African Journal of Education Science and Technology (AJEST) Vol. 8 No.1 (October, 2024)
University of Eldoret, Kenya, Mount Kenya University, Kenya, Chukwuemeka Odemegwu Ojukwu University, Nigeria, Kyambogo University, Uganda and University of Makeni, Sierra Leone

Development of Artificial Intelligence (AI)-Driven Aid to Enhance Visual and Hearing-Impaired Students

*Phocas Sebastian, Joseph Sospeter Salawa, Charles Okanda Nyatega, Juma Said Ally, Cuthbert John Karawa, Elizabeth Odrick Koola and Richard Mwanjalila
Department of Electronics and Telecommunication Engineering, School of Engineering, Mbeya University of Science and Technology(MUST), P.O. BOX 131 Mbeya

***Corresponding Email :** phocassebastian@gmail.com

Abstract

This paper introduces a series of works on Artificial Intelligence (AI)-based assistive devices that improve students' learning with visual and hearing impairments. Artificial intelligence technology provides personal help and support for various learning tasks and activities. The system uses computer vision techniques to read visual information such as text and images, which can be made available in usable formats such as audio descriptions. In addition, the system can recognize and respond to the user's voice. commands and requests using speech recognition technologies. Stu- dents can view learning materials, get instant help with classroom activities, and participate in engaging learning exercises designed for their specific needs through an intuitive user interface. Ensuring equal access to educational resources. The project focuses on the effective and efficient teaching and mentoring of students with visual and hearing impairments. To understand the impact and applicability of the proposed AI-based tool in enhancing students with vision or hearing impairments and overall educational engagement, user surveys, and feedback are taken, and it is clear that, to a greater extent, it shows the potential and utility of the system to be included in real-world classroom settings. The importance of this paper is particularly based on the contention that it will bring about a major change as far as education is concerned in a bid to make these noble provisions available for students with disabilities. It speeds up access to the required information, fosters differentiation in delivering the instructions, offers quick help, helps improve academic achievement significantly, and helps learners develop confidence in themselves. Further refinements and extensive user evaluations are ongoing to ensure the system meets the diverse needs of students with visual and hearing impairments.

Keywords: *Artificial Intelligence, Optical Character Recognition, Speech-To-Text, Text-To-Speech*

INTRODUCTION

Eyes and ears are the most essential part of the human body, our eyes relay the location and movement of objects in our visual field, while our ears convey information about the actions of objects both in and outside of our visual field. Though our ears are intended to detect sound and our eyes are meant to see light, there are some remarkable similarities in how the two organs process the physical phenomena occurring in and around us. When sound and light pass through their respective organs, they undergo several changes before being absorbed by the body's neurological pathways. Since the ears and eyes are two subsystems that cross-reference information for certain crucial functions, using information about sound and light is not a

mutually incompatible function. The visual system receives signals from the balance receptors in the inner ear to maintain object focus when we move. Similar to this, abrupt, loud noises cause the eyelids to blink to shield them from the impact. According to the World Health Organization, at least 220 million people world- wide are visually impaired. Blind people have a disease or accidental injury to the optic nerve or eye that has left them blind in both or one eye. In most people's perceptions, people with visual impairments are all blind. Blindness is a type of visual impairment (World Report on Disability, n.d.). Today, we embark on a journey to explore a groundbreaking project at the intersection of technology and inclusivity. Our focus is on leveraging artificial intelligence to create a transformative aid for individuals with visual and hearing impairments in Tanzanian local schools. In our local schools, the pursuit of knowledge is a shared ambition, however, the road to education re- mains unequal. This article proposes support innovations to assist those with visual and hearing impairments (Nemade, Patil, Bijwe, Bhangale, & Mapari, 2022). The framework is based on the Raspberry Pi 3 B+ model and continuously coordinates many important subsystems, including the camera module, Tesseract OCR (optical human recognition) engine, and Python-based TTS (text- to-speech) synthesizer. The camera section is an important information widget that can be used to take photos with messages. These images are then processed using the Raspberry Pi's advanced image processing calculations. The processed image is then transferred to Tesseract's OCR engine, which performs character recognition and converts the image's text into parallel encrypted text. Semantic checks are conducted to ensure accuracy and meaning when extracting printed con- tent. The analyzed text is then transferred to the TTS module, which converts the text into audible conversation. With this comprehensive approach, Raspberry Pi can recognize real images of text and produce speech relevant to the important goals of students with visual and hearing impair- ments. The significance of this paper rests in addressing a fundamental gap: the accessibility and inclusion of education for students with visual and hearing impairments.

