PROVISIONAL SPECIFICATION

Field of the invention:

Children with slight or severe impairments of any physical organ or children who are mentally unstable are given special talents to aid them in absorbing the normal life process. According to UNESCO's report on children with disabilities, 27% of impaired children between the ages of 5 and 19 do not have access to education. We consider providing them with a supportive and growing environment where they can live and grow. In the educational ecosystem, the needs of such an environment need to be catered according to the category of ailments that exist in the students and provisions need to be made to make them feel special, and cared for and set the right mindset while they grow.

Vision plays a major role among all senses. With just little aid to the visually impaired people existing currently, there is a need to implement a device that aids them in their educational purpose. The goal is to provide inexpensive solutions to the visually impaired and make their lives better and more self-sufficient. Techocular is made to aid school children with total or partially impaired eyesight without needing a guardian to improve their knowledge. The configuration of hardware and software units is to be designed, which leads to the recognition of texts from the given source and converts it to audio input for the user. We present a method, which uses Optical Character Recognition on the live stream of images and Natural language Processing for further conversion for expected output.

Background of the invention:

The camera used in the device is capable of autofocusing and capturing the image of whatever is placed in front of it. We access the camera by using voice commands recognized by the microphone and the program executed in the Edge computing device. Further, we get audio output

via headset or speaker. A general-purpose input/output (GPIO) pin is used where the user can simply switch on the device, and then give a voice command to take a photo, then the contents will be extracted and output is given to the person as audio, where they can Interrupt to read it again, pause, stop and continue. Our proposed work can access real-time inputs and give instant and accurate output. Students can get benefitted using our device anytime.

Invention:

 A portable device that acts as a Reading assistant for Visually Impaired people using technologies like Natural Language Processing NLP, Speech recognition, and Optical Character Recognition OCR, which takes voice commands as input and gives audio output instantly, recognizing and converting the text in the given source into machine readable electronic version of the text.

A detailed description of the invention:

