Color Sensor for Visually Impaired People

G.H.M. Liyanage (E/16/207), A.M.P.D. Muthukuda (E/16/243),
R.T. Nanayakkara (E/16/246), L.B.I.P. Thilakasiri (E/16/367)

Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya

***Abstract*—Visual impairment is a major limitation for a large number of people in their attempt to completing basic day-to-day activities. Total blindness and Color blindness are two of such issues that has to be addressed. This paper addresses these two issues using an Artificial intelligence based object detection method and a grid based dominant color information extraction method. These two are presented using a low cost and reliable device that can be used to complete multiple tasks.**

***Index Terms*—Multi-mode operation, Color identification, Object detection, 3D audio generation, Power saving, Ergonomics**

I. Introduction

Our sense of sight is responsible for most of the information we absorb from our five combined senses. Many of the movements we perform, tasks we complete, and personal interactions we make rely on vision in some way. But unfortunately, according to the WHO, over 36 million people are completely blind and 217 million people suffer from moderate to severe visual impairment (MSVI). In our project we address this problem by giving sound feedback on the objects and colors they see. We chose sound feedback because such individuals develop better sensory systems because they depend on other sensory information such as sound.

II. Related Work

Over the years, many researches has been done to address the above issues and people with color blindness have difficulties only in distinguishing certain colors. Re-coloring the images after processing is a technique that has been proposed in multiple researches considering the deficiency of color perception. Surveys has been done on the effectiveness of the recoloring algorithms for the color blind [1], and several other user studies has been done to understand more about the interface recoloring aspects of the colorblind [2]. Enhanced re-coloring method with an information preserving property using effective color clustering [3] has been done and other novel attempts at image re-colorization [4] [5] have also been implemented. On the other hand, building augmented reality systems for better user experiences has been conducted in a number of instances [6] [7]. Use of smart glasses has been cleverly implemented by companies such as Google, Amazon in their smart glass lineups but the exact use of those for the colorblind can not be considered as convenient considering the built and cost structures. Other attempts to introduce smart glasses for the colorblind has also been conducted [8]. Considering the sound feedback systems, experimental systems for the conversion of images into sound patterns [9] has been conducted and other devices with assistive/ interactive interfaces and technologies as well [10] [11] [12] [13].

Researches with devices with a screen and a microprocessor has been done by [14] and mobile applications have also been developed to introduce an adaptive interface scheme with sensing of the colorblindness type and switch to respective color-blind mode [15]. Furthermore, mobile applications which records and converts the color information into beep sounds and vibration pulses with the aid of a smartphone has also been undertaken [16]. In this paper we have presented a low cost method which is convenient for the user in a number of avenues such as low power usage, both object detection and color information extraction, and multi-mode operation.

III. Idea

In this paper the above issues are addressed using sound feedback about the surroundings of the individual using a user friendly manner and a control panel. Sound feedback is used to give information about the color & surrounding objects because, visually impaired individuals develop better sensory systems as they depend on other sensory information such as sound.

IV. Product

The product is focused on helping visually impaired individuals to detect objects and colors and give a basic understanding about the surroundings. This is accomplished by capturing the images of the eyesight using a camera and using image processing techniques to detect the dominant color of the targeted area and by using machine learning techniques to identify objects.

Raspberry Pi module was chosen as the main processing unit and other attachable components such as the camera module, control panel with buttons, were chosen as a part of a simple design in mind to ease the use of the device.

Key features of the product :

1) Use of 3D sound which enables the user to get a perception of location
2) Scan mode, which generates a caption of explaining the image captured from the camera
3) Simple functionality which allows the user to operate the product easily
4) Ergonomic considered design which makes the handling of the device easy to all users

5) Operability upon request by the user allows power saving and higher durability of the product.

A. Technologies Used

- Python - Used as the code base and was chosen for the application considering the ease of use, support for many different packages and because of its overall versatile nature.
- Raspberry Pi - Raspberry Pi 3B model is used as the processor of this product. Image processing, sound generation is done by the Raspberry Pi and its Tensorflow support is another reason for this choice.
- Computer Vision - Used for the processing of the camera feedback and other visual inputs.
- Spatial Audio Generation - Once the colors are identified using the software, a predefined sound corresponding to the color will be played. This sound will be generated in a manner such that the sound intensity will be maximum in the region of the color. This will allow the user the to get a spatial perception of the color.
- Text-to-Speech Generation - When camera captured the image, the objects of the image is identified by the corresponding Neural Network and after identifying the object class, Text-to-Speech generation is used to generate the object class name as a sound output.