Recent Methods of AI

With huge technological advances in computer vision and deep learning, artificial intelligence is being used in medical imaging (Villata et al., 2022) (Feuerriegel, Shrestha, Von Krogh, & Zhang,
2022). The addition of artificial intelligence methods to the treatment of visual impairments has made doctors more accurate and faster in clinical examination, diagnosis, and prognosis. In diagnosing eyes, fundus images are easily accessible and contain rich biological information, which is suitable as input datasets for CNN, GAN, and transfer learning. All three methods are suitable for processing images for analysis. CNNs can be directly convoluted with image pixels to extract im- age features from them. In addition to the classification and feature extraction of traditional neural networks, GANs can generate new images based on the characteristics of real data. In practice, this does not require additional mathematical hyperparameters. Transfer learning lowers the requirements for an independent and identical distribution of training and test data. It also eliminates the need to manually enter data records during training. Transfer learning can effectively solve the problems of poor generalization results, time-consuming and labor-intensive processes, and small data sets.

Convolutional Neural Network (CNN)

One important branch of deep learning X. Wang et al. (2021), CNN, is often used for image and video processing. CNNs differ from normal neural net- works in that they contain the feature learning stage consisting of one or more convolutional and pooling layers J. Wang,

Wang, and Zhang (2023). A convolutional neural network (CNN) is a regularized type of feed-forward neural network that learns feature engineering by itself via filter (or kernel) optimization. Vanishing gradients and exploding gradients, seen during backpropagation in earlier neural networks, are prevented by using regularized weights over fewer connections (Venkatesan & Li, 2017). CNNs are also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps (Convolutional neural network, 2024). Counter-intuitively, most convolutional neural networks are not invariant to translation due to the down-sampling operation they apply to the input (Convolutional neural network, 2024). Figure 1 below shows the convolutional neural network.

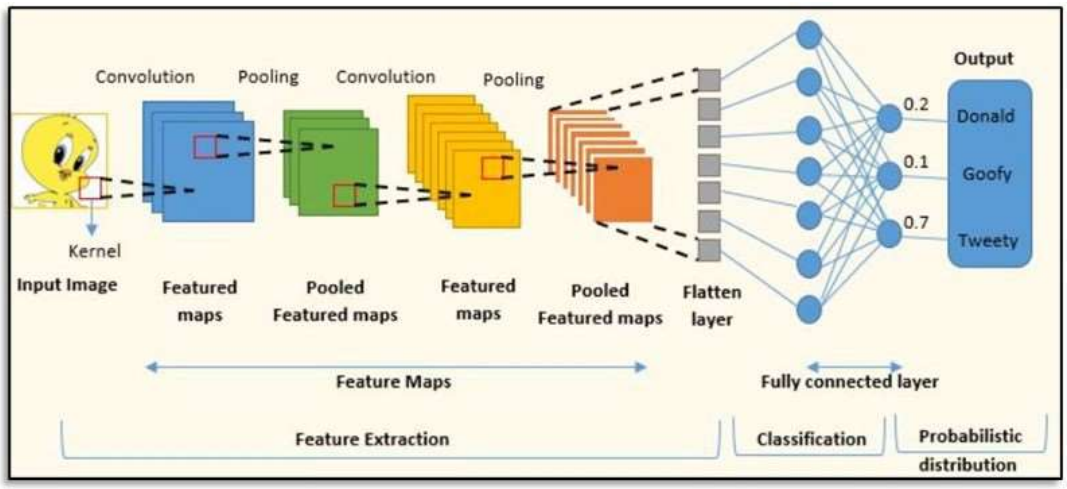

Figure 1: Convolution Neural Network

Generative Adversarial Network (GAN)

Generative adversarial networks, or GANs for short, are an approach to generative modeling that uses deep learning techniques such as convolutional neural networks. Generative modeling is an unsupervised machine-learning task that involves the automatic discovery and learning of regularities or patterns. input data in such a way that the model can be used to generate or output new examples that could probably be from the original dataset (Brownlee, 2019). Generative adversarial networks (GANs) (Iqbal, Sharif, Yasmin, Raza,& Aftab, 2022) can be divided into two main parts: generator and discriminator, whose roles and functions have been explained as follows:

Discriminator

The discriminator in a GAN is simply a classifier. Figure 2 below tries to distinguish real data from the data created by the generator. It could use any network architecture appropriate to the type of data it's classifying (The Discriminator | Machine Learning, n.d.).

