"PRO-EYE" - A VIRTUAL ASSISTANT TO FACILITATE VISUALLY IMPAIRED INDIVIDUAL

Yogalingam Senthuran, Pushpakumare T.P.I.S, Amaranayake W.P.K, De Silva H.L.T., Dr. Anuradha Jayakody, Shashika Lokuliyana

Department of Computer Systems Engineering, Sri Lanka Institute of Information Technology, Sri Lanka

ABSTRACT

Human senses play a crucial role in everyday life. Senses proactively let the brain knows about the environment, protects by alerting of any dangers. Out of the five senses, humans have 'sight' is the most dominant sense. More than 80% of what we perceive comes through our eyesight. In a fast-paced world to acquire knowledge and to survive independency has become a significant factor, but for a visually impaired person this independency is a critical question. As per studies visual loss a double the risks of depression as well as they can also develop many other mental disorders. Each day to do even routine tasks a visually impaired person needs the assistance of another person which makes the impaired person a dependent personality. Decision-making is a key factor that keeps a human's life intact. Even though some technologies and devices are advanced enough to help vision-impaired they are either not affordable by most or complex in interaction. The proposed system compromises state-of-the-art technologies at low cost to aid visually impaired personalities in outdoor navigation, image recognition, text recognition using human alike conversations, providing effortless interactions. This improves the visual perception and the awareness of an ambiguous environment. This system analyzes the data and categorizes it into sections using Machine learning and Artificial Intelligence technologies to provide clear guidance to the user. The Proposed System includes a text recognition system which classifies text, tables, and graph in the material separately, and convert them to "life like speech" format, which can help visually impaired persons to read printed materials that are not available in braille format. Furthermore, the proposed solution provides outdoor navigation functionality for the visually impaired which integrates with the global navigation system to guide the user to a particular designation and avoiding static obstacles in the path. The proposed solution computes the optimized routes based on user preference, temporal constraints such as traffic congestion, road closings using real-time processing. The system constantly guides the blind user to navigate based on static and dynamic data. Users can interact using voice and the Assistant will be able to understand it using natural language processing.

Keywords: Virtual Assistant, Object Recognition, Intelligent Character Recognition, Natural Language Processing, Inference Engine, Expert System.

INTRODUCTION

Blindness, the most extreme type of visual disability, can diminish individuals' capacity to perform day-by-day tasks, and move about independently. Great quality restoration permits individuals with various degrees of visual impairment to completely benefit from life, accomplish their objectives to be active and productive in the present society.

An Artificially Intelligent assistant will be the core of the Pro-Eye which bridges the gap between the user and the other provided services such as navigation, environment identification (Image identification, Object tracking), and text recognition (Symbols, Charts). The assistant will proactively interact with the user and will guide him as well as help him by understanding his/her requirements. Users can interact using voice and the Assistant will be able to understand it using natural language processing. Pro-Eye acts like a human-like guider when the user steps to the outside environment. Navigational Functionality is a combination of two sub functionalities, which guide the user globally with geographical routes and guide the user to walk in the path considering physical temporal data. The proposed system is also capable of tracking obstacles in the walking lane and providing a user with navigational rules to avoid obstacles by understanding user requirements.

LITERATURE REVIEW

A key problem in the current developed assistive technologies in the market does not provide centralized control to perceive various functionalities of the system and most developed technologies provide only limited capabilities such as text reading. Implications of a survey done under “Needs of Visually Impaired Users and Requirements for a Virtual Assistant in Ambient Assisted Living” [1] states that for a visual impaired to study he would at least require Audiobooks or Braille Books, talking newspapers, and other assistive technologies to facilitate their education, this also implies the impaired person will also require helpers to provide help to utilize these tools available to them which are not efficient.

An electronic travel aid was introduced to provide enhanced navigational functionalities over the traditional white cane. This electronic travel aid is based on a set of sensors to provide the capabilities. An ultrasonic sensor for discovering obstacles, limiting detection range is achieved by coupling a gyro sensor and inclinometer, and to identify the color of the obstacles a color is used. When an obstacle is detected a vibrator is used to inform the user [2]. “The Guide Cane – Applying Mobile Robot Technologies to Assist the Visually Impaired” [3] elaborates about a device designed to help the visually impaired navigate locally using ultrasonic sensors. Even though the local navigation is provided by the proposed solution the user is unable to get a brief idea about the surroundings. The paper also suggests that there is a requirement to use computer vision to assist the user in ways that a traditional sensor-based navigation assistive technology unable to cater. Visual Pal a mobile device that provides object recognition functionalities via a hybrid algorithm combining both Artificial neural networks and Euclidean distance measures. The hybrid use of algorithms provides more accuracy and the colors are classified according to brightness and robustness [4]. This paper suggests providing a touch-free interaction to users for future improvements.

