iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: V Month of publication: May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62551

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Haptic Summarizer for Visually Impaired Using Rasberry PI

Gauri Dhanorkar[1], Minal Zope[2], Sanika Mane[3], Rudra Pingle[3], Ramit Desai[4]
AISSMS Institute of Information Technology, Pune 411001, Maharashtra, India

Abstract: The project endeavours to ameliorate the constraints inherent in conventional reading modalities for visually impaired individuals through the development of a self-contained reading apparatus. This device is designed to facilitate access to printed textual material while augmenting its capabilities with image captioning functionality, thus extending its applicability beyond mere textual content.

To achieve this objective, the project will leverage sophisticated deep learning methodologies, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and ResNet architectures, for image analysis and caption generation tasks.

Furthermore, optimization of the system will be carried out using stochastic gradient descent, ensuring optimal performance and efficacy in practical usage scenarios.

Additionally, the device will include voice output capabilities for delivering prompts, notifications, and user interaction feedback, enhancing the overall usability and accessibility for users with visual impairments.

Keywords: Text-to-speech, Deep learning- Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), ResNet, Stochastic Gradient Descent.

I. INTRODUCTION

The project endeavours to revolutionize the reading experience for visually impaired individuals by introducing a state-of-the-art self-contained reading device powered by Raspberry Pi.

This device is meticulously designed to address the inherent limitations of conventional reading methods faced by the visually impaired community. By seamlessly integrating cutting-edge technologies such as text-to-speech (TTS) conversion, voice output capabilities, and advanced image captioning using deep learning techniques, the device not only enhances access to printed textual material but also expands its functionality to comprehend visual content.

Raspberry Pi, a small yet powerful single-board computer, serves as the core component of the reading device, providing the computational capabilities necessary for executing complex tasks with efficiency and precision. This versatile platform empowers developers to create innovative solutions tailored to specific needs, making it an ideal choice for developing assistive technologies like the proposed reading device.

The current scenario of reading for visually impaired individuals primarily relies on the Braille system, which poses challenges due to the limited availability of Braille materials. This limitation restricts the range of accessible content and hinders the independence of visually impaired individuals in accessing information. Consequently, there is a pressing need for innovative solutions that enable better manageable, eyes-free operation for reading tasks.

In response to this need, the project sets out to create a comprehensive reading device that empowers visually impaired individuals with a versatile tool for accessing and comprehending both textual and visual content. Leveraging sophisticated deep learning methodologies, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and ResNet architectures, the device excels in image analysis and caption generation tasks, thereby enriching the reading experience beyond traditional text-based approaches.

The integration of voice output capabilities further enhances the usability of the device by providing auditory prompts, notifications, and user interaction feedback. This not only improves accessibility but also fosters a seamless and intuitive user experience for individuals with visual impairments. Through meticulous optimization utilizing stochastic gradient descent, the project ensures that the device delivers optimal performance and efficacy in real-world usage scenarios. By bridging the gap between textual and visual information, this innovative reading device, powered by Raspberry Pi, is poised to make a significant impact in empowering visually impaired individuals to access and comprehend a broader range of content with greater independence and efficiency.

II. LITERATURE SURVEY

Author	Technique	Results	Limitation
"S. Srija, *P. Kawya, T. Akshara Reddy, M. Dhanalakshmi"	Miniaturized Camera, Image Processing, Raspberry Pi, Vibration Mo tors.	The Raspberry pi based wearable reader captures printed text as images and then segments the image to obtain the text alone.	proposed work ad- dresses only upward deviations from the current text-line.
"Megha P Arak- eri, Keerthana N S,Madhura M, Anusha Sankar ,Tazeen Munnavar[2008]"	Raspberry Pi, Optical Charac- ter Recognition (OCR), Object Recognition, Text to Speech, Language Trans- lation.	The proposed so- lution is to de- sign an inexpen- sive wearable de- vice that uses com puter vision to read out any form of text around the user in various alignments and lighting conditions.	A few extensions that can be added to this product are Speech In- teraction, Facial Ex- pression Recognition and Human Face Rec- ollection.
"Usman Ma- sud 1,2, Tareq Saeed3, Hunida M. Malaikha4, UL Islam1, AND Ghulam Abbas1[2022]"	Smart system, visual losses, biomedical sensor, object recognition, ten- sorflow, Viola Jones, ultrasonic sensor.	we plan to add a text-to-speech system and embed it with the GSM module so that blind person can actually hear the directions in form of voice.	Limited object recog- nition, potential false alarms, limited depth perception, and de- pendency on technol- ogy.
"Xiaochen Zhang, Bing Li, Samleo L. Joseph, Jizhong Xiao, Yi Sun andYingli Tian,J. Pablo Mun˜oz, Chucai Yi[2020]"	assistive naviga- tion; semantic path planning; SLAM; wearable device	Based on the lo- calization and the semantic digital map, the user is navigated to the desired room. The user can be guided by the audio output command to the destination easily and conveniently with usability humanistic audio interface.	Relies on depth cues, which may limit its performance in sce- narios where accurate depth information is not available.