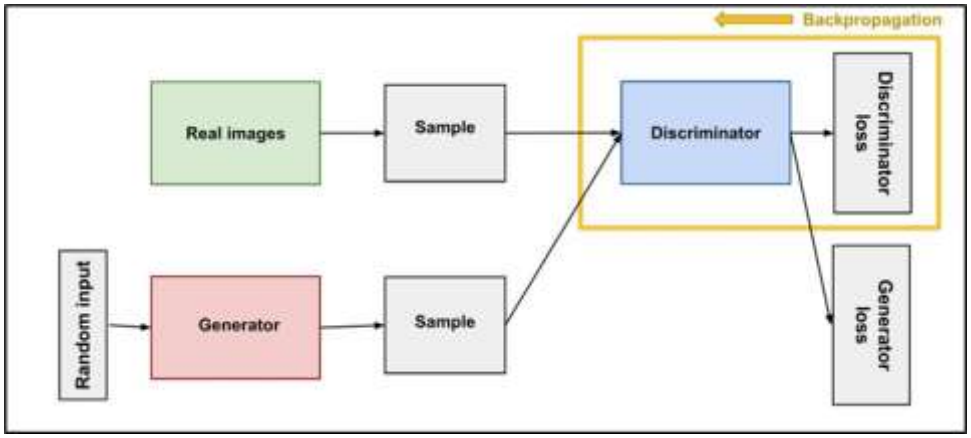

Figure 2: Backpropagation in the Discriminator Training (The Discriminator | Machine Learning, n.d.)

ii) Generator

The generator part of a GAN learns to create fake data by incorporating feedback from the discriminator. It learns to make the discriminator classify its output as real (Ramyasri et al., 2023). Generator training requires tighter integration between the generator and the discriminator than discriminator training does. The portion of the GAN that trains the generator includes:

- random input
- generator network, which transforms the random input into a data instance
- discriminator network, which classifies the generated data
- discriminator output
- generator loss, which penalizes the generator for failing to fool the discriminator

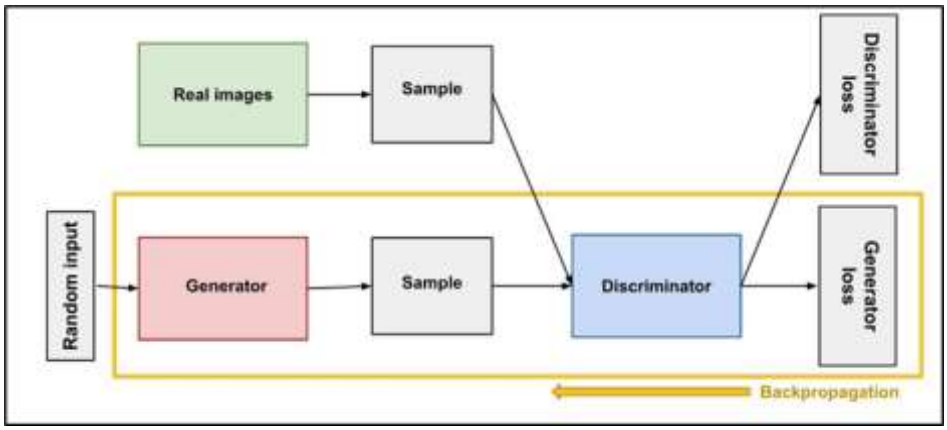

Figure 3: Backpropagation in the Generator Training (The Generator | Machine Learning | Google for Developers, n.d.)

The above figure 3 shows the backpropagation in Generator training as the article (J. Wang et al.,2023) suggested that the training of a GAN is the process by which the generator and discriminator learn from each other. The generator network is expected to produce as much realistic data as possible. With a discriminator, the realism of the data can be evaluated quantitatively. The generator generates some images based on a bunch of random vectors. The generator is trained with feedback from the discriminator. It receives a positive reward when it successfully fools the discriminator and a negative reward when it fails. The generator calculates the loss of the discriminator classification and uses backpropagation to obtain the gradient. Performing backpropagation will fine-tune the parameters of the generator and tend to stabilize it. We can define the number of iterations, and the training can be stopped after a certain number of alternating iterations. At the end of the training, the GAN is given a sufficiently messy generator and a discriminator that distinguishes between true and false images. Therefore, it is also commonly used to improve the quality of medical images.

