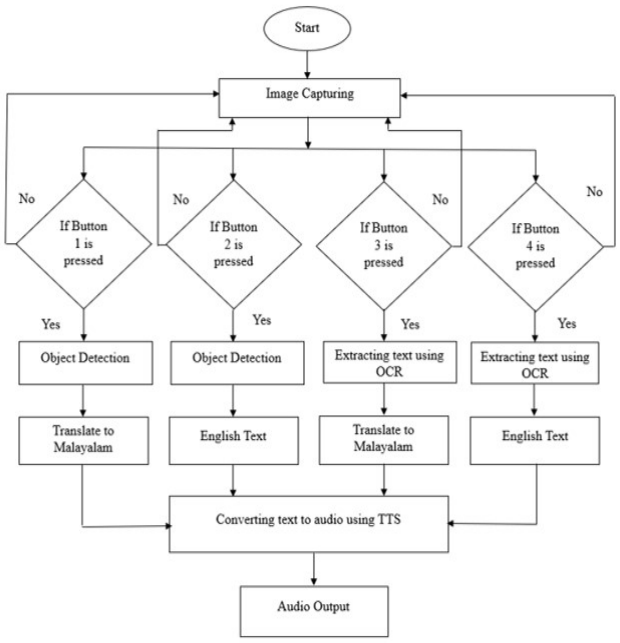

Fig. 2 Flowchart of proposed system

The YOLO algorithm works by dividing the captured image into grid cells and comparing each grid with a predefined dataset of images of various objects and gives a score named MOS (Mean Opinion Score) [4] to the grid when the any image in the dataset is found in the grid cell. This also comes with the audio output. The YOLO uses datasets which consists the list of several objects. These datasets help in identifying the objects in front of the camera by comparing the objects captured by camera with those present in the dataset. And provide an audio version of the names of these identified objects to the visually impaired individuals. For example, pen, book, mobile phone etc. In this way the blind people be aware about the objects sitting in front of them with the help of this proposed assistive device. The Fig. 2 shows the flowchart of the proposed system named iAssist.

The flowchart shown in Fig.2 gives a better overview about working of the proposed system. As the system starts the raspberry pi cam starts to capture the image. This is done with the help of the OpenCV library. Then if any oe of the 4 switches is pressed then the activity assosciated with that particular activity is executed. For example, if button 4 is pressed then text extraction is to be done. For this, the OCR tool recognizes the text present in the image and extract these texts. The extracted text is then converted to audio output providing English audio using text-to-speech engine. And if the user requires translation of the text,the button for that is to be pressed, then this extracted text is converted to Malyalam text with the help of Googletrans library. The converted Malayalam text is then obtained as Malayalam audio. This could be taken out as audio from the earphones or speaker connected to the audio jack of Raspberry Pi. And also, the when the button for object detection is pressed then yolov9 does the object detection task and the identified object's names are read out as audio.

IV. Results and Discussion

The proposed assistive device helps the visually impaired to be independent in doing tasks like reading. The python libraries like OpenCV, OCR, googletrans etc. are installed. The OpenCV helps in accessing the webcam and it auto captures the image seen. The saved image is resized and noise reduction is done. The path of the tesseract OCR is assigned and the saved image path is given to the OCR. The OCR does text extraction by converting the image to string. The extracted is obtained as the output. Here an input text image is given as illustrated in Fig. 3. The input image is a page of a text book which is having English text in it. The captured image is then processed using OpenCV and text extraction is done with the help of OCR. The extracted text from the input image is provided as the output. Fig. 4 shows the primary output which is the extracted text obtained from the input image. The audio version of this text is then provided in English using TTS.