"Amey Hengle,NirajKulkarni,Atharva Kulkarni,Nachiket Bavadekar ,Rutuja Udyawar[2020]"	Assistive Technologies, Raspberry Pi, Face Recognition, Image Captioning, Text Recognition, OCR, News Scraping	The Smart Cap acts as a conversational agent bringing together the disciplines of Internet of Things and Deep Learning and provides features like face recognition, image captioning, text detection and recognition, and online newspaper reading.	the addition of more robust hardware support like GPUs will not only improve the device's response time but also pave the way for the inclusion of faster and more accurate deep learning models. OCR can be coupled with Document Image Analysis (DIA) for getting more optimal results.
"Anmol Bhat,Aneesh C Rao,Anirudh Bhaskar,Adithya V,Prof. Pratiba D	Raspberry Pi; Object detection; Text extraction; Speech-to-text; Natural Language Processing.	Results of the different modules were generated and Analyzed for different test inputs. The object detection classifier, as Mentioned before, was trained using two models. In the initial training run, with the ssdmobilenet v1 coco model, the accuracy obtained was pretty less.	It is during this process that several essentials in the speaker's delivery are missed out.
"1 S Gayathri, K.Jeyapiriya, K.JeyaPrakash"	RFID reader, Raspberry Pi, loT, RFID scanner.result.	The merchandise is scanned by visually challenged people using RFID cards, and the audio is	The technical complexity of the system, involving Raspberry Pi, Bluetooth, and

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

		transformed into voice by the Rasp-berry Pi in figure 3, which the person hears.	IoT, couldpose challenges for maintenanceand repair, which may be particularly problematic for visually impaired users who depend on the technology.
”Simon L. GAY 1, Edwige Pissaloux 2, Katerine Romeo 2,AND NGOC-TAN Troung 2. [2021]”	propose a novel and inclusive haptic architec-ture to access and interact with 2D data, relying on the force-feedback principle, and named Force- Feedback Tablet (F2T)	Four layers achieved the highest accuracy of 98.60 in both train-ing and testing	Complex mechan-ics, potential for reduced precision, and limited adapt-ability for 3D data.
”Nathan Dunkelberger, Student Mem-ber, IEEE, Jennifer L. Sul-livan, Member, IEEE, Joshua Bradley,Indu Manickam, Gau-tam Dasarathy, Richard Bara-niuk, and Mar- cia K.[2011]”	wearable hap-tics, language communication, multi-sensory haptics, tactile device, phoneme coding.	Haptic Morse code was explored using an electrodynamic minishaker that represented dots or dashes by the du-ration of a fingertip displacement.	Spe constraints, ed for factor limita- m tion and learning s, cur challenges for ve haptic language trans-mission.

”Arthur K. K. Wong, N. K. Fong [2021]”	Automatic Flame Image Segmentation, Image Recognition, Tracking Fire Spread Direction, Flame Movement Prediction.	The multi-threshold algorithm with Rayleigh distribution analysis (modified segmentation algorithm) successfully improved flame image segmentation.	The low image resolution of 320x240 pixels may lead to inadequate fire detection for small or distant fires.

III. RESEARCH LIMITATIONS

Research on a Raspberry Pi-based wearable reader aimed at aiding visually impaired individuals through haptic feedback encounters multifaceted limitations across technical, usability, environmental, accessibility, ethical, and evaluation domains. Technical constraints stemming from the Raspberry Pi's computational limitations, including processing power and memory restrictions, pose challenges for implementing sophisticated image processing and object recognition algorithms, potentially limiting the system's accuracy and responsiveness. Moreover, the precision and adaptability of haptic feedback mechanisms present usability hurdles, as users may require time to acclimate to the sensory cues, and feedback granularity might be insufficient for discerning intricate details. Wearability concerns arise from factors such as device size, weight, and comfort, impacting long-term user acceptance and adoption. Additionally, the device's performance may be influenced by environmental factors like lighting variations and obstacle detection reliability, affecting its effectiveness across diverse scenarios. Accessibility issues emerge concerning the device's affordability, technical complexity, and ethical considerations, including privacy safeguards and equitable access. Finally, evaluation limitations, such as small sample sizes and the need for extended longitudinal studies, challenge the comprehensive assessment of the device's efficacy, usability, and user satisfaction. Mitigating these limitations necessitates a holistic approach encompassing meticulous design iterations, extensive user testing and feedback integration, ethical guidelines adherence, and rigorous evaluation methodologies to ensure the device's practicality, effectiveness, and ethical integrity in assisting visually impaired individuals.