Transfer Learning

Transfer learning is a technique that utilizes a trained model's knowledge to learn another set of data. Transfer learning aims to improve learning in the target domain by lever- aging knowledge from the source domain and learning tasks. Different transfer learning settings are defined based on the type of task and the nature of the data available in the source and target domains [10].

Comparison of AI-based approaches to Assist Reading

Comparing AI-based approaches to assisted reading involves evaluating various techniques and technologies aimed at improving access to reading for people with visual impairments. These approaches leverage artificial intelligence (AI) to provide individual support and assistance in accessing written content. Some common AI- based approaches include text-to-speech Optical Character Recognition (OCR) and Natural Language Processing (NLP). Table 1 shows the comparison between different AI approaches to assisted reading.

Table 1: Comparison of AI-Based Approaches

Ref	Year	Hardware	Features	Disadvantages
(Rattanaphinyowanich &	2021	Raspberry Pi, camera, sensor and phones	Generating results from Keeping a safe tance from objects.	No mention of text being gual
(Punith et al., 2021)	2021	Raspberry Pi, camera, sensor and phones	Generating results from Keeping a safe tance from objects.	No mention of text being gual
(Yeo, Bae, Lee, Kim, &	2022	The VR mounted, Smartphone, plication	Customizable. Clinically to be more tive.	The reading has not changed

		and Remote		
(Mukhiddinov & Cho,	2021	Smart glass, phone/ tactile pad	More versatile suitable for a range of ments	Some errors in recognition nent object tion models contain some during extraction
(Sarkar et al., 2021)	2021	The Raspberry Pi board	Fixed recognition tiple fonts nized	Higher prices
(Nemade et al., 2022)	2022	Raspberry Pi Web camera phone/	For web tion	Not applicable paper-based texts
(Pandya & Goradiya, 2021)	2021	Raspberry Pi Cam-	LED light and sensor era, Power booster and	

Power supply

Multi-purpose Fast, efficient, and robust

Only available in selected areas

MATERIALS AND METHODS

As referred to in figure 4 works on button interaction and provides a simplified and easy-to-use user interface for the visually and hearing impaired. This means that the product remains idle for the user to switch the button on once after its reset is performed. As much as he/she gives a command, something happens in the system. After activation, there is a confirmation that the command has been received and the prompt asking for the command to be typed is sent. By doing this step, you can be assured that the user knows that commands will be described to the system. If for instance, the user wishes to ask the system to perform a specific operation or to start listening, the user can press the button again.

Regime: When a user requests the system to perform an activity, the system initiates the corresponding action for the task like using the input device or intrinsic camera to capture a photograph. Finally, the system exploits OCR to read the obtained image recognize text from it, and store it in its database. The technology reads the decoded text, using voice output, and provides graphic information in audio form. In this mode, the system checks continuously for new commands given by the user. If a user chooses to stop listening mode, the system exits listening mode and waits for further commands. However, if the user confirms that he intends to stop using the system (e.g., ”yes”) it will stop working gracefully and wait for further activation. During communication, the system provides clear feedback to ensure a smooth user

experience. And quick answers to user commands. The system uses a push-button interface, providing a simple and intuitive control mechanism that allows visually and hearing-impaired people to effectively use and interact with the system.