B. Modes of Operation

There are two modes of operation.

- Normal Mode
  Under normal mode operation, the image fed by the inbuilt camera is partitioned into a 3x3 grid. Each segment is processed separately to identify the dominant color inside each grid. The processing is done starting from the left to right from the top row and continue to the next two rows. Thereafter, a unique sound corresponding to the detected dominant color will be generated with the horizontal location information. These sounds are used to create the 3D sound effect to give a perception of location to the user.
- Scan Mode
  Scan mode has color scan & object scan modes. Under the object scan mode, the center area of the image from the camera feed is analyzed to generate a caption stating the objects in the image and their properties. In the color scan mode, the dominant color of the center area of the image is conveyed to the user via the generated sound.

C. Color Identification & 3D Audio Generation

The dominant color information is saved for each grid and its inner workings is described further in Section VI. The color feedback about the respective area is conveyed to the user with the use of 3D sounds generated by the system, according to the input of the image. These 3D sound effects are predefined with each sound representing a different color. Usually, visually impaired individuals develop better sensory systems because they depend on other sensory information such as sound than the visual input. This is the basis of our sound effect system and when outputting the audio of the relevant color, 3D sound is used in order to provide a better and more user-friendly perception of the surrounding. To give better information, the image input is divided into a 3x3 matrix and the dominant color within each unit cell is identified and the sound corresponding to each cell is returned.

D. Object Identification

The object identification of the image captured is carried out under the scan mode. The object within the center most grid is identified using machine learning techniques and the object is notified to the user via audio (example: “dog”,”car”). If there are more than one object present, information about all those classes is passed on to the user. Apart from this, the dominant color of the center most grid is sounded allowing the user to get a better idea of the object.

E. Other Competitive Products

Caption generating devices using camera feeds are already available in the market. But, these products are typically in the range of \$1800 - \$6000 and are not affordable for middle class and low income people, whereas our product can be produced below \$400.

Some similar products that are currently available in the market :

- Iris Vision
- Orcam
- Envision
- eSight

V. Ergonomic Aspects

The hardware component designing of the product focused mainly on the usability and comfort of the user. The product consists of two major components,

- Wearable pair of glasses
- Processing unit

Wearable pair of glasses are indistinguishable from an ordinary pair of shades which does not cause any discrimination in the society for the user.

The processing unit weighing a maximum 0.2kg, is of dimensions 9cm x 6cm x 3cm, which allows the user to carry it around easily in a pocket. The processing unit will be placed inside a casing of rating IP55, which enhances the durability and protection of the processing unit.

The ability to use the device with any pair of headphones/earbuds with a 3.5mm jack adds another user friendly aspect of the product.

VI. Progress

A. Gantt Chart

1) Problem defining and possible solutions consideration: For the problem defined in Section I, few possible methods of identifying colors and providing the user with an output

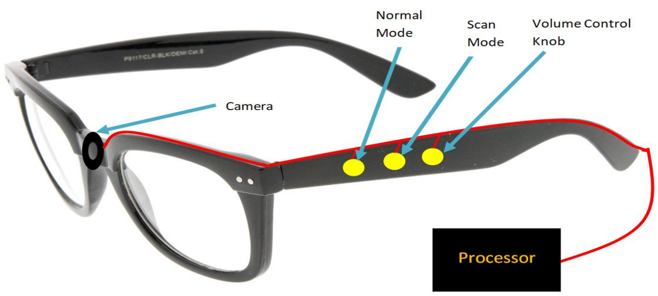

Fig. 1. Wearable pair of glasses with control knobs

Fig. 2. Final Prodcuct

were considered.

- ***Kmeans clustering method and Elbow Plot approach***
  First, the image was fed to the program and all the pixel values were considered for computation. The RGB values of each of those pixels were categorized and clustered using KMeans clustering and the Elbow plot approach was taken to identify the number of dominant colors. This approach was computationally expensive and the elbow plot approach is comparably tedious rather than an automated approach.