IV. PROPOSED WORK

This innovative research combines state-of-the-art technologies in computer vision and deep neural networks to tackle critical challenges. The device's workflow encompasses tasks such as image capture, processing, text extraction, and conversion to speech, all integrated with feedback mechanisms through both speech and haptic interfaces.

Fig. 1. Level 0 Data Flow Diagram

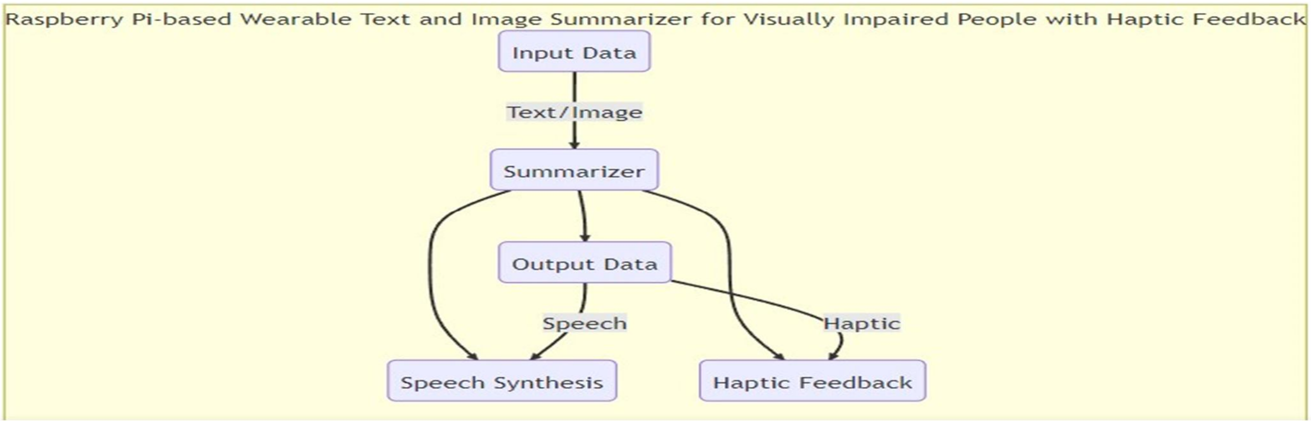


Fig. 2. Level 1 Data Flow Diagram

A. *Text-to-speech Conversion into Haptic Feedback*

The extracted text is further processed by text-to-speech (TTS) software to convert it into speech output that is played through the speaker. Simultaneously, haptic motors provide tactile feedback or alerts to the user.

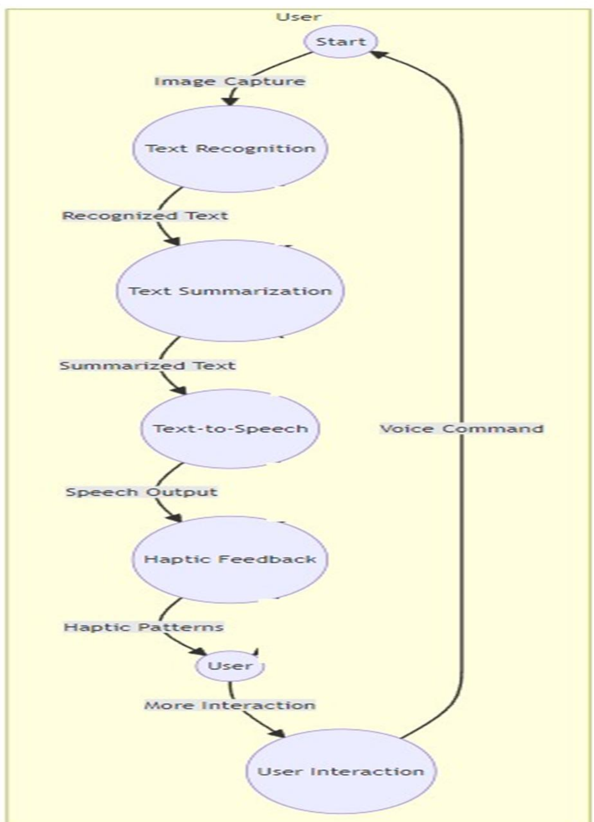


Fig. 3. Flowchart

B. *Algorithms used for Image Captioning:VGG16 (CNN)*

In our project, we use VGG16 as the convolutional neural network (CNN) architecture. The deep structure of VGG16, which consists of 16 layers comprising fully linked and convolutional layers, is well recognized. It employs max pooling layers for downsampling and 3x3 filters with a stride of 1 in convolutional layers. Using its acquired features from extensive datasets, the network makes use of the ReLU activation function and is frequently utilized as a pretrained model for applications such as picture categorization.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

1) *LSTM:* In image captioning jobs, LSTMs are used to produce coherent and descriptive captions for images. In this case, to produce relevant captions, LSTMs process the visual features that are retrieved from the image (usually using a convolutional neural network such as VGG16 or ResNet) and integrate them with contextual data. The LSTM is well-suited for comprehending the sequential nature of language and producing precise captions that match the visual content of the images because of its capacity to grasp long-range relationships and temporal dynamics.