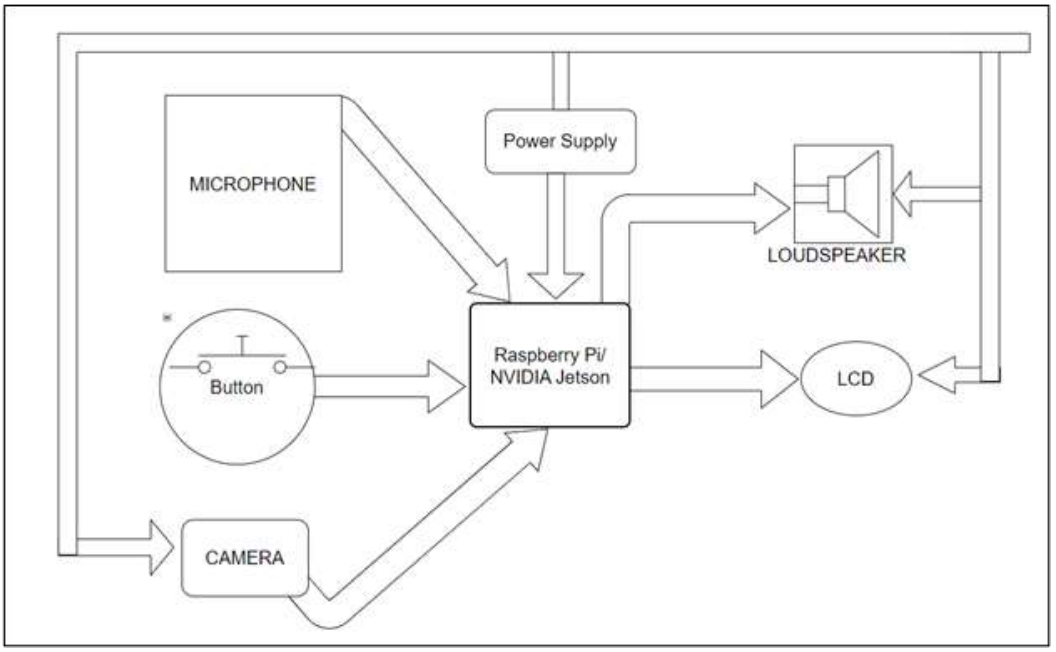

Figure 4: Block Diagram of The system

RESULTS AND DISCUSSION

Figure 5 shows the flow chart of the system operation, whereas it describes the sequential operation of a system designed to assist visually and hearing-impaired students. After reset, the system waits for the button input to start. After activation, the system gives the user a voice and text response on the screen, indicating readiness to receive. The system then listens for user commands with two main branches: ”hello” and ”listen”. When the user gives the command ”hello”, the system takes a photo with the webcam, performs optical character recognition (OCR) to extract the text on the captured image, converts the detected text into speech, and completes the process. Alternatively, when the user issues a”listen” command, the system acknowledges the command and waits for further input. If the user asks to stop listening, the system will stop listening. Otherwise, it will continue to listen for commands. The following are the steps followed during this simulation procedure:

