iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45740

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Machine Learning Based Text to Speech Converter for Visually Impaired

Prof. Vidhyashree. C[1], Supriya. A. M[2], Supriya. H[3], Vedala Dinesh[4], Kavya. R[5]
Electronics and Communication, Dr.Ambedkar Institute Of Technology, Bangalore
[1]Guide.Assistant Professor, Dept.Of ECE, Dr. AIT, Bangalore
[2, 3, 4, 5]Student. Dept.Of ECE, Dr. AIT, Bangalore

Abstract: ***This project aims to help visually impaired persons sleep in the modern environment, regardless of their disability. To make the material easier to read and understand, we're going to create a standalone text-to-speech engine. Take a snapshot of the printed text using a USB webcam, and then look at the image using optical character recognition (OCR). The recognised text is converted into native-language voice using a text-to-speech algorithm. It is suitable not just for the visually impaired but also for the general public who want to read the content aloud as rapidly as possible. With the use of local language translation, this project attempts to let persons who are blind read typed English documents.***
Keywords: Optical character recognition(OCR), Text-to-Speech, Raspberry Pi, Finger gestures, Camera.

I. INTRODUCTION

In the form of camera-based gadgets that incorporate computer vision technology with already-available commercial products like OCR systems, these folks can gain from recent advancements in computer vision, digital cameras, and portable computers. Accessing written documents can be challenging for blind people in various situations. When accessing text while on the go and in less than optimal circumstances. The purpose of this technology is to enable blind people to touch textual content and instantly hear audio output. A computer-based technology called the Text-to-Speech (TTS) Synthesizer may be used to. Text information extraction is a crucial component of OCR since it affects how understandable the output speech is, making it a large and crucial function of any auxiliary reading system. Visually impaired students are a diverse collection of people who have different reading and learning challenges, require error-free reading, and are placed in an environment that is highly typical or well-established. helps you after you do to attain good academic performance. It also demonstrates how challenging it is for those who are blind to succeed in both education and the workforce. We suggested eyewear with a built-in camera for the blind in order to solve this social issue. useful for interpreting photos that have been captured in the manner of audio output. Additionally, even regular people are capable of quickly reading lengthy documents. In order to provide the ability to scan and detect text, such systems integrate with Character Recognition Through Optics software. Some devices have an audio output built right in. No other product, as far as we all know, offers the same ease of use, accuracy, and high performance for those who are blind. greater expressiveness of synthetic speech and improved text-to-speech quality. I divided the video block into text and non-text regions using automatic teletext detection and extraction. This is now achievable thanks to recent advancements in computer vision, digital cameras, and computers as well as the appearance of computer vision-based camera solutions.

II. OBJECTIVE

Text-to-speech technology is that the process of talking to a computer. A variety of techniques can be used to improve text-reading software. The solution to this issue is the development of finger reading techniques that do away with previously created and saved datasets and offer a pre-response to read the text provided as the input capture image.

Our project's primary objective is

1) Developing a framework that uses machine learning to assist the visually impaired.
2) Development of a module that captures images of text supported fingertouch.
3) Design of a module that extracts text from a picture, converts the extracted text to voice, and so converts it into an area language.
4) Design a TTS module that converts text to speech.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

III. PROBLEM STATEMENT

There are many challenges in translating archaic writing styles into voice for visually handicapped individuals. There is a need for a device that enables successful voice-based computer interaction for blind or visually impaired people. The issue is that the system's fundamental design is an embedded one that takes an image, extracts a text-containing image from it, and then converts that text into the necessary language. Use a series of procedures for image processing to search out the text and take away the background. one in every of the drawbacks here is that edge detection can misidentify some alphabets, which may cause incorrect output generation by OCR.

To enable people to be more effectively committed local and / or remote services, Text-to-Speech (TTS) provides computer-generated speech to the text that users "read". It absolutely was first developed to assist the visually impaired. In this project, text to speech conversion will be seen and heard. Optical character recognition is used. We advise using an Android or Raspberry Pi camera that can connect via Ethernet, Bluetooth, or Wi-Fi to take pictures. The image is captured by the webcam or Raspberry Pi camera in high resolution, and is then saved on the Raspberry Pi system. Use Python's Tesseract OCR package to convert the captured image to text. An image is typically captured and processed by cameras in between 5 and 7 minutes. The taken image is translated to text on the Raspberry Pi's Winchester drive using Python's Tesseract OCR package. The TTS system receives the transformed text and formats it according to the language.