METHODOLOGY

Smartphone adaption now is immense, this also can be seen common among the visually impaired as well. The major reasons for the adaption are accessibility and mobility.

In Pro-Eye, we propose a system that provides key assistive functionalities to the visual impaired to carry out their day-to-day activities. An application specifically designed as an assistive technology running based on a smartphone regardless of the hardware can be very beneficial and reachable to the visual impaired.

Images are automatically taken by the application itself periodically to identify the environment and pro-actively assist the user. Taken images are pre-processed

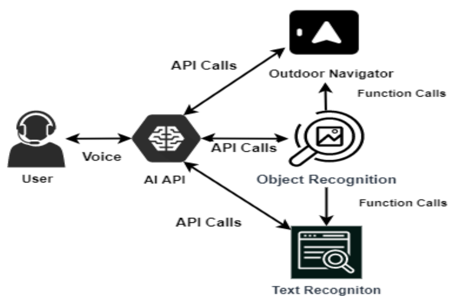


Figure 1: Logical View of the System

I.User interaction using NLP

The approach of handling the user interaction is one of the core concerns since when a vision-impaired user progressively uses any assistive technology its preferable to have a human-like companion which can bring randomness into responses as well as understanding random user queries rather than having a set of rules which predicates the user to remember the questions or patterns that can be requested from the assistant.

The system uses no SQL database to store user details and uses Redis cache to improve performance. When the mobile application is invoked by the user a docker container will be triggered by a lambda function that will serve the requests. The goal is to identify the question or the query of the user into subcategories on which the

to make sure its in minimal resolution and processed in decouple manner to recognize objects and text parallel to increase efficiency and response time. The visual impaired can directly interact with the application via voice and get the required assistance which is outdoor navigation, object recognition, and text reading the application assists with voice communication as well. As Fig. 1 illustrates the system compromises decoupled modules provide the functionalities without stressing the whole system.

assistant can provide services. Spacy and open-source library for advanced natural language processing is optimized to classify the queries as specified above [5]. Using convolutional neural networks on named entity recognition to identify and classify the class of the query.

Pro-eye is trained on sentence based NER and entity label representation as follows,

Table 1: NER Entity Labels of Pro-Eye

Entity Label	Query Objective
NVG	Assist on navigation
TXT	Assist on Reading
OBJ	Assistant on recognition

In addition to the above mentions, entity labels pro-eye also include other labels which are based on the model English multi-task CNN trained on Onto Notes [6] the pre-trained model that is trained to understand the above entities.

Pro-eye uses the efficient implementation of word embedding strategy using subword features and Bloom embeddings, where the named entity parsing is handled through a deep convolutional neural network with persisting interrelation. Being trained on a new entity label the model will initially predict the unlabeled text and since we know the correct answer the feedback of the prediction made by the model can be given in the form of error gradient of the loss function which is a contrasting prediction expected value and output. The NER is trained with a hinge-loss objective which is used for the identification of loss function [7].

The application uses the trained model in the live scenario when a query comes in to classify, done by predicting the text by converting it into a spacy type of document. These entities are objectified as ents in the document type of the space. Once the query is classified it uses the parts of speech tagging to identify the object the user requires from the assistant to achieve for example if the user has questioned the assistant as “How can I go to library?” the NER model identification will identify as “NVG” which represents navigation which is the objective hence the object will be library to identify the object POS tagging [8] is used. Tokenization the process of chopping a unit of the document into pieces which are called tokens. The tokens are a set of series of characters that can be used in semantic processing. Once the tokenization of the document is completed pro-eye uses spacy to parse and tag a provided document in this case a query. The parsed document of the “How can I go to library?” query is shown in Fig. 2.

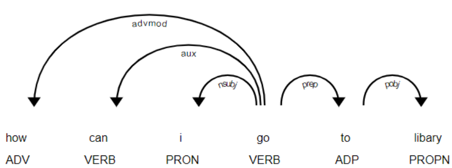


Figure 2: Semantic Relations in a sentence

Using the Parts of speech tagging the actual meaning of the text can be understood and the object the user trying to achieve can marked as library hence the final output of natural language processing component of the pro-eye will be both action and the object in the format of action as NVG and object as library.