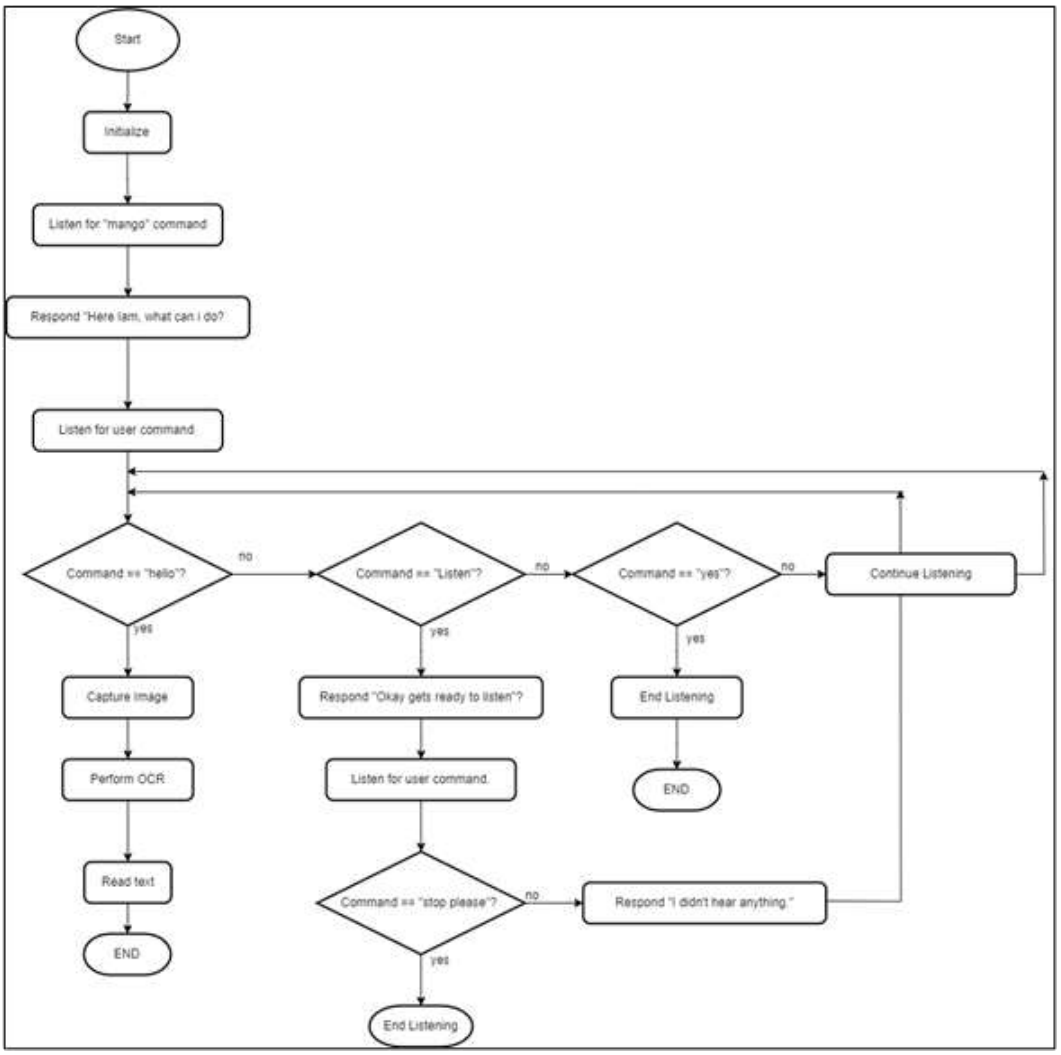

Figure 5: Flow Chart on the System Operation

Image Capture and Processing

The simulation results demonstrate the successful capture of images from the webcam using the OpenCV library. The captured images are processed to extract text using specialized image processing algorithms. This step ensures that the text is accurately extracted from the captured images, laying the foundation for subsequent OCR processing. The figure 6 below shows the captured image

Figure 6: Image Capturing and Processing

Optical Character Recognition (OCR)

The simulation results indicate the effectiveness of the Tesseract OCR engine in accurately recognizing text from the captured images. The OCR engine converts the extracted text from the images into binary-coded text, which serves as input for semantic analysis. The high accuracy of the OCR process ensures reliable conversion of visual information into machine-readable text, enabling further analysis and processing. The below figure 7 shows how the Tesseract engine reads the text from the captured image.

Figure 7: Text Extracted from the OCR performed

Text-to-Speech (TTS) Conversion

Following successful OCR processing, the simulation results demonstrate the seamless conversion of the recognized text into audible speech using the Pyttsx3 library. This conversion enables the synthesized speech to be heard by users, providing accessibility to individuals with visual impairments. The clarity and naturalness of the synthesized speech reflect the effectiveness of the TTS module in conveying textual information audibly.

User Interaction and Control

The simulation results also highlight the user-friendly interface and interactive nature of the system. Users can initiate various functionalities, such as image capture, OCR processing, and TTS conversion, through voice commands or push button input. The system's responsiveness to user commands enhances its usability and accessibility, catering to individuals with diverse needs and preferences.

CONCLUSION

Based on the simulation results, the system is robust and reliable, suggesting that it can effectively support both visually and hearing-impaired students. Integrating various subsystems such as image processing, optical character recognition (OCR), and text-to-speech (TTS), the system seamlessly converts visual information into accessible formats and also enables audio input for hearing-impaired students. The accurate text capture, processing, and transformation shown in the results highlight the system's ability to facilitate the use and interaction of text content with students with visual and hearing impairments, improving their learning.

ACKNOWLEDGMENTS

Sincere gratitude is accorded to everyone who contributed to this project.

REFERENCES

Brownlee, J. (2019, June). A Gentle Introduction to Generative Adversarial Networks (GANs). Retrieved 2024-02-29, from https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/

Feuerriegel, S., Shrestha, Y. R., Von Krogh, G., & Zhang, C. (2022, July). Bringing artificial intelligence to business management. Nature Machine Intelligence, 4(7), 611–613. Retrieved 2024-02-29, from https://www.nature.com/articles/s42256-022-00512-5 doi: 10.1038/s42256-022-00512-5

Iqbal, A., Sharif, M., Yasmin, M., Raza, M., & Aftab, S. (2022, September). Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. International Journal of Multimedia Information Retrieval, 11(3), 333–368. Retrieved 2024-02-29, from https://link.springer.com/10.1007/s13735-022-00240-x doi: 10.1007/s13735-022-00240-x

for Blind. International Journal of Innovative and Emerging Research in Engineering, 5(1). Retrieved 2024-02-26, from http://ijiere.com/FinalPaper/finalPaperRaspberry%20Pi%20based%20reader%20for%20Blind191540.pdfdoi:10.26769/IJIERE.2018.5.1.191540