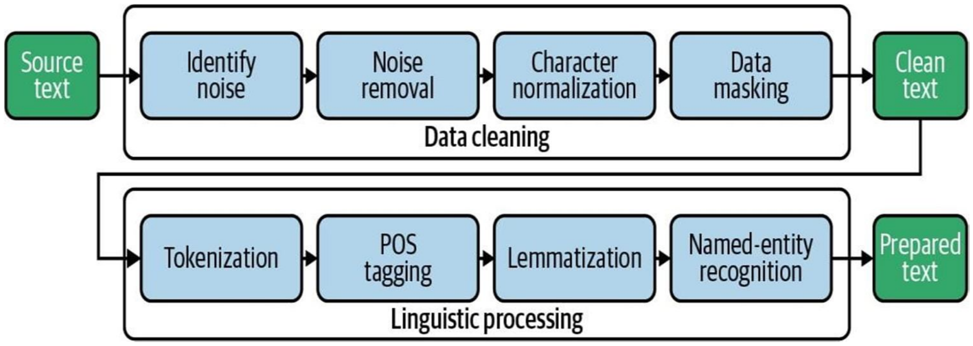


Fig. 4. Steps for Text Preprocessing

C. Text Summarization

Text preprocessing is the process of cleaning and converting unprocessed text data by managing numbers, normalizing words, lowercasing, tokenizing, and deleting punctuation, stop words, and special characters. When it comes to text summary, you can use extractive or abstractive techniques to distill the text and produce succinct summaries based on the important information that has been extracted or, alternatively, by crafting new sentences that encapsulate the main content.

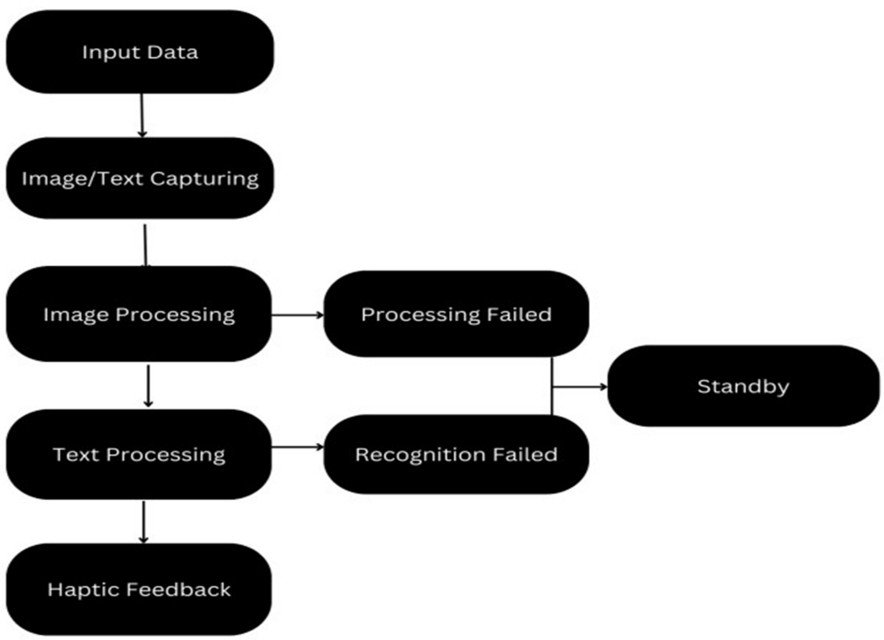


Fig. 5. Activity Diagram

D. Training and Testing Dataset

Constructing this robust system required a crucial element: a diverse and realistic training dataset comprising:

In this project, the dataset has been divided into a 90:10 ratio for training and testing purposes. This means that 90% of the data is used for training the machine learning model, allowing it to learn patterns and relationships from the data. The remaining 10% of the data is reserved for testing the trained model to evaluate its performance and generalization ability on unseen data. This split ensures that the model is trained on a sufficient amount of data while also providing a separate set of data for unbiased evaluation, helping to assess the model's effectiveness and performance metrics accurately.