- ***Using an Online Color Map Dataset directly***
  Directly using some of the online available datasets containing RGB color information was also considered for dominant color information extraction. But, as these datasets contain several millions of colors, but are not categorized into a limited number of color classes, using this to give a visually impaired person a more generalized impression about color and is not convenient. Therefore, it was decided to create our own dictionary of colors during image processing.

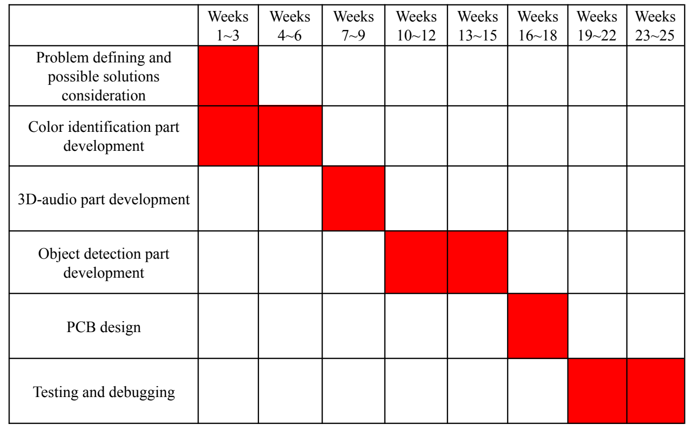

Fig. 3. Gantt chart

2) Color identification component development: After considering different methods, the situation of the user-base & speed, the following method was chosen as a combination of the above mentioned methods to make an improved & more suited approach for extracting color information.

- ***Clustering, Sliding Window based dominant color identification and Euclidean Distance based method***
  At first, the image is divided into a 3x3 grid as shown in Fig. 4. and taking a similar approach to how the Convolutional Neural Networks work, a small window is passed through the image. Here the dominant color information of each instance is recorded, and divided to 10 clusters. This process is done for all 9 grids separately the RGB values corresponding to dominant color is recorded. At the initial stage a dictionary containing more than 160 colors were created by us and the dominant color recorded for each grid is then passed into another function where the euclidean distance is calculated between the dominant colors and the colors available in the dictionary created by us. The color with the minimum distance is then selected as the output.

The above mentioned selected method was developed using python and the code showed promising results against the test images. The output generated by the program at the end of processing the Fig. 4 is shown in Fig 5.

3) 3D-audio component development: Once the images are scanned to identify colors, the corresponding color information is passed on to the user from left to right with audio feedback. The corresponding color information is conveyed using a predefined Color-Sound map and as shown in Fig 6. The generated 3D sound wave generated for the left and right earbuds for some test images is shown in Fig 7.

4) Object detection component development: The main issue with using low computational power devices for machine learning or computer vision tasks is its lack of performance. Therefore, normal neural networks can not be implemented in such devices and have good results. Quantized neural network

Fig. 4. Test image for color identification

```
Dominant Color Matrix:

 green,    orange,    red,
 green,    orange,    red,
 green,    orange,    red,
>>> |
```

Fig. 5. Results obtained for Fig. 4

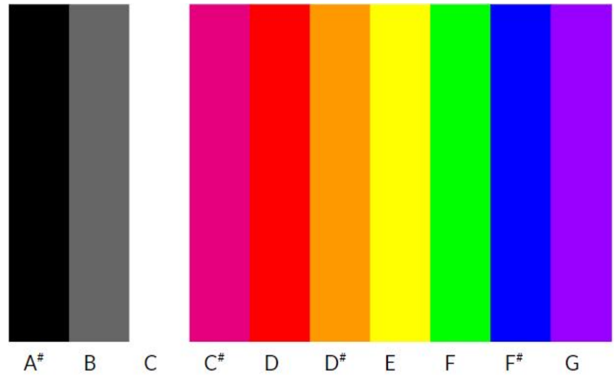

Fig. 6. Color-sound map

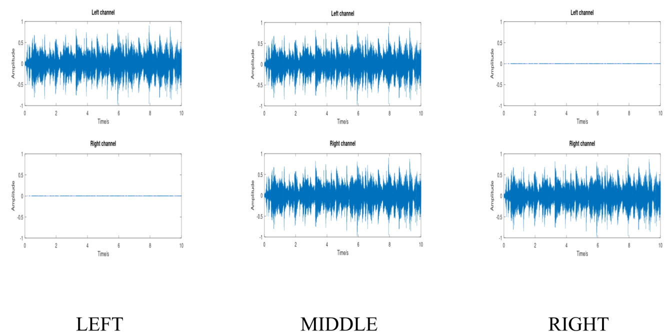

Fig. 7. Sound wave outputs at left and right earbuds for grids scanned from left to right