Jadhav, M. S., Koli, S. M., & Kulkarni, S. S. (2018, March). Raspberry Pi Based Reader

Mukhiddinov, M., & Cho, J. (2021, November). Smart Glass System Using Deep Learing-

& Wyner, A. (2022, December). Thirty years of artificial intelligence and law: the third decade. Artificial Intelligence and Law, 30(4), 561–591. Retrieved 2024-02-29, from https://link.springer.com/10.1007/s10506-022-09327-6 doi:10.1007/s10506-022-09327-6

Nemade, S., Patil, R., Bijwe, I., Bhangale, K., & Mapari, R. (2022, May). Artificial Vision- Raspberry Pi Based Reader for Visually Impaired. Artificial Vision-Raspberry Pi Based Reader for Visually Impaired, 1(1), 1–3.

Pandya, K., & Goradiya, D. B. C. (2021). Audio Assisted Electronic Glasses For Blind & Visually Impaired People Using Deep Learning. , 16.

Punith, A., Manish, G., Sumanth, M. S., Vinay, A., Karthik, R., & Jyothi, K. (2021). Design and implementation of a smart reader for blind and visually impaired people. In (p. 060002). Kuching, Malaysia. Retrieved 2024-02-25, from https://pubs.aip.org/aip/ acp/article/1002159 doi: 10.1063/5.0036140

Ramyasri, M., Sridevi, G., Lavanya, K., & Sindhuja, S. (2023, June). Raspberry PiBased Reader for Blind. Journal

on Electronic and Automation Engineering, 2(2), 08–WIBORD as computer assisted learning facilities for children with visual impair- ment. Journal of Physics: Conference Series, 1835(1), 012080. Retrieved 2024-02-29, from https://iopscience.iop.org/article/10.1088/1742-6596/1835/1/012080 doi: 10.1088/1742-6596/1835/1/012080

Rattanaphinyowanich, T., & Nunta, S. (2021, March). Development of DAISY-

Sarkar, S., Pansare, G., Patel, B., Gupta, A., Chauhan, A., Yadav, R., & Attula, N. B(2021, April). Smart Reader for Visually Impaired Using Raspberry Pi. IOP Conference Series: Materials Science and Engineering, 1132(1), 012032. Re- trieved 2024-02-29, from https://iopscience.iop.org/article/10.1088/

The Generator | Machine Learning | Google for Developers. (n.d.). Retrieved 2024-03-24, fromhttps://developers.google.com/machine-learning/gan/generator

Venkatesan, R., & Li, B. (2017). Convolutional Neural Networks in Visual Computing: A Concise Guide (1st ed.). Boca Raton ; London : Taylor & Francis, CRC Press, 2017.: CRC Press. Retrieved 2024-02-29, from https://www.taylorfrancis.com/books/9781498770408 doi: 10.4324/9781315154282

Villata, S., Araszkiewicz, M., Ashley, K., Bench-Capon, T., Branting, L. K., Conrad, J. G.

Wang, X., Wang, C., Liu, B., Zhou, X., Zhang, L., Zheng, J., & Bai, X. (2021, December). Multi-view stereo in

the Deep Learning Era: A comprehensive review. Displays, 70, 102102. Retrieved 2024-02-29, from https://linkinghub.elsevier.com/retrieve/ pii/S0141938221001062 doi: 10.1016/j.displa.2021.102102

Yeo, J. H., Bae, S. H., Lee, S. H., Kim, K. W., & Moon, N. J. (2022, June). Clinical performance1757-899X/1132/1/012032 doi: 10.1088/1757-899X/1132/1/012032uploads/2023/05/10.46632-jeae-2-2-3.pdf doi: 10.46632/jeae/2/2/3

Wang, J., Wang, S., & Zhang, Y. (2023, April). Artificial intelligence for visually impaired.Displays, 77, 102391. Retrieved 2024-02-28, from https://linkinghub.elsevier.com/retrieve/pii/S0141938223000240 doi: 10.1016/j.displa.2023.102391

World Report on Disability. (n.d.). Retrieved 2024-03-22, from https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/world-report-on-disabilityof a smartphone-based low vision aid. Scientific Reports, 12(1), 10752. Retrieved 2024-02-29, from https://www.nature.com/articles/s41598-022-14489-z doi:10.1038/s41598-022-14489-z