1) *Dataset Variability:* An additional effort was made to ensure the representation of real-world scenarios within our dataset. This required gathering photographs from the internet, including images from various web sources, and merging random images from the Flickr8k dataset. The variety of the training dataset is proof of the model's flexibility and capacity to effectively handle real-world problems.

E. Scope of the Research

The scope of the results of the project encompasses several key objectives aimed at delivering a comprehensive and impactful solution for visually impaired individuals:

1) *Functional Wearable Device:* The primary goal is to develop and deliver a fully functional wearable reading device. This device will possess the capability to accurately recognize and read both printed text and visual content aloud to users with visual impairments. The device's functionality will be optimized to ensure reliable and efficient performance in real-world scenarios.
2) *Improved Accessibility:* By providing visually impaired individuals with greater access to printed materials and visual information, the project aims to significantly enhance their independence and overall quality of life. The device will bridge the accessibility gap, enabling users to access a wide range of information independently and efficiently.
3) *Enhanced Learning Opportunities:* The device's capabilities extend to facilitating educational activities. By enabling users to access and comprehend a broader range of educational materials and resources, the project seeks to enhance learning opportunities and promote educational inclusivity for visually impaired individuals.
4) *Increased Social Inclusion:* An important outcome of the project is the empowerment of visually impaired individuals to participate more fully in social and professional settings. By equipping users with a tool that enhances their ability to access and interact with information, the device contributes to increased social inclusion and participation.
5) *Potential for Expansion:* Beyond the initial scope, the project creates opportunities for expansion and adaptation. This includes the potential integration of additional features and functionalities to further enhance the device's capabilities. Moreover, the project envisions adaptation for use in diverse cultural and linguistic contexts, making the solution globally accessible and inclusive.

V. MATHEMATICAL MODEL

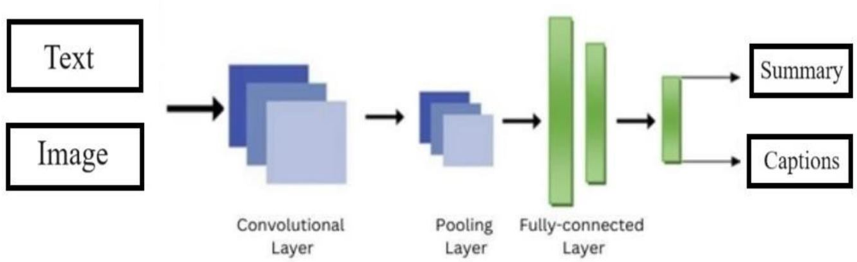


Fig. 7. Mathematical Model 1

1) Convolution entails the multiplication of a 5x5 image matrix by a 3x3 filter matrix, resulting in a "Feature Map" generated through dot products spanning the entire image.
2) Padding, or the addition of zeros, is required when the filter doesn't align perfectly with the input image.
3) Pooling layers serve to downsize large images, decreasing parameter count to facilitate smoother processing.
4) Normalization commonly involves ReLU activation, employing the function $f(x) = \max(0,x)$ to introduce non-linearity and enhance the interpretation of real-world data.

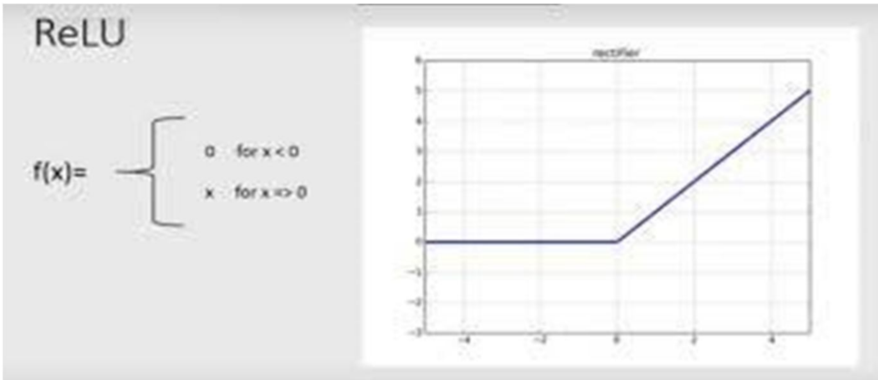


Fig. 8. Mathematical Model 2

Fig. 8, illustrates the mathematical formula for the Rectified Linear Unit (ReLU) function, which is utilized in artificial neural networks. It creates non-linearity by directly outputting the input if it is positive; if not, it outputs zero. This technique is frequently used for feature transformation and activation in a variety of machine learning-models.
ReLU preserves inputs greater than 0 for additional network complexity while setting outputs to 0 for inputs less than 0. • Since negative numbers are assumed to exist in real-world data, you can skip this step if they don't.