Purposed Model:
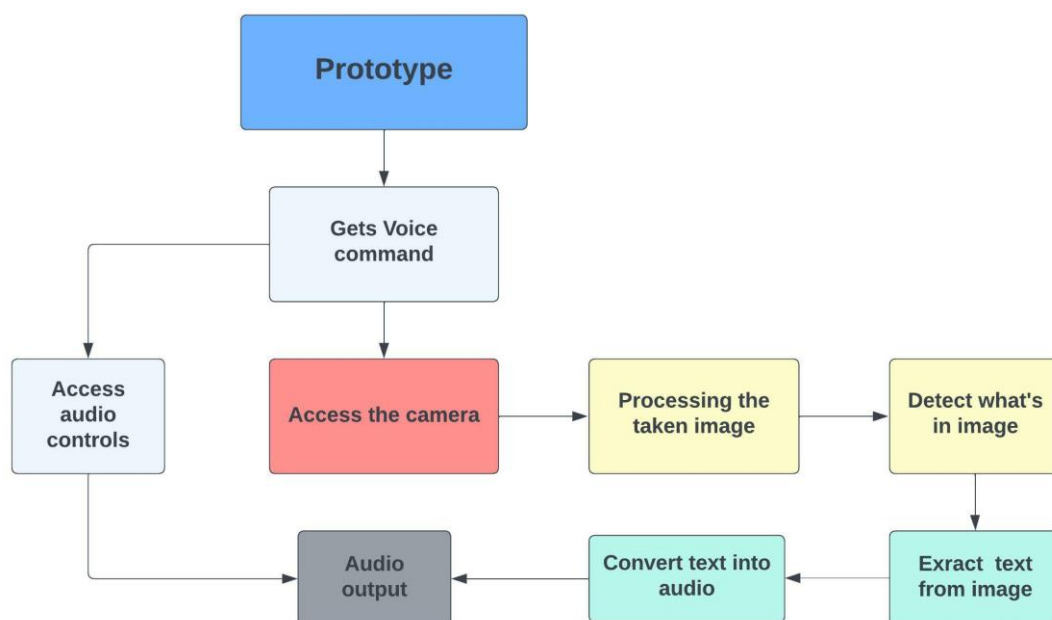
Entire Workflow

1. Abstract:

For individuals with impaired vision, accessing study resources is very difficult which makes them get out of the natural education system. They need the education delivery mechanism to be so engaging that despite the ailments, children can reflect on their learnings and grow. Consciously improving Education for especially abled children will need planned efforts with the help of technology and human power. Our proposed work provides them with a learning environment, where they can access the device themselves without any guidance and improve their knowledge. Our model is a portable device that gives visually impaired students reading assistance which can be accessed by simple voice commands since voice communication is the only effective way for the visually impaired. Technology like Artificial Intelligence is rapidly implemented in every system to enhance the process. Artificial intelligence-driven speech recognition processing is the method of making software accessible using human input language. We use this technology to make users interact with computers or devices. Getting simple voice commands from the users through microphones and accessing the camera which has been set up. Pre-customized simple voice commands have been incorporated into the model. The picture which has been taken by the camera is processed using Optical Character Recognition (OCR). It is a technological method for identifying text in digital images. Text in scanned documents and photos is frequently recognized using this technique. A real paper document or an image can be transformed into a text-rich, accessible electronic version using OCR software. Then the converted text is stored in a separate file. Using text-to-speech conversion, the stored text is given as audio input to the user. We use pygame-mixer the audio can be accessed, and it can be paused, stopped, and continued according to the user's command. To inform the user that the content has been read, we attached the page completed command. If no content is found in the given resource/book, the user will be intimated that there is no text found. Our device can access multiple languages and give instant output. This efficient product runs effectively. It can be

effectively used by visually impaired people to enhance their knowledge and independent access, which gives them support for the educational purpose in a natural flow.

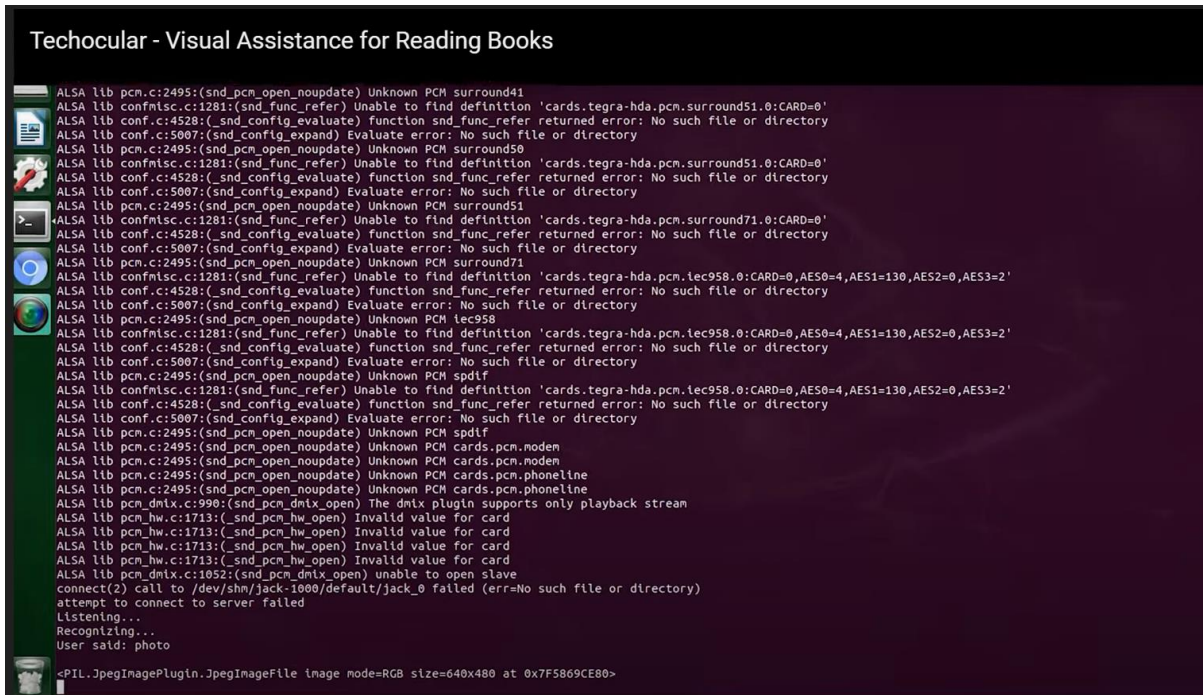# 2. <u>Setting up the device:</u>

The developed product has an edge computing device- Jetson nano playing a major role. The microphone is used to recognize the voice and it is connected to the jetson nano which accesses the camera and takes pictures. The model can be static where the camera can be fixed to the stand and the books are placed under or it can be a portable mobile one where the whole setup is made as a wearable device like normal glass. The output is given via headphones or through a speaker which has been connected. The power supply is given through a Power bank to make it portable and simple. Wi-fi module and Bluetooth module are used.

3. <u>Voice command and Natural Language Processing:</u>

It captures the image of the object with the user's voice command ("Take a photo") which is recognized by the microphone and saves the image in jpeg format.