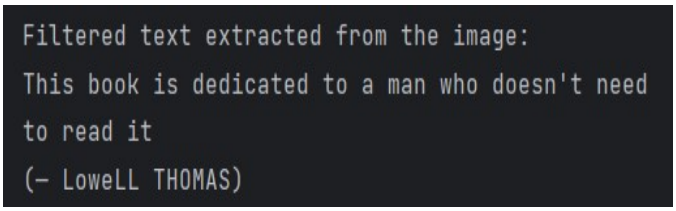

Fig. 4 Extracted text from image

And then the extracted text is converted into Malayalam text with the help of Google trans library. The translation of English text to Malayalam will help the users for better understanding of the text in front of the them. The translated Malayalam text is shown in Fig. 5. The translated text is then spoken out as Malayalam audio.

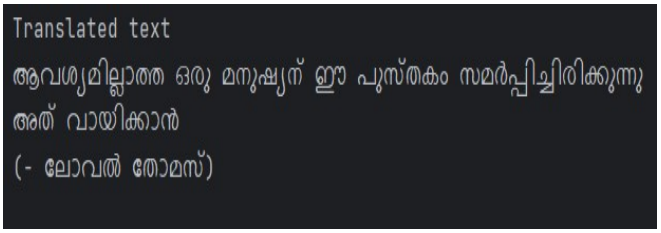

Fig. 5 Translated text.

The input image for object identification is shown in Fig. 6 The YOLOv9 identifies the images present in the image and list out them as shown in Fig.7. Hence,the proposed system helps the blind in being independent in reading text in reading the ordinary text. Also, this technology helps bridge the accessibility gap, allowing the blind users to access a wide range of written content that would be otherwise inaccessible to them. This reading assistant could save time and effort compared to traditional methods of accessing printed materials

Fig. 6 Input image for object detection

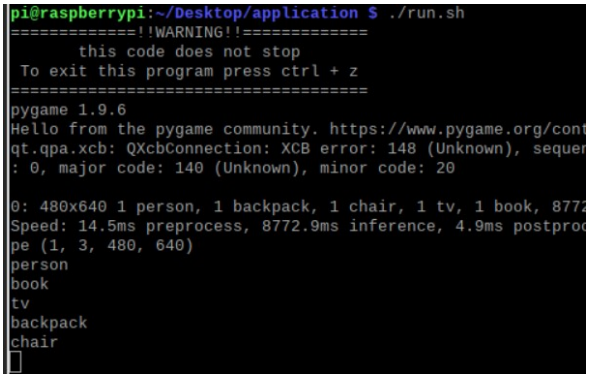

Fig. 7 Identified objects

V. CONCLUSION

The development of the proposed assistive technology solution marks a significant stride towards enhancing the autonomy and daily lives of individuals living with visual challenges. By integrating Optical Character Recognition (OCR) technology for text reading and computer vision algorithms for object detection, this system aims to address fundamental struggles that the community of blind people encounters. The implementation of OCR coupled with text-to-speech synthesis enables users to access a diverse range of printed and handwritten materials, fostering independence in education and work. Moreover, the object detection component empowers individuals by providing auditory feedback about objects in their immediate environment. Yolov2 enhances the functionality of the device by promoting greater independence and safety. The proposed system is a wearable glass for blind which could be easy and convenient to use. The simplicity and portability of the system, comprising a camera, Raspberry Pi, power bank, and earphones, ensure its accessibility and usability for a wide range of users, including elderly individuals with visual impairments. And the features like translation to Malayalam could be very useful for people who lack the knowledge on English language.

VI. FUTURE SCOPE

In future iterations or expansions of this project can be done by continuously training the system to recognize a broader range of objects and their specific characteristics. Implement machine learning algorithms that learn from user interactions, adapting to individual preferences and environments. Include real-time scene description and face recognition. Implement system which could read text in any language and also provide translation of text from any language to another language as per user needs. Also include bar code reader which could create a better shopping experience for blind. Integrate voice commands to facilitate hands-free operation, allowing users to navigate menus and functionalities using voice prompts. Ensure adherence to international accessibility standards, facilitating adoption in various regions without barriers. Explore cloud-based solutions for seamless updates, data storage, and additional functionality. By focusing on these future improvements, this project can continue to evolve, providing even greater support, autonomy, and inclusivity for the visually impaired community.