After the action and object identification is completed pro-eye will use the artificial linguistic internet computer entity to generate random responses for the user while triggering the assistive service, in this case the navigation service.

II. Object Detection and Color Detection

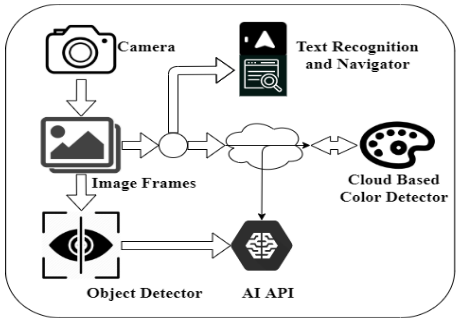


Figure 3: System Diagram of Camera and Object detection

As illustrated in Fig. 3. this function has several steps. AI API interacts with the user and provides the relevant instruction to the other components. To access the camera and take the image frames, CameraX API has been used [9]. Since the users are visually impaired It does not have a preview layout. All the functions have been automated due to these special users. When image frames are saved in the storage, those are converted into bitmap objects to work on object detection. Object detection has been built on the device. TensorFlow Lite interpreter has been applied to work on object detectors because it helps developers run TensorFlow models on mobile, embedded, and IoT devices and It enables on-device machine learning inference with low latency and small binary size [10]. A model for object detection is trained to detect the presence and location of various groups of objects

III. Outdoor Navigation

Walking securely in the outside environment is one of the main needs for visual impairers. The proposed method for outdoor navigation in Pro-Eye includes mainly three stages: supply user with directions from user location to the destination [11], identify two edges in the walking lane to navigate user inside the lane, generate navigation rules for the user to help the user avoid collisions [12].

Directions are supplied to the user using Google Directions API. And the lane line edge detection functionality has been mainly implemented using three steps.

1. Find and define the ROI (Region of Interest)
2. Image masking
3. Apply Hough Line Transform technique

Pro-Eye uses Hough Line Transform Algorithm to identify the final output or the straight lines which are road edges. The Hough Transform is an algorithm that was initially implemented to perceive complex lines in photos. To improve the accuracy in this lane edge detection functionality, Pro-Eye has used background subtraction and filtering.

The system uses CLIPS ide to develop the expert system for the navigation rule generation purpose. An expert system is a computer system that emulates the decision-making ability of a human expert [13]. The overview diagram of the expert system is illustrated in Fig. 4.

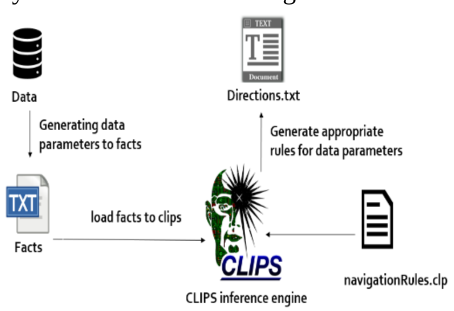


Figure 4: Overview of generating navigational rules

The data parameters use in Pro-Eye inference engine is presented in Table II.

Table 2: Description of the data parameters

Parameter Name	Description
FrameNo - A	Number of the video frame of the relevant video stream
StreetWidth - B	Starting width of the lane in the relevant frame
EndWidth - C	Ending width of the lane in the relevant frame
ObjectWidth - D	Width of the object
ObjectHeight - E	Height of the object
ObjectDepthFromBeginning - F	Object depth from the beginning of the lane
ObjectToLeftBoarder - G	Length to the left edge of the lane from the object
ObjectToRightBoarder - H	Length to the right edge of the lane from the object
FramePoint - I	Depth to the frame pointer from the beginning of the frame (normally, frame point is in ½ of lane width and ¼ of lane height)

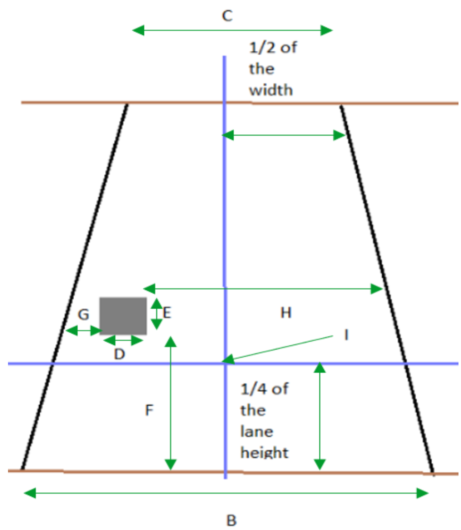


Figure 5: Data parameters preview

As illustrated in Fig. 5. the data parameters are taken by the system.