IV. BLOCK DIAGRAM AND DESCRIPTION

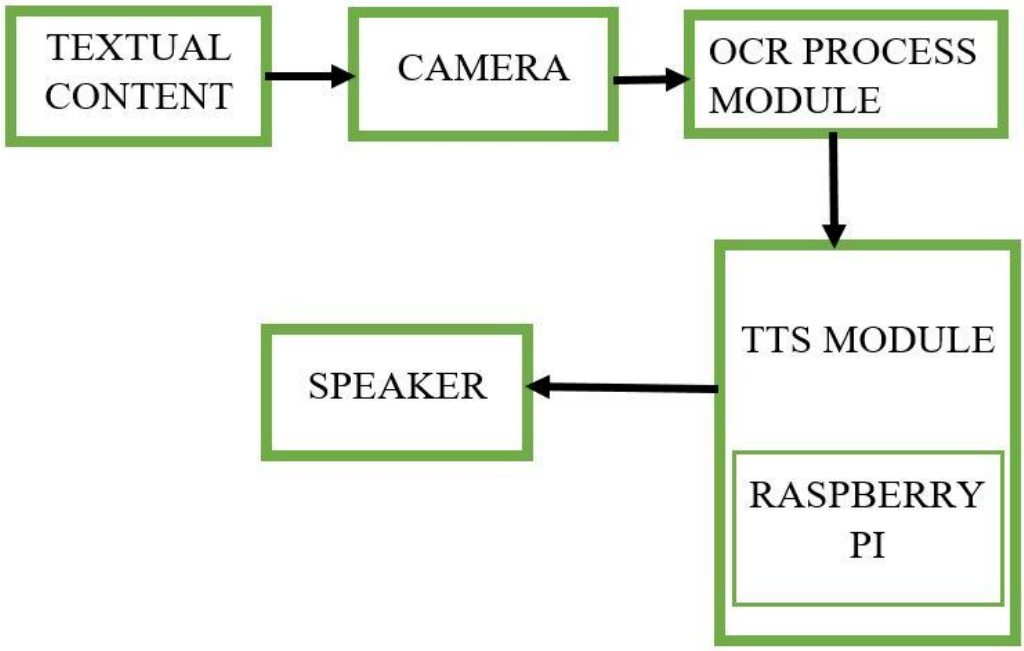


FIG : 1 - Proposed Project Block Diagram

A. Working

The finger module has an integrated camera that can take pictures and turn them into text . It then converts the text into speech using speech synthesis algorithms and offers local language translation for those who are blind or visually challenged to understand other languages.

B. Finger Reader Module

1) In this prototype, the IR sensor for finger detection is connected to the Pi board by the camera's connection to the RaspberryPi.
2) Use a wrist-mounted Raspberry Pi board with a voice-enabled text reader to accommodate the visually impaired
3) TTS-style headphones allow you to regulate haptic feedback.
4) On a laptop computer, the camera's video output will be processed in real time.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Optical Character Recognition

1) The optical character recognition engine may convert an image from input, handwriting, or printed text to machine-encoded text mechanically or electronically.

2) This is a standard method for converting printed text into an electronic format for use in text-to-speech systems and other automated procedures.

3) The input is entered as text using a camera that can be manipulated with a finger device. The OCR algorithm then speaks the text aloud after being photographed by the camera and delivered there.

D. TTS Module

1) Conventional text is converted to speech via text-to-speech (TTS) systems, while some systems also perform the same for symbolic language expressions like pronunciation notation.

2) It has a front end and a back end. There are two primary functions for the frontend. First, it transforms unprocessed text that contains symbols like numbers and abbreviations into the equivalent of spelt out words. The frontend then separates the text into prosodic units like phrases, phrases, and sentences and marks the content by giving each word a pronunciation notation.

V. HARDWARE AND SOFTWARE REQUIREMENTS

A. Hardware Requirements

1) Raspberry pi 3b+

2) USB Camera

3) Adapter

4) IR sensor

5) Headphone

6) Laptop

B. Software Requirements

1) Raspbian OS

2) Python software

3) VNC viewer

VI. HARDWARE AND SOFTWARE IMPLEMENTATION

A. Software Analysis

Flow Chart

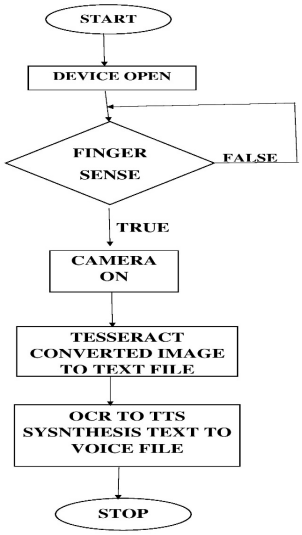


FIG :2-FLOWCHART

1) When the application launches, the device powers on and activates the IR sensor used in the finger sensor module.
2) If the module responds, the image will be captured by the USB camera. Otherwise, it will return to the initial state.
3) Next, the text is extracted from the image captured using OCR-Tesseract and saved in a file.
4) Finally, the text in the file is translated into the language you need using the text-to-speech converter.

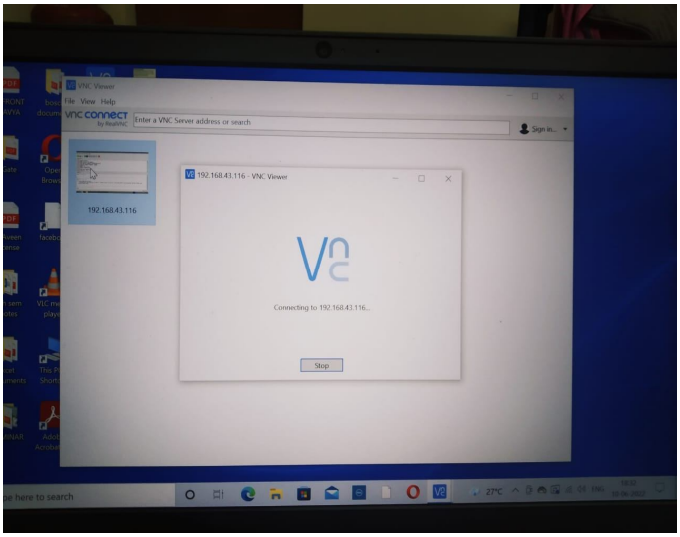


FIG:3-VNC VIEWER

With the help of the VNC viewer and WiFi, the raspberry pi's display is displayed on the computer screen.
Steps for configuring a Raspberry Pi over WiFi

a) Set up the OS on your SD card.
b) Download: Ssh & WPA-Supllicant
c) Edit the Name and Password of your Wi-Fi Router in Wpa-Supplicant.
d) Then copy additional documents into your SD card.
e) Connect a 5V charger to your Raspberry Pi and insert a Micro-SD card.
f) Open your router and load a page in your browser.
g) You can find the Raspberry Pi IP address.
h) Copy this IP address and any subsequent ones in VNC viewer.
i) Press open to visit the open command window after that. Therein Type Raspberry Pi Login & Password.
j) Following that, launch Terminal Server Access from the start menu.
k) Click Connect after specifying the IP address of the Raspberry Pi.
l) On this website, kindly provide your username and password. Pi is the user name, while raspberry is the password.
m) The Raspberry Pi Screen may now be viewed on a laptop.

B. OCR Process Using Mobilenet Café Model (SSD Algorithm)

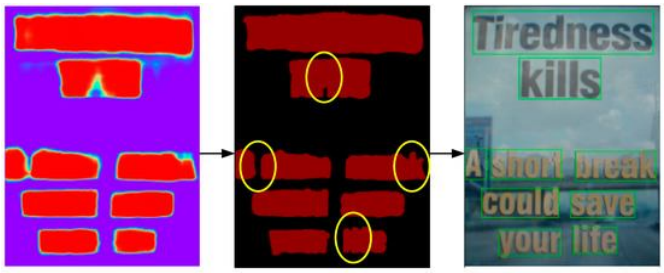


FIG :4 - Flowchart for the semantic bounding box module

CRF preparation and bounding box look are the two modules found in the SBM. The problem of boundary obscurity and stickiness in content semantic division is understood by the CRF handling. According to the CRF preparation results, the bounding box look yields the best content semantic division bounding box. When words are close together, it is simple to create grip, which is one flaw in the semantic division conclusion. The semantic division result's pixel-level forecast is corrected using the CRF calculation. After CRF preparation, content edges are more refined and less sticky. Maps of noisy division have been tamed with CRF. These algorithms favour same-label assignments to spatially nearby pixels by coupling nearby hubs via short-range CRF. Instead of attempting to smooth over the neighbourhood structure throughout this effort, the goal should be to recover it. In this way, the CRF show is coordinated with our organisation in its entirety. Two modules make up the SBM: CRF preparation and bounding box look. The CRF handling is aware of the issue of border stickiness and obscurity in content semantic division. The optimal content semantic division bounding box is obtained using the bounding box look, according to the CRF preparation results. Because it is straightforward to create grip when words are close, the semantic division result has this flaw. To correct the pixel-level prediction of the semantic division result, the CRF calculation is used. Following CRF preparation, content edges are better sharpened and less sticky. The rowdy division maps have been tamed using CRF. The same-label assignments to spatially close-by pixels are preferred by these techniques, which use short-range CRF to couple nearby hubs. In this work, regaining the intricate neighbourhood structure should be the goal rather than simply assisting it. In this way, our organisation coordinates the fully associated CRF exhibition.

C. Hardware Setup

The idea on hardware design for Machine learning based to text to speech conversion is shown in below figure.



FIG:4-Hardware Setup

VII.ADVANTAGES AND APPLICATIONS

A. Advantages

1) Reduce operating costs. this implies more production.
2) Improves operational efficiency, health and safet
3) Remote inspection during operation.
4) Increased on-the-job learning
5) Improved word recognition skills, fluency and accuracy.

B. Applications

1) Aerospace and Avionics

Goggles and optical head-mounted displays make the utilization of goggles and optical head-mounted displays very convenient within the aerospace and aero electronic engineering industries. Here, the foremost demanding operations and maintenance are possible at the nano and micro levels. With smart glasses, you'll easily give virtual instructions.

2) Atmospheric Survey

Smart glasses allow the wearer to survey the visual patterns of things within the atmosphere. you'll be able to identify the parameters that affect the environment.

3) Navigation and Travel Experience

Navigation experience will be improved with the assistance of smart glasses. you'll be able to easily identify the situation map and find the shortest and safest route. By integrating the traffic management system with smart glass, travellers can estimate and visualize the time required.

4) Games

Augmented reality and computer game features of smart glass and optical head-mounted displays facilitate your experience the gaming experience.

5) Entertainment

The Entertainment section includes movies, news, and more. Users can experience the entertainment they have, like color adjustments, language changes, and voice-controlled film experiences.

VIII. RESULTS AND OUTPUT ANALYSIS

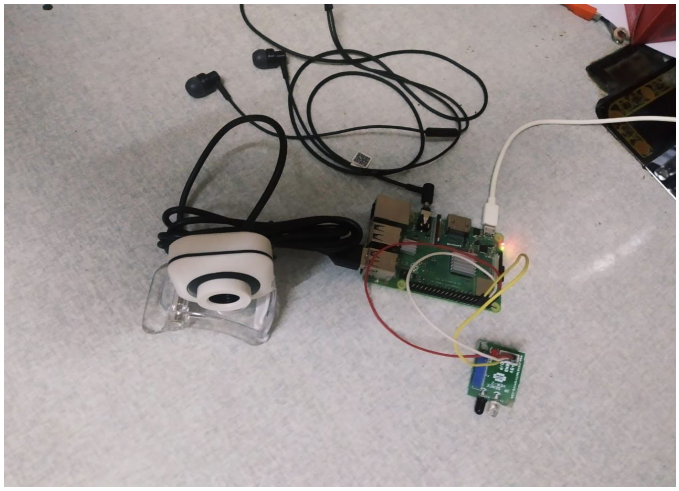


FIG :5-RASPBERRY PI WITH CAMERA

Tesseract is being used in this instance to transform the image into a text file, and TTS is being used to process the text into voice.

1) The figure says that we are capture image by focusing the image.
2) The device (Raspberry pi) converted image file into text file.
3) The OCR processing the image file to text. File.
4) Earphone is used for listening of voice processed text file.