ACKNOWLEDGEMENT

We would like to thank the Department of Electronics, Toc H Institute of Science and Technology for their guidance and proofreading.

REFERENCES

[1] P. Viet, D. L. Duy, V. A. T. Thi, H. P. Duy, T. V. Van and L. B. Thu, "Towards An Accurate and Effective Printed Document Reader for Visually Impaired People," 2022 14th International Conference on Knowledge and Systems Engineering (KSE), Nha Trang, Vietnam, 2022, pp. 1-5, doi: 10.1109/KSE56063.2022.9953768.

[2] 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.

[3] T. Saeki, S. Takamichi and H. Saruwatari, "Incremental Text-to-Speech Synthesis Using Pseudo Lookahead With Large Pretrained Language

Model," in IEEE Signal Processing Letters, vol. 28, pp. 857-861, 2021, doi: 10.1109/LSP.2021.3073869.

[4] Y. Choi, Y. Jung, Y. Suh and H. Kim, "Learning to Maximize Speech Quality Directly Using MOS Prediction for Neural Text-to-Speech," in IEEE Access, vol. 10, pp. 52621-52629, 2022, doi: 10.1109/ACCESS.2022.3175810.

[5] M. A. Khan, P. Paul, M. Rashid, M. Hossain and M. A. R. Ahad, "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind," in IEEE Transactions on Human-Machine Systems, vol. 50, no. 6, pp. 507-517, Dec. 2020, doi: 10.1109/THMS.2020.3027534.

[6] L. Xue, F. K. Soong, S. Zhang and L. Xie, "ParaTTS: Learning Linguistic and Prosodic Cross Sentence Information in Paragraph-Based TTS," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2854-2864, 2022, doi: 10.1109/TASLP.2022.3202126.

[7] R. Prabha, M. Razmah, G. Saritha, R. Asha, S. G. A and R. Gayathiri, "Vivoice - Reading Assistant for the Blind using OCR and TTS," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 01-07, doi: 10.1109/ICCCI54379.2022.9740877.

[8] S. Karthikeyan, A. G. S. de Herrera, F. Doctor and A. Mirza, "An OCR Post-Correction Approach Using Deep Learning for Processing Medical Reports," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 5, pp. 2574-2581, May 2022, doi: 10.1109/TCSVT.2021.3087641.

[9] V. Adusumilli, M. F. Shaik, N. Kolavennu, L. B. M. T. Adepu, P. A. V and I. R. Raja, "Reading Aid and Translator with Raspberry Pi for Blind people," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 327-331, doi: 10.1109/ICACCS57279.2023.10113042.

[10] J. Park, E. Lee, Y. Kim, I. Kang, H. I. Koo and N. I. Cho, "Multi-Lingual Optical Character Recognition System Using the Reinforcement Learning of Character Segmenter," in IEEE Access, vol. 8, pp. 174437-174448, 2020, doi: 10.1109/ACCESS.2020.3025769.

[11] A. Jamal, L. Aljiffry, N. Alhindi, R. Nahhas and S. Al-Amoudi, "Image Assistant Tools for Extracting, Detecting, Searching Images and Texts," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-6, doi: 10.1109/CAIS.2019.8769582.

[12] T. Shah and S. Parshionikar, "Efficient Portable Camera Based Text to Speech Converter for Blind Person," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2019, pp. 353-358, doi: 10.1109/ISS1.2019.8907995.

[13] R. Dhar and S. Mukherjee, "Android-based Text Reader for Partial Vision Impairment," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, India, 2018, pp. 1-5, doi: 10.1109/UPCON.2018.8597074.

[14] K. Patil, A. Kharat, P. Chaudhary, S. Bidgar and R. Gavhane, "Guidance System for Visually Impaired People," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 988-993, doi: 10.1109/ICAIS50930.2021.9395973.

[15] I. H. Shanavas, P. B. Reddy and M. C. Doddegowda, "A Personal Assistant Robot Using Raspberry Pi," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 2018, pp. 133-136, doi: 10.1109/ICDI3C.2018.00038