CLIPS ide can represent data using rules and facts formats. Therefore, the application transforms these data parameters to the facts format. Pro-Eye has stored navigational rules inside CLIPS ide working memory and the relevant rule will be triggered according to the variations of facts or the data parameters.

Finally, the inference engine creates an extra slot in facts which is the output that is used to navigate the blind person. Rules have been implemented in a manner to guide the user in situations encountered by a person while moving in an outdoor walking lane, and the actions that a person has to take to avoid walking away from road edges as well as to avoid collisions with obstacles. Each example rule consisting input elements such as the distances of the obstacles to left, to right, and in front of the person and the output element giving the change in the walking direction of the person being required in response to the input data. The first rule in the inference engine has been built so, if the length to object from the beginning is greater than 1m and the frame point is near to the left border, which is less than 0.5m, a rule will be created as a guide to the right. This is just a basic rule.

IV. Text Detection and Graph Reading

When the user requests a material to be read using voice the AI API framework will invoke the test recognition functionality which uses the outputs of the object recognition to recognize text. The pro-processed image frames will be taken as the input to the trained models in the application which will be then processed. It will give the text and description of the image in JSON format. The text-to-speech (TTS) system will convert the text in digital format to an audio format which the user will be able to listen to using headphones. The text reading system is implemented under two major categories.

- Text and symbols recognition.
- Graph reading.

Since the real-life, text is more sophisticated, the system uses an ML kit on java which uses a detection feature on TensorFlow conventional ML approach. To help the computer understand and match the character, a detection feature needs to be introduced. The detection feature runs through neural networks of inferences hence, detection feature or intelligent character recognition (ICR) is used to identify characters' traits [14]. It searches for letters that meet the specified requirements.

For example: If the system recognizes a longer vertical line meeting a shorter horizontal line at a right angle, that is the letter "L". These angles and lines are recognized as mentioned above.

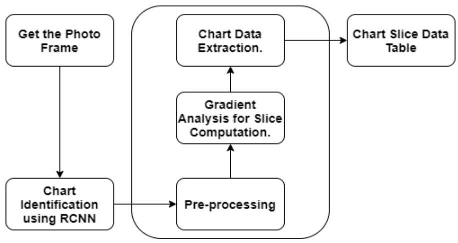


Figure 6: Graph reading system

Fig. 6. illustrates the steps carried out in detecting charts in the frame, if the diagram picture containing the border of the pie chart and specific borders of various neighboring slices is acquired, the next step is to classify each pie chart slice [15]. For that, invert the image. Then all the slices of the pie chart get separated. But the method of inversion added another dimension that exists outside the boundary of the pie chart and always touches the boundary of the image [16]. For automatic data extraction from a pie chart, boundary touching components should be removed. Therefore, Connected Component Analysis (CCA) is used to recognize the boundaries of the pie chart [17]. CCA removes the boundaries and contains only the chart slices as illustrated in Fig. 7.

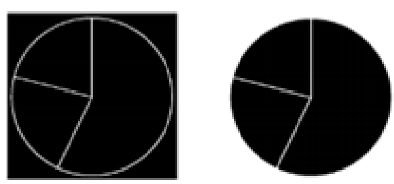


Figure 7: CCN converted chart.

For data extraction, CCN carries out the total number of pixels in all the slices. And CCN provides the number of pixels in each slice.

Let us assume,

Total number of pixels in the pie chart = X

Number of pixels in a slice = Y

Therefore, we can compute the percentage of the slice as:

Percentage of the slice = Y/X×100 % (1)

Using the above algorithm, this system will be able to fetch the percentages of slices when providing a photo of a pie chart. After fetching data from the pie chart, the data packets will send to the cluster as illustrated in Fig. 8.

Research discussion and results

A survey was carried out among 44 visually impaired personalities to measure the usage and effectiveness of an Assistive Technology.

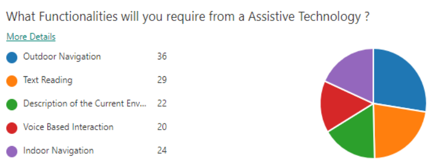


Figure 8: Survey Result 1

Fig. 8, Shows that almost 50% of the response included all the functionalities provided by the proposed system.