approach can be seen as a solution to this issue and there are a few models that has already been made publicly available by Tensorflow for general use. Such an object detection quantized neural network is used in the product for the object detection component. It has been trained on the MS COCO data set and around 80 item classes can be identified using this data set. Once the item classes are identified, the class name is conevrted into "text-to-speech" format using a native python library and the results of the object detection and classification run for a test image is shown in Fig. 9 & Fig. 10.

5) Hardware Implementation: The hardware implementation of this device was completed and it was tested for various different conditions and images that were captured from the camera. The setup consists of the processor unit, the button array for the multi-mode operation, the camera module, and the power supply. Unfortunately due to the current situation of the country, we could not get the required smaller camera modules, therefore we used a smartphone camera as the camera unit in the hardware implementation. All the software implementations were tested and validated using the hardware setup.

6) PCB design: In the beginning of the project, a single PCB unit was decided to be attached to the processing unit with the controls, placing all the knobs and the buttons on the surface of the processing unit. But considering the unfriendliness it brings to the product, all the knobs and buttons were decided to be attached to the side of the wearable pair of glasses by separating the two PCBs. The finalized PCB layouts are shown in Fig. 11 & Fig. 12.

7) Testing and debugging: All components of the software were tested and the results obtained for some images are included in the video.

B. Awards and Achievements

This product was presented to two competitions which was held during the semester and their information is mentioned below.

- Third Place, IEEE Electronic Design Competition (EDC) - IEEE Sri Lanka Section (Fig. 14) [1]
- Top 5 Finalists out of 150+ Teams - HackX 2021 (Fig. 15)

VII. Future Expansions

- Wireless data transmission
- Multi-language support

[1] https://www.facebook.com/IEEESriLanka/videos/348328730369252/

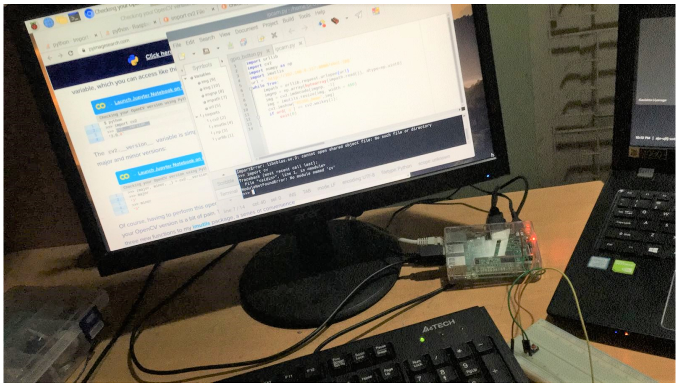

Fig. 8. Hardware setup used for testing

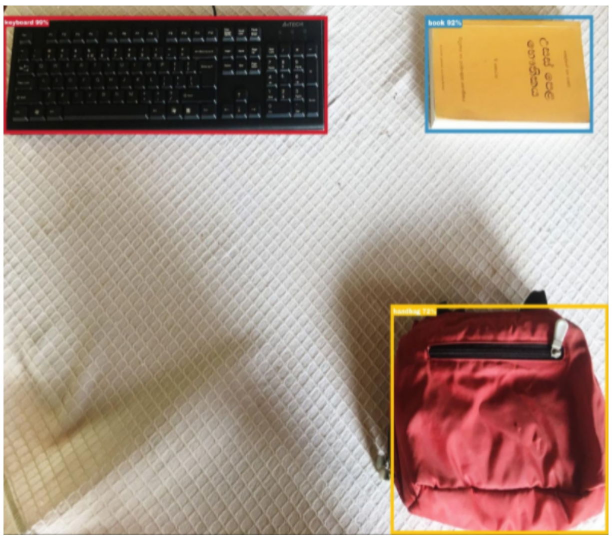

Fig. 9. Test image for object identification

- Higher number of object classes identification
- Descriptive caption generation
- Real time video processing
- Voice control capability