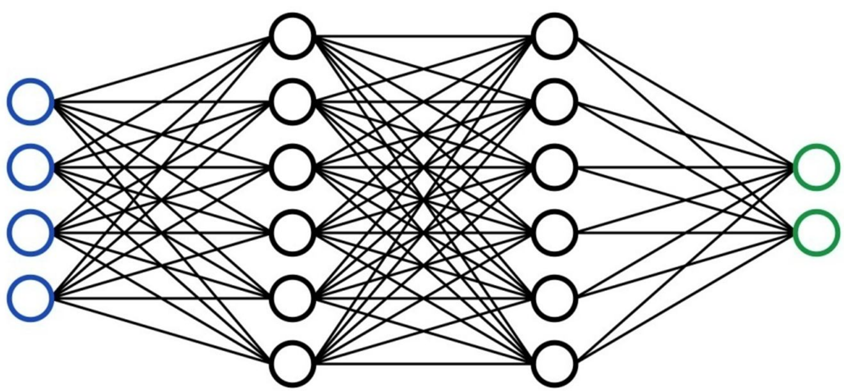


Fig. 9. Mathematical Model 3

VI. RESULTS

The development of a self-contained reading device powered by Raspberry Pi and integrated with advanced deep learning methodologies and voice output capabilities marks a significant advancement in enhancing reading accessibility for visually impaired individuals. By seamlessly merging state-of-the-art technologies such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), ResNet architectures, and stochastic gradient descent optimization, the device transcends the limitations of conventional reading modalities. It not only facilitates access to printed textual material but also extends its capabilities to comprehend and interpret visual content through image analysis and caption generation.
Furthermore, the incorporation of voice output features enhances the usability of the device, providing auditory prompts, notifications, and user interaction feedback. This not only fosters a seamless and intuitive reading experience but also promotes greater independence and efficiency in accessing diverse content for individuals with visual impairments.
The proposed reading device represents a significant step forward in empowering visually impaired individuals to navigate the ever-expanding landscape of information and knowledge. By bridging the gap between textual and visual information, it opens doors to a wealth of educational resources, literature, and multimedia content previously inaccessible to this demographic.
As we look towards the future, continued refinement and optimization of the device, along with the exploration of additional deep learning techniques such as Transfer Learning, GANs (Generative Adversarial Networks), and Attention Mechanisms, hold promise for further enhancing reading accessibility and inclusivity for visually impaired individuals.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

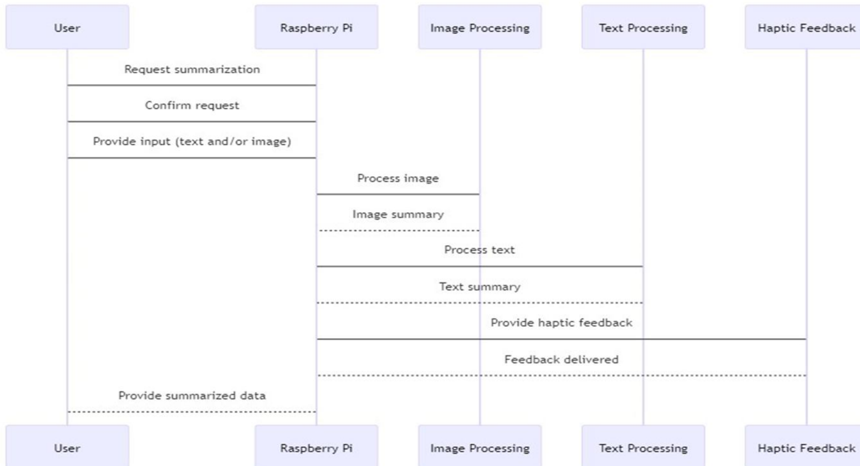


Fig. 10. Sequence Diagram

Our project is centered around creating a self-contained reading device tailored specifically for enhancing accessibility among visually impaired individuals. This innovative device amalgamates a range of hardware components, notably a Raspberry Pi serving as the core processing unit. Complementing the Pi is a camera module adept at capturing images of printed text or visual content, haptic motors designed to provide tactile feedback, a speaker for rendering synthesized speech, and optionally, a microphone enabling voice input or commands. This hardware ensemble forms the foundation of a system designed to empower users with enhanced accessibility to printed materials and visual information. On the software front, your project leverages the versatility of Python programming alongside OpenCV, a powerful tool for image processing tasks such as image capture, preprocessing, and text extraction from images. The pivotal role of optical character recognition (OCR) software, particularly Tesseract OCR, cannot be overstated. Tesseract OCR facilitates the conversion of images containing text into machinereadable text data, which is subsequently fed into sophisticated deep learning techniques like Convolutional Neural Networks (CNNs),Recurrent Neural Networks (RNNs), and possibly Residual Networks (ResNet). These neural networks play a crucial role, especially in tasks such as image captioning, further enriching the device's functionality. The workflow of your device encompasses a series of seamlessly integrated steps, starting with image capture, followed by sophisticated image processing, text extraction, and conversion into speech output via text-to-speech (TTS) software. The synthesized speech is then channeled through the speaker, delivering auditory feedback to the user. Simultaneously, haptic motors come into play, providing tactile feedback or alerts as needed. This comprehensive workflow encapsulates the device's ability to facilitate real-time access to printed text and visual content, thereby augmenting the overall reading experience and information retrieval capabilities for visually impaired individuals.