➢ The user is instructed to give pre-customized voice commands such as "Take a photo", "Play audio", "pause", "Continue", "Replay", and "Stop".

➢ The audio signals are recognized by the microphone. The sentences are recognized by using the SPEECH RECOGNITION module.

➢ The recognized words are being understood by the machine using NATURAL LANGUAGE PROCESSING where the input words are. If the customized phrases are recognized, for example, "Take a photo", the camera automatically takes the picture and it saves in jpeg format.

➢ If the words are not been recognized properly the machine replies with the response command "Say that please again" by using SPEECH SYNTHESIS and it waits for the voice command again.

4. Image processing and Text extraction:

Using OCR (Optical Character Recognition) module the text can be extracted from the image which has been scanned.

➢ The text from the saved picture is recognized and extracted by using Tesseract and OCR (Optical Character Recognition) module. It follows the matrix scanning and recognizes character by character.

➢ If the text is present in the paper or document which has been scanned, then the text will be saved in a separate text file. If the text is not present, it responds "no text found".
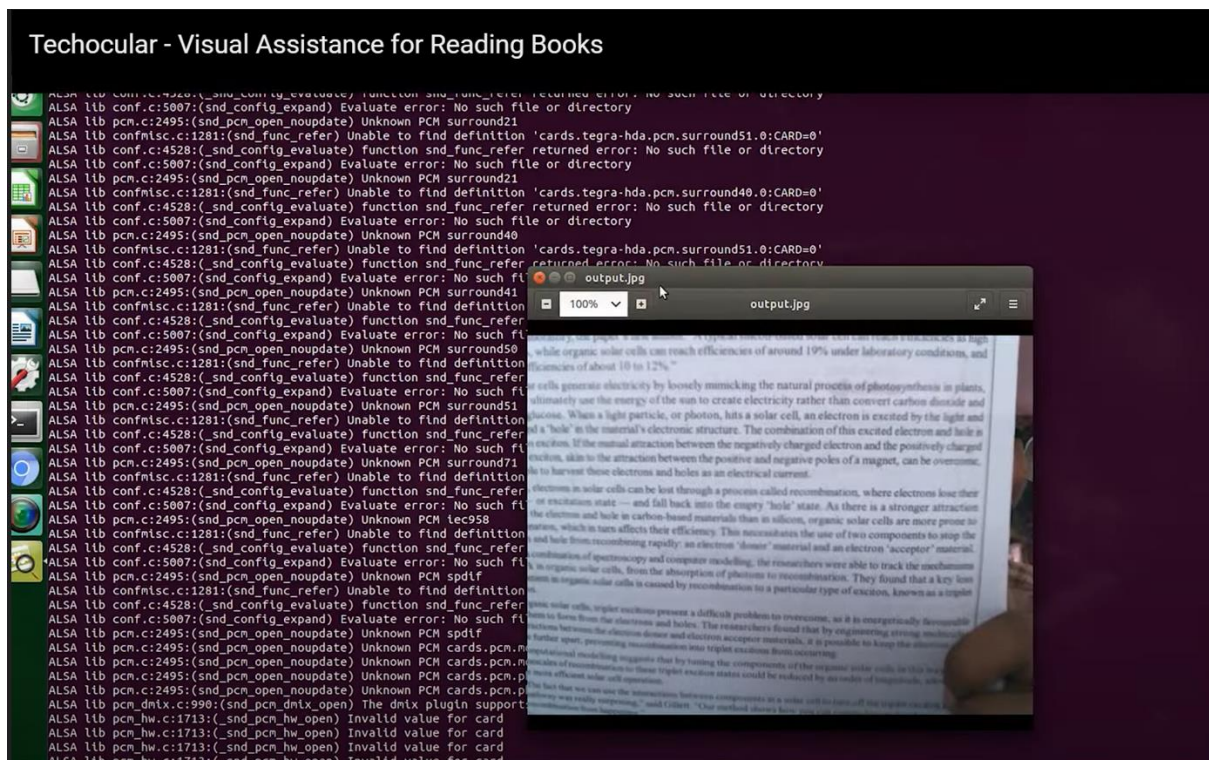
5.<u>Text-to-audio conversion:</u>

Text which has been extracted is given as output through a speaker or headset using Google Text-to-Speech (TTS) and Pyttsx3 module.

➢ If the input voice command is "Play audio", by using GOOGLE TEXT-TO-SPEECH conversion the text is converted into an audio file.

➢ By using the pygame-MIXER the audio can be played, paused, continued, or can be stopped by which the customized commands are being mapped to the outputs.