VIII. FAQ

1) Why do we use a sound color map instead of text to speech?
   - Both modes can be used depending on the preference of the user
   - The sound information can be passed to the users as the name of the color and also as the respective sound mapped to that particular color
   - Blind people are a lot more sensitive to sound
   - The artistic approach of mapping the colors to distinct sounds was first proposed by one of the blind persons that we talked to
2) Why don't we do live video stream for processing?
   - Processing video in real-time is a highly computationally cost process & mobile devices are not really suitable for the task

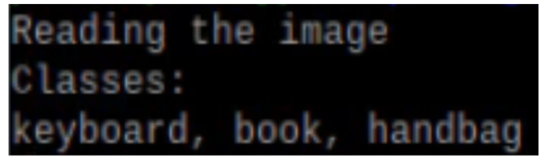

Fig. 10. Results obtained for Fig. 9

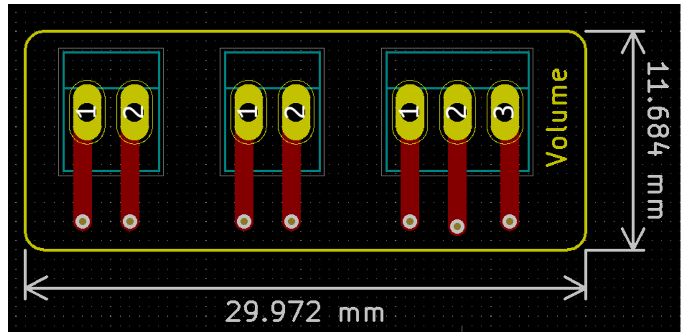

Fig. 11. PCB attached to the spectacle frame

   - Live or continuous image & video processing requires a lot of power to have an all day battery life
   - Continuously running too many heavy tasks could reduce the lifespan of the device
3) Why haven't we used other types sensors available to get a better sensor input?
   - Having a bulk of sensors and having another special unit for the users to wear is inconvenient and it also creates a social gap between them and the rest of the society.
   - Too many inputs to the user can be distracting & unnecessary
   - Our way is cleaner
4) Why have we not included obstacle avoidance?
   - Blind people rely heavily on physical feedback for obstacle detection
   - They feel uneasy to completely depend on audio feedback
5) How do you power up your device?
   - Main power supply is a power bank
   - The model requires roughly 2A during the full mode of operation.
   - If we use a 10000mAh power bank, the device can operate around 5 hours at continuous full mode of operation. Therefore, the user will be able to use the device comfortably for 7-8 hrs.
   - An audio signal to indicate if that the battery level is below 10% has also been included.

ACKNOWLEDGMENT

Project team members are grateful to the project supervisors, Dr. J.V. Wijayakulasooriya, Dr. R.D.B. Ranaweera, Dr. U.S. Navaratne for giving us important insights and information on how to make the product more complete and fruitful. In