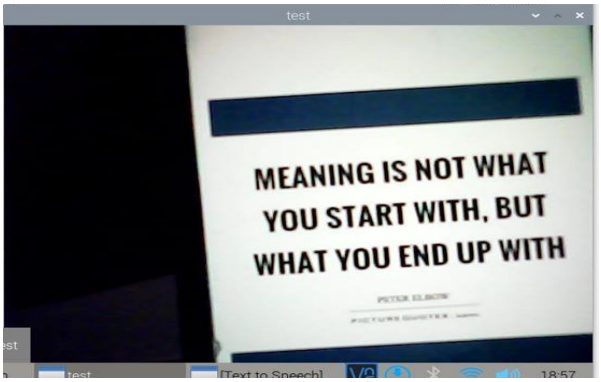


FIG :6-INPUT IMAGE

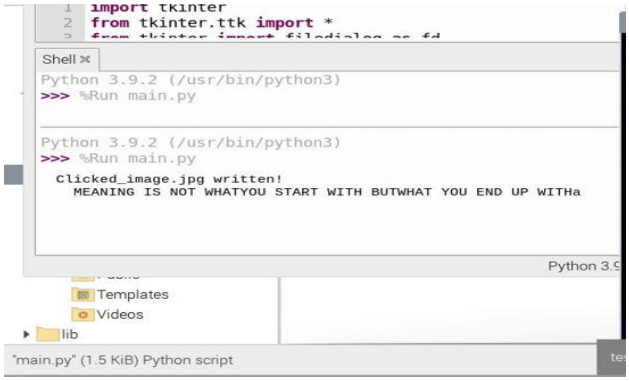


FIG :7-OUTPUT WINDOW

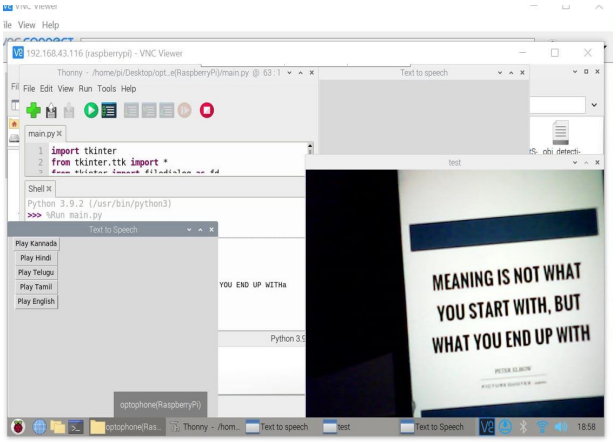


FIG :8 Language Translation

IX. CONCLUSION AND FUTURE SCOPE

A. *Conclusion*

The raspberry pi is used in this system implementation to execute a text-to-speech conversion that is entirely knowledge-based. Men and women who are visually handicapped can focus on the document's text without any trouble. By employing translation software, one can change the text's language, and then with the aid of a TTS voice recognition device, one can turn the new text into audio. With the use of a raspberry pi, we have distributed a text to speech conversion method. After the simulation outcomes were successfully verified, the hardware output was examined using distinct samples. The image is efficiently approached and properly read using our set of principles. By expanding these systems, human and computer courtship will get a lot closer. Thus, it makes it possible to get over the obstacle of the virtual gap.

B. *Future Scope*

The creation of a tool that performs beholding rather than still photos and extracts text from video could be part of future development. We can focus on creating effective tools that accurately translate text from photos into speech. The automatic area trimming and scaling of text regions using computer vision may be included in future work, as well as the use of image processing along with a computer vision library to recognise the components in a picture. Text recognition might be an upcoming duty in addition to the advanced handwritten text recognition that is now being done.

REFERENCES

[1] R. Ani; Effy Maria; J. Jameema Joyce; V. Sakkaravarthy; M.A Raja, “Smart Specs: Voice assisted text reading system for visually impaired persons using TTS method”, 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), 2017, pp. 1-6, DOI: 10.1109/IGEHT.2017.8094103.

[2] Chitra S ,Balaji N, “Enhanced portable text to speech converter for visually impaired” ,January 2018 International Journal of Intelligent Systems Technologies and Applications 17(1/2):42. DOI:10.1504/IJISTA.2018.091586

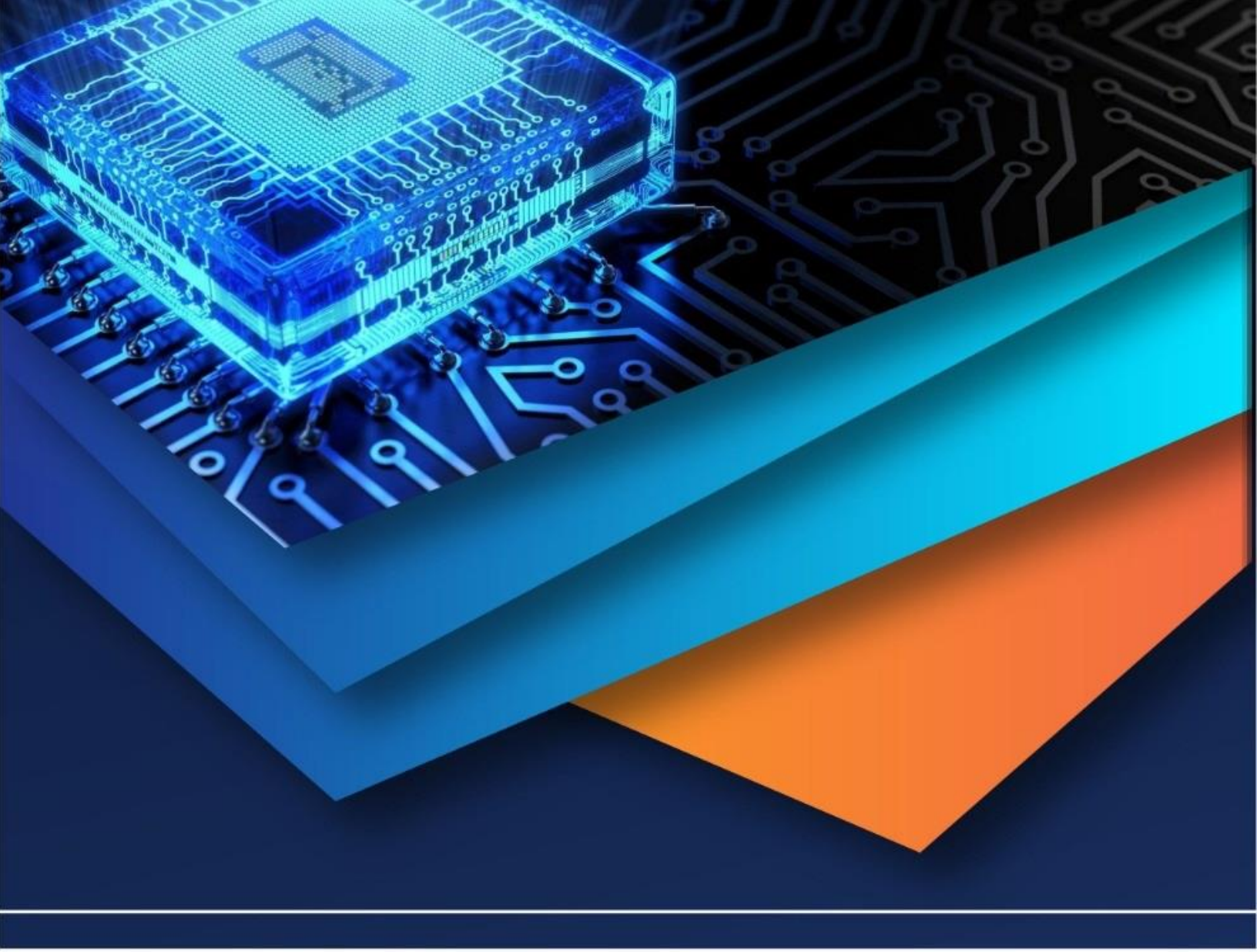
[3] Krithi sharma, “Enhanced portable text to speech converter for visually impaired”, November 2020 International journal of machine design and manufacture.

[4] Chaitali S. Hande, Vijay R.Wadhankar , “Voice Assisted Text Reading and Google Home Smart Socket Control System for Visually Impaired Persons”, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 06 | June 2019.

[5] Sanjana.B , J.RejinaParvin , “Voice Assisted Text Reading System for Visually Impaired Persons Using TTS Method”, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 6, Issue 3, Ver. III (May. - Jun. 2016), PP 15-23 e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197

[6] Hawra Al Said ;Lina Alkhatib ; Aqeela Aloraidh ; Shoaa Alhaidar , “Smart Glasses for Blind people”, Spring 2018/2019

[7] Sharvari S, Usha A, Karthik P, Mohan Babu C,” Text to Speech Conversion using Optical Character Recognition”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 07 | July 2020



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)