VII. RESULTS

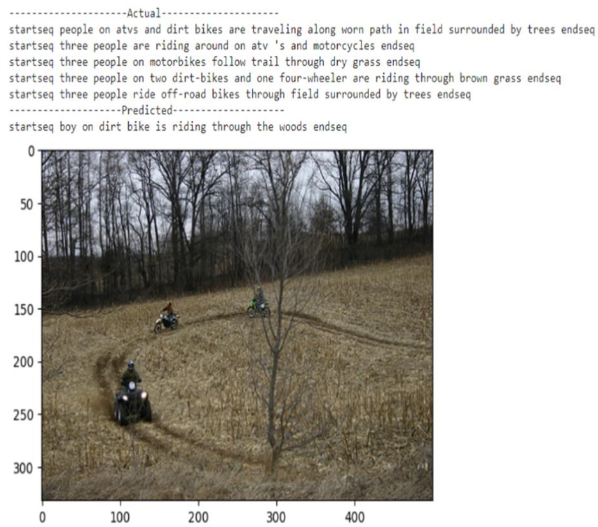


Fig. 11. Expected Outcomes (1 & 2)

VIII. CHALLENGES AND FUTURE SCOPE

The challenges faced by the project include technical complexity in integrating various technologies, ensuring accuracy and reliability of image captioning, designing an intuitive user interface, managing battery life and portability, addressing accessibility needs, balancing cost and affordability, complying with regulations, and gathering user feedback for iterative development. These challenges require a multidisciplinary approach and commitment to user-centered design principles to overcome and create a transformative reading device for visually impaired individuals.

the future scope of the project includes ongoing enhancement of features, miniaturization of hardware, integration with wearable technology, fostering collaborative platforms, personalization options, educational tools, remote assistance capabilities, and ensuring global accessibility. These advancements aim to continuously improve the reading experience and overall quality of life for visually impaired individuals.

IX. CONCLUSION

In conclusion, the development of a wearable reader utilizing a high-resolution miniature camera, Raspberry Pi microcontroller, and vibration motors represents a pioneering solution to address the limitations faced by visually impaired individuals in accessing printed text. By enabling real-time image capture and processing, the device converts printed text into computerized format, enhancing accessibility and expanding learning resources beyond those available in Braille. The incorporation of tactile feedback through vibration motors aids users in maintaining reading orientation, while the conversion of text to audible speech offers a hands-free and versatile reading experience. This innovative device holds great potential to empower visually impaired users with increased independence and access to a diverse range of educational materials, thereby fostering greater inclusivity and opportunities for learning.

REFERENCES

[1] J. N. Madhuri and R. Ganesh Kumar, ”Extractive Text Summarization Using Sentence Ranking,” 2019 International Conference on Data Science and Communication (Icon DSC), Bangalore, India, 2019, pp. 1-3, doi: 10.1109/IconDSC.2019.8817040.

[2] Rahul, S. Adhikari and Monika, ”NLP based Machine Learning Approaches for Text Summarization,” 2020 Fourth International Con- ference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC- 00099.