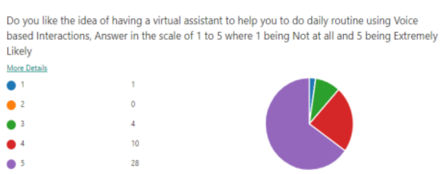


Figure 9: Survey Result 2

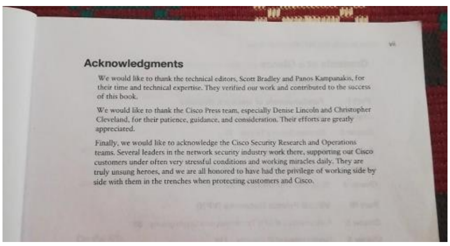
Fig. 9 illustrates that most of the participants are extremely likely to use a voice-based virtual assistant to help to do daily routines.

Results of the Outdoor navigation functionality provide the directions to the destination from the user's current location from the JSON format and Identified edges of the lane. Which is converted to voice and conveyed to the user in a human-understandable format.

Figure 10: Input frame

Figure 11: Output frame

As illustrated in Fig. 10. and Fig. 11. the input frames and the output frame after the application identified the lane edges. When the user is 0.5m near to a lane edge, the application generates a beep sound, so the user knows he is near to the lane edge. And then all the data parameters used in the training phase to generate navigation rules must be received to the application in real-time to generate rules using the inference engine. To that, coordinates of the obstacles and the road must be tracked using sensors. The more accurate the coordinates are, the more trustworthy rules become. That is the major limitation of the outdoor navigation functionality.


Acknowledgments

We would like to thank the technical editors, Scott Bradley and Panos Kampanakis, for their time and technical expertise. They verified our work and contributed to the success of this book.

We would like to thank the Cisco Press team, especially Denise Lincoln and Christopher Cleveland, for their patience, guidance, and consideration. Their efforts are greatly appreciated.

Finally, we would like to acknowledge the Cisco Security Research and Operations teams. Several leaders in the network security industry work there, supporting our Cisco customers under often very stressful conditions and working miracles daily. They are truly unsung heroes, and we are all honored to have had the privilege of working side by side with them in the trenches when protecting customers and Cisco.

Figure 12: Text in a Tangible Material

AcknowledgmentsWe would like to thank the technical editors, Scott Bradley and Panos Kampanakis, for
their time and technical expertise. They verified our work and contributed to the successof this book.We would like to thank the Cisco Press team, especially Denise Lincoln and Christopher
Cleveland, for their patience, guidance, and consideration. Their efforts are greatlyappreciated.Finally, we would like to acknowledge the Cisco Security Research and Operations
teams. Several leaders in the network security industry work there, supporting our CScocustomers under often very stresstul conditions and working miracles daily. They aretruly unsung heroes, and we are all honored to have had the privilege of working side byside with them in the trenches when protecting customers and Cisco.

Figure 13: Text Recognized by the system

The system was able to recognize the text on a printed material shown in Fig. 11 using ICR and the output produced is illustrated in Fig. 13. Furthermore, the accuracy of the recognized text was evaluated by providing 50 text materials with different levels of font sizes and calculated the word-level accuracy.

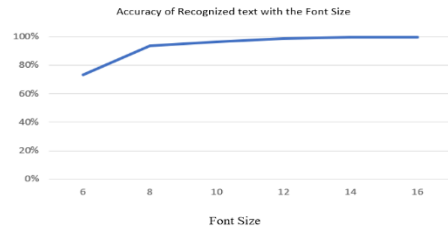


Figure 14: Accuracy of text recognition on various font sizes

As shown in Fig. 14. the result proved the trend accuracy of the system increases exponentially when the font size is above 8.

Table 3 shows a feature comparison between GuideCane [3], VisualPal [4], DAVID [18], and Pro-Eye, with this comparison we can derive that the proposed system accommodates more assistance than others.

Table 3: Feature Comparison

Feature	DAVID	Guide Cane	Visual Pal	Pro-Eye
Outdoor Navigation	✗	✓	✗	✓
Text and Graph Recognition	✓	✗	✗	✓
Object Recognition	✗	✗	✓	✓
Voice Interaction	✓	✗	✓	✓

CONCLUSION

Efficient and effective decision-making comes through independence, there is a strong gravitational pull the decision will impact when the decision was made independently. The person who is always dependent on someone will tend to make decisions from the input from others which is in the case of the visually impaired is a pretty much-affecting factor. Even to take minor decisions a visually impaired personality will have hesitation, which can rank them behind a visual health person. Pro-Eye will be a Virtual Assistant which helps users to navigate, understand the environment and to read texts using a collection various component located on the device.

Acknowledgment

It is a great appreciation that we convey our gratitude to the Sri Lanka Institute of Information Technology (SLIIT) for providing us an opportunity and appropriate environment to complete this research.

REFERENCES

Mahawariya, K., 2019. A Study of Needs and Requirements of Visually Impaired Students of University of Delhi. International Journal of Information Dissemination and Technology, 9(2), p.83.

Laehyun Kim, Sehyung Park, Sooyong Lee, Sungdo Ha "An electronic traveler aid for the blind using multiple range sensors" IEICE Electronics Express Vol. 6, No. 11, pp. 794-799, 2009.

Iwan Ulrich, Johann Borenstein "The GuideCane - Applying mobile robot technologies to assist the visually impaired" IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans, Vol. 31, pp. 131-136, 2001.

Shagufta Md. Rafique Bagwan, L. J. Sankpal, 'VisualPal: A mobile app for object recognition for the visually impaired', [Online] Available: https://ieeexplore.ieee.org/document/7375665 .

Spacy.io, 'Training spaCy's Statistical Models, [Online] Available: https://spacy.io/usage/training.

Anaconda.org, 'CNN trained on OntoNote, with Glove vectors', [Online] Available: https://anaconda.org/conda-forge/spacy-model-en_core_web_lg.

Hanieh Poostchi, Massimo Piccardi, 'A multi-constraint structured hinge loss for named-entity recognition' Australasian Language Technology Association 2019, [Online] Available: https://www.aclweb.org/anthology/U19-1006.pdf.

Giuseppe Ciaburro and Prateek Joshi, Python Machine Learning Cookbook. Birmingham: Packt Publishing Ltd, 2019.

"CameraX overview | Android Developers", Android Developers, 2020. [Online]. Available: https://developer.android.com/training/camerax. [Accessed: 12- Jul-2020].

TensorFlow. 2020. Object Detection | Tensorflow Lite. [online] Available at: https://www.tensorflow.org/lite/models/object_detection/overview [Accessed 10 July 2020].

SAGE Journals. 2020. Navigation Application Programming Interface Route Fuel Saving Opportunity Assessment on Large-Scale Real-World Travel Data for Conventional Vehicles And Hybrid Electric Vehicles - Lei Zhu, Jacob R. Holden, Jeffrey D. Gonder, 2018. [online] Available at: https://journals.sagepub.com/doi/abs/10.1177/0361198118797805

Balani, Y., Narayanan, D., Parande, S., Birari, A. and Yeole, A., 2019. Drishti – A Smartphone Application for Visually Impaired. SSRN Electronic Journal.

J. Pothal and D. Parhi, "Navigation of multiple mobile robots in a highly clutter terrains using adaptive neuro-fuzzy inference system", Robotics and Autonomous Systems, vol. 72, pp. 48-

58, 2015. Available: 10.1016/j.robot.2015.04.007.

"Simple OCR implementation on Android with Google's ML Kit | TSH.io", The Software House, 2020. [Online]. Available: https://tsh.io/blog/simple-ocr-implementation-on-android-with-googles-ml-kit/. [Accessed: 17- April- 2020].

W. Huang, R. Liu, and C. L. Tan. Extraction of vectorized graphical information from scientific chart images. In Proc. of the 9th Interna-tional Conference on Document Analysis and Recognition (ICDAR'07), pp. 521–525, 2007. [Accessed: 17- June- 2020].

J. Gao, Y. Zhou, and K.E. Barner. View: Visual information extraction widget for improving chart images accessibly. In Proc. of the 19th IEEE International Conference on Image Processing (ICIP'12), pp. 2865–2868, 2012. [Accessed: 17- June- 2020]

"Automatic Data Extraction from 2D and 3D Pie Chart Images - IEEE Conference Publication", Ieeexplore.ieee.org, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8692104. [Accessed: 17- April- 2020]

Marvin, E., 2020. Digital Assistant for The Visually Impaired - IEEE Conference Publication. [online] Ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/document/9065191 [Accessed 12 September 2020].