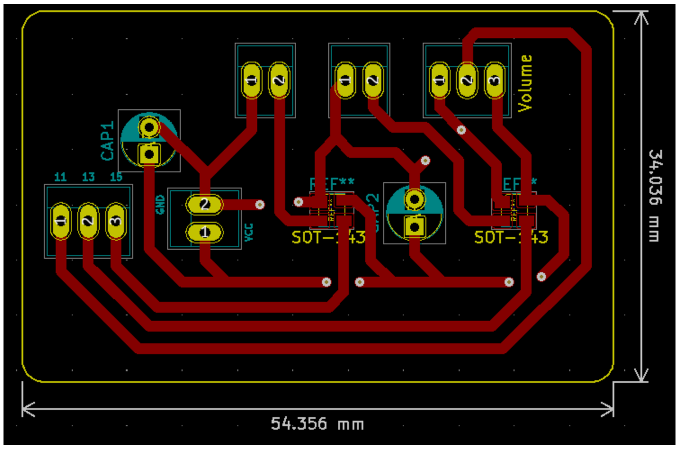

Fig. 12. PCB included in the control box

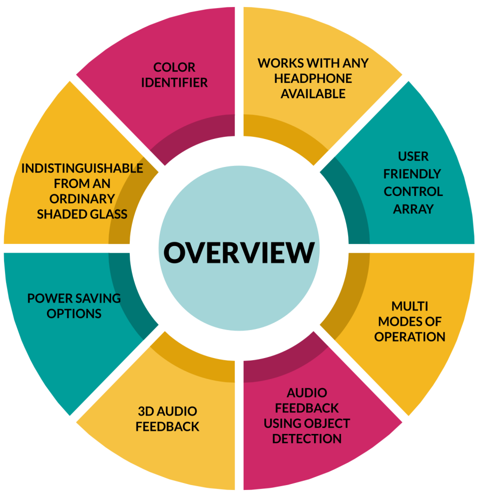

Fig. 13. Overview of the product

addition, members would like to thank our mentor from the HackX competition, Mr. S. Sathananthan from Sysco Labs for helpful suggestions.

References

[1] M. Ribeiro and A. J. Gomes, "Recoloring algorithms for colorblind people: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–37, 2019.

[2] R. J. de Araújo, J. C. Dos Reis, and R. Bonacin, "Understanding interface recoloring aspects by colorblind people: a user study," *Universal Access in the Information Society*, vol. 19, no. 1, pp. 81–98, 2020.

[3] J.-Y. Jeong, H.-J. Kim, Y.-H. Kim, T.-S. Wang, and S.-J. Ko, "Enhanced re-coloring method with an information preserving property for color-blind person," in *2012 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2012, pp. 600–601.

[4] J.-B. Huang, C.-S. Chen, T.-C. Jen, and S.-J. Wang, "Image recolorization for the colorblind," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1161–1164.

[5] J.-Y. Jeong, H.-J. Kim, T.-S. Wang, Y.-J. Yoon, and S.-J. Ko, "An efficient re-coloring method with information preserving for the color-blind," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1953–1960, 2011.

[6] B. S. Ananto, R. F. Sari, and R. Harwahyu, "Color transformation for color blind compensation on augmented reality system," in *2011 International Conference on User Science and Engineering (i-USEr)*. IEEE, 2011, pp. 129–134.

Fig. 14. IEEE Electronic Design Competition - IEEE Sri Lanka Section

Fig. 15. HackX Top 5 Finalists out of 150+ Teams - Team Transients

[7] E. Tanuwidjaja, D. Huynh, K. Koa, C. Nguyen, C. Shao, P. Torbett, C. Emmenegger, and N. Weibel, "Chroma: a wearable augmented-reality solution for color blindness," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 2014, pp. 799–810.

[8] J. Ruminski, "Color processing for color-blind individuals using smart glasses," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 8, pp. 1652–1661, 2015.

[9] P. B. Meijer, "An experimental system for auditory image representations," *IEEE transactions on biomedical engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[10] M. A. Hersh and M. A. Johnson, *Assistive technology for visually impaired and blind people*. Springer, 2008.

[11] R. Velázquez, "Wearable assistive devices for the blind," in *Wearable and autonomous biomedical devices and systems for smart environment*. Springer, 2010, pp. 331–349.

[12] T. L. McDaniel, K. Kahol, and S. Panchanathan, "An interactive wearable assistive device for individuals who are blind for color perception," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2007, pp. 751–760.

[13] J. Brabyn, K. D. Seelman, and S. Panchang, "Aids for people who are blind or visually impaired," in *An introduction to rehabilitation engineering*. CRC Press, 2006, pp. 313–340.

[14] J. McDowell, "Design of a color sensing system to aid the color blind," *IEEE Potentials*, vol. 27, no. 4, pp. 34–39, 2008.

[15] M. W. Iqbal, S. K. Shahzad, N. Ahmad, A. Amelio, and D. Brodic, "Adaptive interface for color-blind people in mobile-phones," in *2018 International Conference on Advancements in Computational Sciences (ICACS)*. IEEE, 2018, pp. 1–8.

[16] S. Bandyopadhyay and B. B. Rathod, "The sound and feel of titrations: A smartphone aid for color-blind and visually impaired students," 2017.