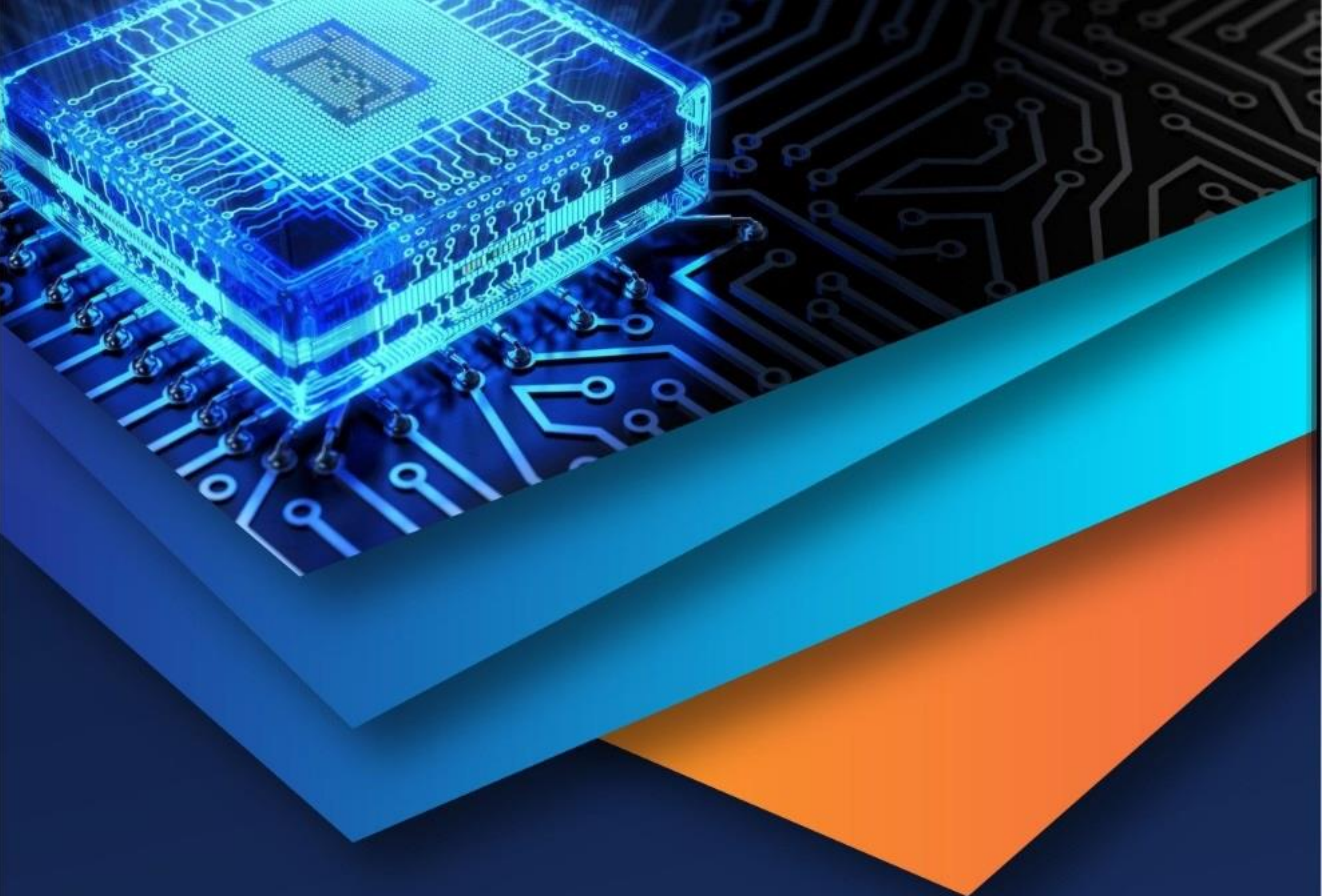
[3] Rahul, S. Rauniyar and Monika, ”A Survey on Deep Learning based Various Methods Analysis of Text Summarization,” 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbat- ore, India, 2020, pp. 113-116, doi: 10.1109/ICICT48043.2020.9112474.

[4] P. Janjanam and C. P. Reddy, ”Text Summarization: An Essential Study,” 2019 International Conference on Computational Intelli- gence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862030

[5] G. V. Madhuri Chandu, A. Premkumar, S. S. K and N. Sam- path, ”Extractive Approach For Query Based Text Summarization,” 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2019, pp. 1-5, doi: 10.1109/ICICT46931.2019.8977708.

[6] C. Xu, ”Research on Information Retrieval Algorithm Based on TextRank,” 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 2019, pp. 180-183, doi: 10.1109/YAC.2019.8787615

[7] A.Ramesh, K. G. Srinivasa and N. Pramod, ”SentenceRank — A graph based approach to summarize text,” The Fifth International Con- ference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), 2014, pp. 177-182, doi: 10.1109/ICADIWT.2014
[8] D. J. Calder, “Assistive technologies and the visually impaired: A digital ecosystem perspective,” in Proceedings of the 3rdInternational Conference on PErvasive Technologies Related to Assistive Environments, ser. PETRA’10. New York, NY, USA: ACM, 2010, pp. 1- 8
[9] W. Chan, N. Jaitly, Q. V. Le, and O. Vinyals, “Listen, attend and spell,” CoRR, vol. abs/1508.01211, 2015
[10] R. Prabhavalkar, K. Rao, T. N. Sainath, B. Li, L. Johnson, and N. Jaitly, “A Comparison of Sequence-to- sequence Models for Speech Recognition,” in Proc. Interspeech, 2017.
[11] Yun Ren, Changren Zhu, Shunping Xiao, “Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures”, in Hindawi. Mathematical Problems in Engineering, Vol. 2018



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY