**DOCUMENT READING SYSTEM FOR BLIND PEOPLE (“READING EYE”)**

E.M.D.D Ekanayake

(IT16102156)

Bachelor of Science (Hons) in Information Technology Specialized in

Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

08th May 2020

**DOCUMENT READING SYSTEM FOR BLIND PEOPLE (“READING EYE”)**

E.M.D.D Ekanayake

(IT16102156)

Bachelor of Science (Hons) in Information Technology Specialized in

Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

08th May 2020

# DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Signature: Date: 04th of May 2020

Signature of the Supervisor: Date: 04th of May 2020

Signature of the Co-Supervisor: Date: 04th of May 2020

# ABSTRACT

Reading is an essential ability a person should have. Without reading you are pretty much futile. Because to complete day-to-day tasks you need to read at least once a day. From the email that were sent by your boss to a bill for something you purchased from store when you are heading home, you have to read. For the most part, we take it for granted. We do not think much about it because we have no problems with reading a bill with our own eyes. But for a blind user, the case is entirely different as it is not an easier task for them to read, even with Braille.

Thus, this proposal proposed a Virtual Reading Assistant for blind and vision impaired people, which can be used to read text, images, charts, equations and table-based content in a document. The proposed mobile application will reduce the time to read a document efficiently. Particularly when comparing with a vision impaired person who usually read using Braille codes, it takes quite some time to read a document. According to several researches that were conducted regarding Braille reading speed, they concluded Braille was a slower method. Therefore, the proposed mobile application will reduce the reading time and make reading a pleasurable experience for the user. Moreover, the security of the application will be enhanced.

In implementation process of this application, we will train different neural network models to detect and understand objects from the digital images of the document, which will be captured by the camera in the app. After analyzing the content of images, specific descriptions will be created using natural language processing techniques. Finally, we generate an audio file based on the previously created content description for the user to listen.

# Acknowledgement

First, I would like to express my gratitude to supervisor **Ms. Suranjini Silva** and co-supervisor **Dr. Anuradha Jayakody** for the guidance and mentoring given throughout this period to make this research a success. Also, I would like to thank senior lecturer and lecture in-charge for Comprehensive Design and Analysis Projects module, **Dr. Janaka Wijekoon** for the valuable assistance given to complete this project. At last but not least I am also thankful to all my family members and colleagues who have been always helping and encouraging me throughout the research period. I have no valuable words to express my thanks, but my heart is still full of the favors received from every person.

# Table of Content

[Declaration i](#_Toc19501641)

[Abstract ii](#_Toc19501642)

[Acknowledgement iii](#_Toc19501643)

[Table of Content iv](#_Toc19501644)

[List of Figures v](#_Toc19501645)

[List of Abbreviations v](#_Toc19501646)

[1. Introduction 1](#_Toc19501647)

[1.1 Background Literature 2](#_Toc19501648)

[1.2 Research Gap 5](#_Toc19501649)

[1.3 Research Problem 6](#_Toc19501650)

[1.4 Research Objectives 7](#_Toc19501651)

[2. Methodology 8](#_Toc19501652)

[2.1 Methodology 8](#_Toc19501653)

[2.2 Commercialization aspects 10](#_Toc19501654)

[2.2.1 Cloud based revenue model 11](#_Toc19501655)

[2.2.2 On premises revenue model 11](#_Toc19501656)

[2.3 Testing and Implementation 11](#_Toc19501657)

[3. Results and discussion 18](#_Toc19501658)

[4. Conclusion 21](#_Toc19501659)

[5. References 22](#_Toc19501660)

# List of Figures

[Figure 1‑1 WebRTC direct browser connection establishment 4](#_Toc19501523)

[Figure 1.2.1 Conceptual model for video routing when using simulcast 6](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501524)

[Figure 2.1.1 High Level Architectural Diagram 8](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501525)

[Figure 2.1.2 Feeds broadcasting within a session 9](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501526)

[Figure 2.1.3 Feed recording process 9](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501527)

[Figure 2.1.4 Post processing of recordings 10](#_Toc19501528)

[Figure 2.3.1 Live Streaming UI for lecturer 12](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501529)

[Figure 2.3.2 Screen Selection UI for Sharing screen 13](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501530)

[Figure 2.3.3 Application Selection UI for Screen Sharing 13](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501531)

[Figure 2.3.4 Preview of what lecturer shares on an ongoing stream 14](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501532)

[Figure 2.3.5 Input Device Selection Window 14](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501533)

[Figure 2.3.6 UI for Current Live Streams Available to Join 15](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501534)

[Figure 2.3.7 Student's view of a live stream 15](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501535)

[Figure 2.3.8 Live Q&A Sessions UI 16](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501536)

[Figure 2.3.9 List of previously recorded lecture streams 16](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501537)

[Figure 3.1 CPU consumption 18](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501538)

[Figure 3.2 Memory Consumption 18](#_Toc19501539)

[Figure 2.3. 11 Network In of AWS instance 19](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501540)

[Figure 2.3. 12 Network Out of AWS instance 19](file:///F:\SLIIT\Year%204\CDAP\Thesis\19-079_IT16131002_ThesisFinal.docx#_Toc19501541)

# List of Abbreviations

|  |  |
| --- | --- |
| WebRTC | Web Real-Time Communications |
| RTP | Real-time Transport Protocol |
| PTZ | Pan Tilt Zoom |
| CPU | Central Processing Unit |
| SFU | Selective Forwarding Unit |

# Introduction

Reading is more than just a pastime for bookworms. It is a skill we each use on a daily basis. A work email, text message, street sign, or even a status on Facebook all require you to read. In addition to these everyday tasks, many people read to learn information or facts, to be entertained, or to understand other cultures or groups.

One advantage of reading is to gain a more thorough understanding of something. Be it a story in a novel or even a technical topic, you get more understanding of the subject by reading a related book than watching a video.

To elaborate more on that, if you code a React project by watching a video, you will by no means learn the syntaxes and complete the project. But the theories behind will not be much clear to you. You will not know why one method was used instead of another. At the same time, if you read a book about React, you will learn all the theories behind the framework more deeply.

A second advantage has to do with reading comprehension, which encompasses the skills needed to understand the meaning of a text. Reading is pointless unless you gain some understanding of the text. Be it for entertainment, learning, or communication, you need to be able to comprehend what you read. The more you read, the more your reading comprehension will improve. Other than that, reading helps us to reduce stress, expand our vocabulary, stronger analytical thinking skills, and of course, free entertainment. There are more like this, just cannot say enough.

Reading is not a difficult task for a person who has the ability to see. However, but how about a person who is lost his/her sight entirely or someone who is not fully but partially blind? How do they read something? Especially a book? How hard could it be for them to read?

Over the years, countless methods were tried to enable blind people to read and write independently. Out of them, the only braille system was successful. Braille is not a language. It's a code. It was developed by a Frenchman who lost his sight as a result of a childhood accident. In 1824, at the age of fifteen he developed a code system for French alphabet and published it. Braille system uses raised dots to represent the letters of the print alphabet. It also includes symbols to represent punctuation, mathematics and scientific characters, music, computer notation, and foreign languages. Ever since the introduction of braille system, blind people were able to read and write without anyone else's support.

Braille is read by moving one hand or both hands left to right along each line and the index fingers do the reading. The average reading speed of a person who uses Braille is about 125 words per minute, but greater speed can be up to 200 words per minute. The average reading speed of most adults(sighted) is around 200 to 250 words per minute. Clearly, there is a significant difference between the two average reading speeds for both types.

These numbers show that the average reading speed of a person who uses Braille is much slower. Of course, it is reasonable. For one thing Braille requires the person to touch and comprehend the words instead of seeing. It is a time consuming and tedious task. For children who are still not good at reading Braille or adults who became blind later in their lives have hard time adjusting to this Braille system to do the reading and writing. As for any other language namely English, French, Braille is easier to learn the younger you are. Though Braille is not a language, it still requires some brainpower to learn and years to dedicate. Besides, you need to have high memorization skills if you want to learn Braille. Because, there are so many codes to remember as in Braille every letter, symbol is a group of raised dots. So, you need to have a high degree of memorization to master Braille.

A sighted adult who became blind later in their life will be reluctant to learn Braille because he/she is more familiar with seeing than touching. And of course, with the effort one must put for learning Braille, the newly blind person may hesitant to learn to use Braille. With the rapid development in technology, various kinds of systems have been implemented for blind and low vision people to use replacing Braille. Most of those systems are implemented to read text documents. They cannot be used to read tabular data, equations, charts or images. In the literature survey, we will describe those systems broadly.

This is important because there must be a way for a blind reader to read these specific types in any kind of document. Therefore, the aim of our research is to develop a system that helps a blind user to read not only text but charts, images, equations, and tabular data in any kind of printed or web document. The smart phone is most common item in day today life in anyone. Same as the vision difficulty community. Therefore, we decided to develop a mobile application. That names “READING EYE”.

## Background Literature

Assistive technology assists people with disabilities to achieve their ordinary life tasks and helps to improve their tasks in working, transportation, and majorly in academic activities at the same time, it guides them to accomplish better independence and makes the life more comfortable. According to the World Health Organization (WHO) official statistics, there are 285 million visually impaired people (partially or completely blind) in the world in which 39 million of them are blind and 87% of them are from developing countries [1]. At the same time, globally, it is estimated that approximately 1.3 billion people live with some form of vision impairment [2]. With the great advances in technology, there are several developed products for visually impaired people to improve their reading ability.

However, a major obstacle for visually impaired people is there not ability to improve readability and knowledge of learning, as they have limited resources to access many types of documents.

Basically, in this research we are providing solution for that problem, there having most of major research areas for me.

1. Detecting printed document and auto capturing the image.
2. Region detection and split upping.
3. Uniquely identify the type of document.
4. Describe the text-based content.

Due to the great advances in modern technology, there are many existing document reading applications for visually impaired people. Some of these are Amazon Kindle, BARD Mobile, Capti Voice, KNBF Reader, etc. [3]. Those accessible mobile apps are applications or specialized programs downloaded onto mobile devices for smartphones and tablets, that have accessibility built-in to compatible with screen reading or screen magnification software that is built into or installed on the device for focusing visually impaired people [3],[4],[5],[6],[7].

However, since the useful information obtained from the developed applications is not significantly higher than the white cane, the market acceptance is low and the answers from the tools are not very user-friendly.

Therefore, the latest research efforts are being turned to come up with new Audio Assistance for Vision Impaired Individual to Recognize Graphical Content on Print Disable Documents called Schmoozer [8] which should reduce the impediments of previously mentioned applications.

Text, graphical images, equations, and tables are the main regions in print disable documents considered in the existing applications. The identification of the graphical regions identification done by the authors collected images of mathematical equations, text, tables and graphical images and stored them in separate folders. Then apply HOG feature extraction and Support Vector Machine (SVM) algorithm on the collected set of images to convert them into a trained data set [8]. Some of the applications have done pre-processing and OCR classification and detect the texts [9].

International Journal of Scientific & Engineering Research provided a solution for Text Detection and Recognition with Speech Output in Mobile Application for Assistance to Visually Challenged Person for using a camera-based assistive text reading system to read text labels and product packaging from hand-held objects. Text detection is to detect regions in an image that contain text characters. Methods of feature descriptor can broadly be classified as using a Histogram of the oriented gradient (HOG) descriptor, Scale-invariant feature transform (SIFT), Speeded up robust features (SURF), Gradient location and oriented histogram (GLOH) algorithms [10].

## Research Gap

The research gap could demonstrate why should we need to implement the proposed system and this will indicate what are the various drawbacks and needed improvements of the previously developed systems as a comparison with proved system and the previously developed systems by referencing the gathered information in the literature survey. This could show basically, how much the proposed system deviates from previously conducted researches.

As discussed in the literature review previously it seems that many visually impaired people also can improve their reading ability and, they can exposure to new areas by gain knowledge through listening to audio document reading systems.

Next level, discussing about key comparison with “schmoozer” and “reading eye” in main research components;

1. Detecting printed document and auto capturing the image.

Table 1 : Research gap (Detection and Auto Capturing Model)

|  |  |  |
| --- | --- | --- |
| Feature | Schmoozer | Reading Eye |
| Edges detection | Using mathlab image processing.   * Low accuracy. * Slow processing. * High weight RAM usage. | Using OpenCV image processing.   * High accuracy. * Speed-up processing. * Low RAM usage. |

1. Region detection and split upping.

Table 2 : Research gap (Region identify and separate the image)

|  |  |  |
| --- | --- | --- |
| Feature | Schmoozer | Reading Eye |
| Identify the regions and split upping | Manually cropping images. | Automatically identify and separated. |

1. Uniquely identify the type of document.

Table 3 : Research gap (Type Identification)

|  |  |  |
| --- | --- | --- |
| Feature | Schmoozer | Reading Eye |
| Identify the type of images. | Using Image processing methods with ML.   * Low accuracy. * More time to identify. | Using Deep Learning methods and neural networks.   * Higher accuracy. * Very quickly identifying, more time saving. |

## Research Problem

Globally, Braille still is the primary reading method for blind people to access information and education independently. In this system, each character represented by a combination of one to six raised dots. A dot may be raised at any of the six positions to form 64 possible subsets. Without Braille codes, blind people would have never been able to read or write. Although it was and still is the best method, there are problems with the Braille system too.

For blind students, some subjects can be unattainable because textbooks and exams may not be readily available for those courses in a braille format. And for subjects like science, engineering, and mathematics, require advanced codes. These subjects are heavily contained with maps, charts, diagrams, figures and equations that have to be redesigned in order for the braille reading students to feel and understand a concept. It will be even harder for students who enrolled in biology classes. As there should be books which include images of human body, molecules, and cells, for students to refer. Because of the complexity of the contents in those subjects, it is hard to create the same book in braille format.

Not only the unavailability, but braille books are also usually more expensive than most college textbooks. It will cost up to $15,000 to convert just five chapters of a science book making it even harder to publish a science book in braille format because the conversion is actually difficult. Another issue with braille books is, they take up more physical space than normal printed books. A 1000-page math book could easily be 5000 pages in braille format. If a book is around this much larger, having thousands of pages, it will be difficult for a student to use the book.

As mentioned in the previous part, braille reading speed is relatively slower than the normal reading speed. Several kinds of research have been done regarding braille reading speed. And they concluded that braille is much slower. According to these facts and findings, it is obvious that there are some major issues with braille method. Since it was designed and introduced centuries ago, it may not be compatible with the needs and desires of a modern person.

We strongly believe these issues can be solved with the help of modern technologies. Therefore, our research is going to be conducted to provide a solution for the above-mentioned problems.

## Research Objectives

Implementation of a document reading system called “Reading-Eye” to facilitate visually impaired people to read printed documents that are not written using the braille system and help them to improve their reading capability as normally sighted people.

Using Image processing and machine learning techniques, Conventional Neural Networks of developed algorithms to identify the closest objects of the captured photograph. Those algorithms will be trained with pre-processed data sets to identify the captured photographs in the storage of the smartphone.

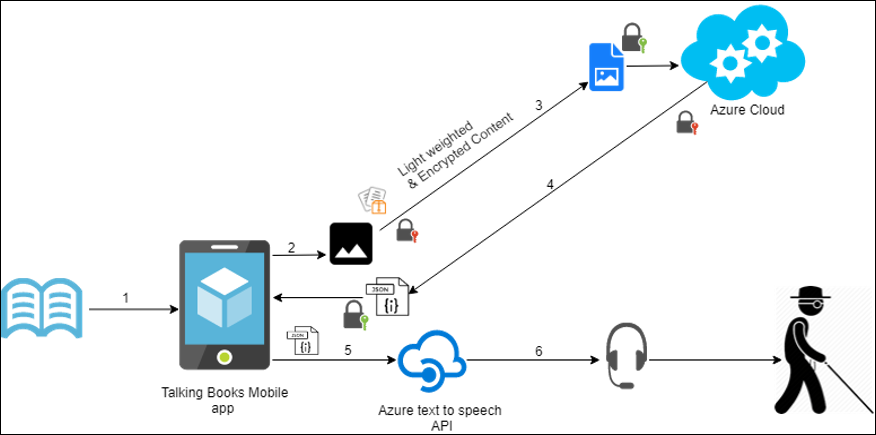
* Analyze captured photos and auto-rotate the image in the actual direction and identify where the captured image should be in the actual position.
* Make the single command to start analyzing and make the option after analyzing the document
* Start the auto play and make the user listen to the audio of the document.
* Identify text, images, charts, tables, and the equations uniquely and create another digital image for different identical matches.
* Identify language patterns and create a detailed description in the text.

# Methodology

## Methodology

Methodology section describes each functionality in the mobile application in more details. Particularly the process of detecting each graphical content in a document, analyzing & reading the contents, security aspects of the application and the technologies used in implementation.

According to the system overview of the proposed “Reading Eye” application, it has the capability to auto capture a printed document and upload it to a server automatically, and it provides better solution for detecting text, natural images, equations, charts and table contents in a document. As the outcome, this application generates the audio of the detailed description of the identified content for the user to listen. Furthermore, the application is equipped with a secure data connection.



Reading Eye App

Figure 1 : High Level Architecture Diagram overall

Basically, “Reading Eye” application having two parts.

1. Local side (mobile phone)
2. Cloud side (azure cloud storage back-end)

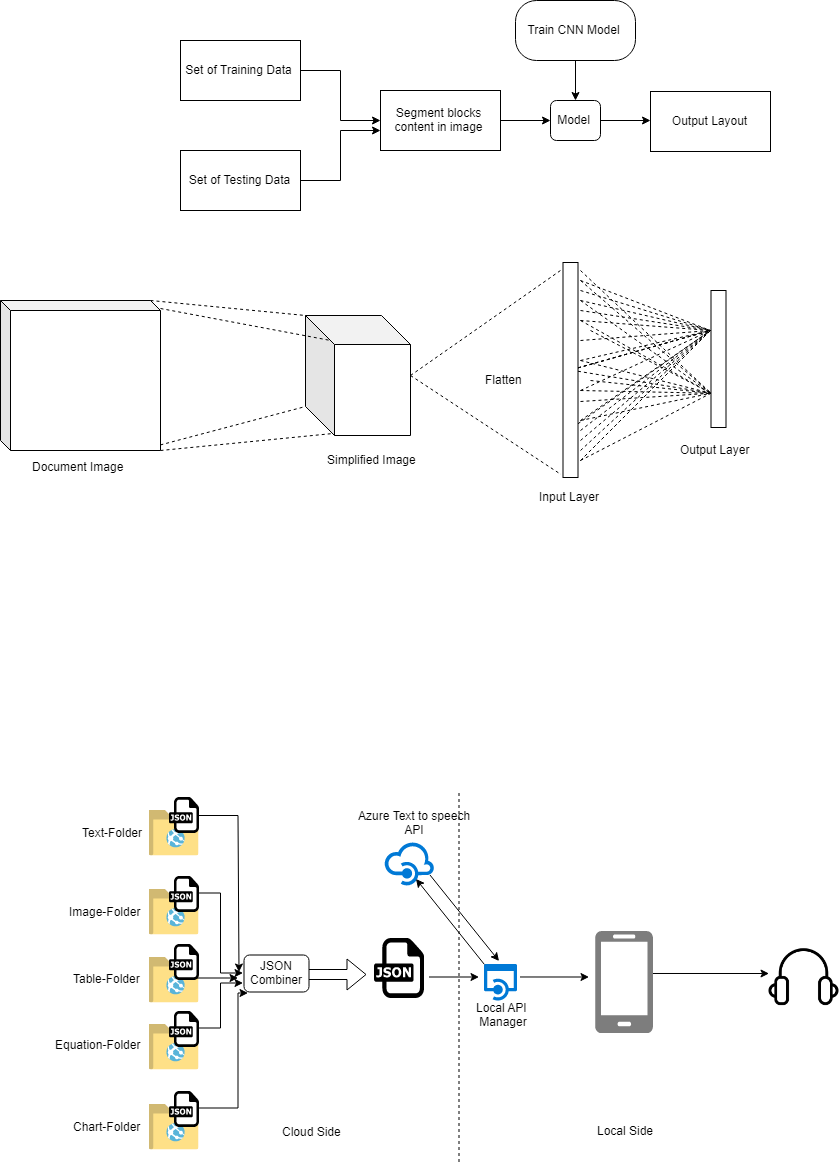
As the mobile side we can captured the image of a printed document. Blinded user can get more user experiences by voice assistance service. All of the identification processes, describing processes provide as the back-end service. Therefore, those modules running on the cloud. By the using the JSON Combiner it will be generated a detailed descriptive combined JSON file to the local API manger. By using the text to speech API a true the mobile application.

Figure 2 : High Level Architecture of Cloud Side and Local Side process

In our solution there are having significant tasks to accomplish. Detailed description of the mentioned bellow.

### Detecting printed document and auto capturing the image (DACI)

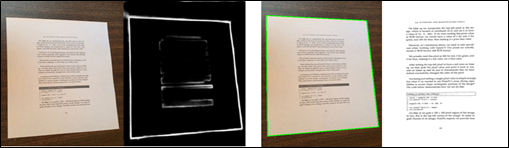
The developed mobile application (Reading-Eye) has the capability of auto capturing photos with the use of REA main interface. This module uses deep learning-based edge detection in OpenCV called as holistically nested edge detection (HED) for real time edge detection. HED make use of the side outputs of intermediate layers of images to generate final predictions. It is also capable of giving prediction of images with the help of convolutional neural network. The auto capturing images are stored in the phone mass storage.

Figure 3 : Sample output of the edge detection

### Region detection and split upping (RDASU)

The main task is image segmentation. The segmentation is a process to separate meaningful individual objects from a captured digital image. There are many machine learning techniques to achieve this process. There were founded most of them are very slow and give low accurate outputs. Therefore, we decided to use computer vision, based image processing techniques with machine learning methods.

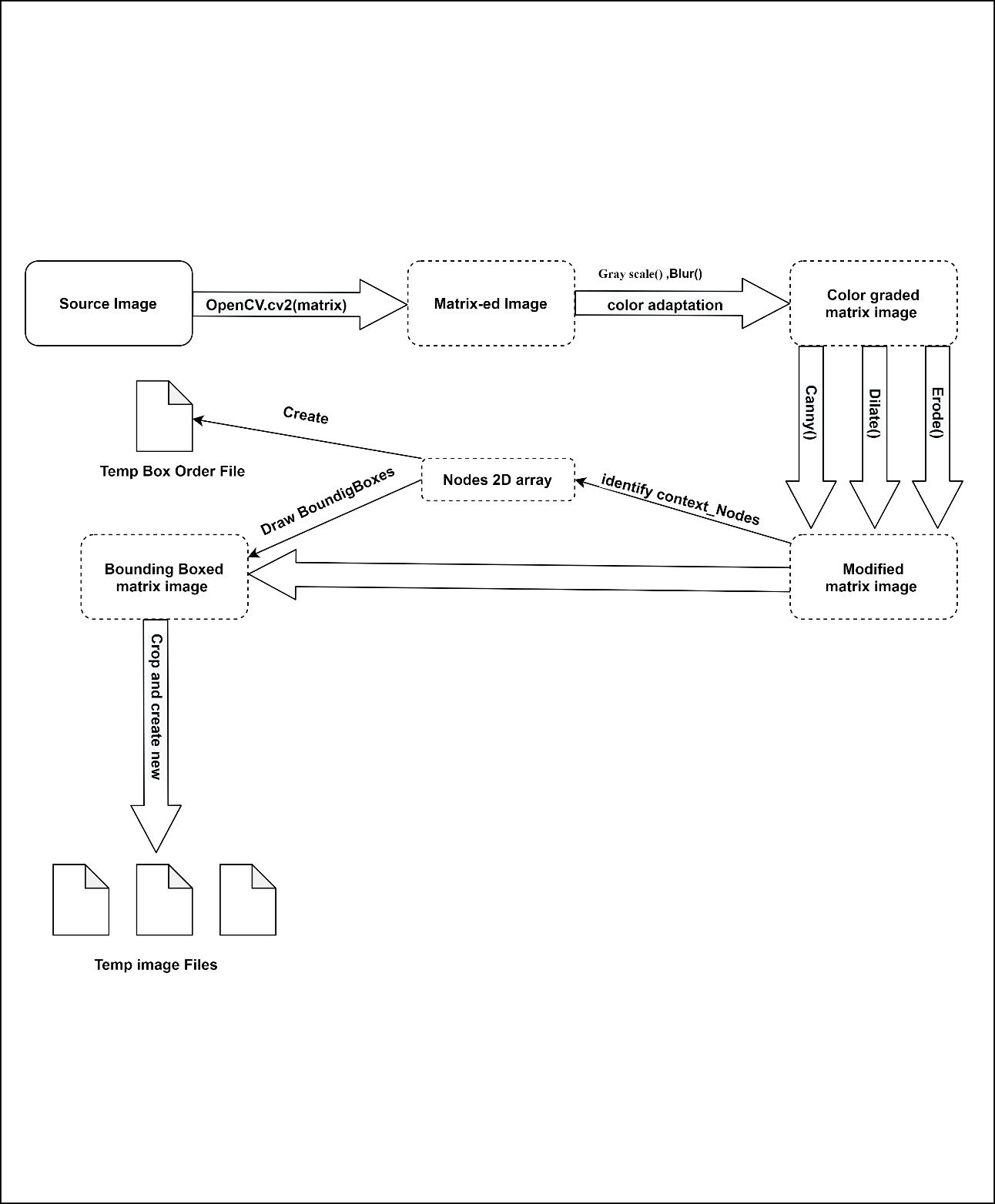
Firstly, load the image as OpenCV cv2 matrix. After that color adaptation to grayscale and Gaussian Blur, Canny, Dilate, Erode functioning. After identifying the context\_Nodes. After that, stored nodes pass the 2D array and draw Bounding Boxes. After crop by bounding box outline and renaming save temp jpg file. Also, parallelly maintain an order of the bounding box node index and save to temp file.

Figure 4 : Region detection and split upping

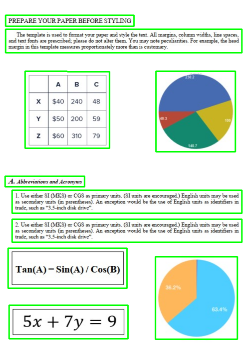
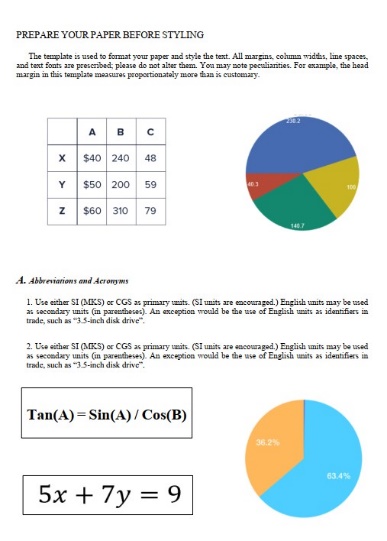
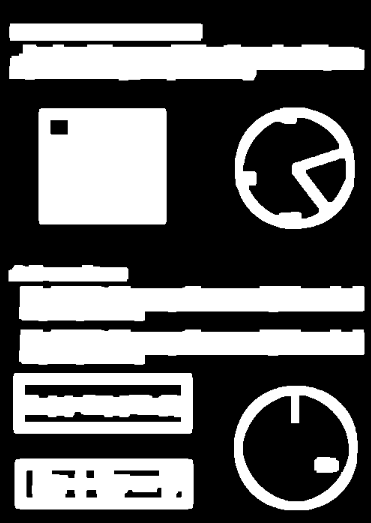


Figure 5 : Region detection output

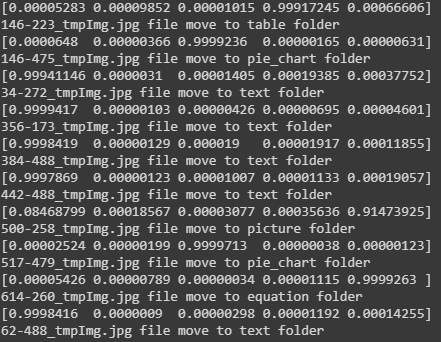
### Uniquely identify the type of document (UITD)

The main task of that section is identifying type of images. There having been many previous machine learning techniques for image classification. The problem was, that having no data sets fully filed and also vary accuracy levels. Therefore, selected deep learning based convolutional neural networks (CNN) classification method.

First stage, we collected the data and created a dataset. It includes images of paragraphs, basic mathematical equations, bar charts, pie charts, tables and various natural images. Then we divided them for training and validation. Then, we started labeling and annotating all images. We used an online labeling tool called VGG Image Annotator.

Then, we trained a new model using pre-trained COCO weights [11] and after that applied ImageNet model and, PubLayNet model. Later we fine-tuned models. After that created keras module converted mobile net to reduce the model file size.

Finally, load the model and pass the previous step created temp image files for the prediction. In order to prediction file moves to correct repo.

Figure 6 : Uniquely identify the type of document output

### Describe the text-based content

The main purpose of the text identification function is to identify and read the textual contents in printed documents. This was done using Tesseract OCR engine, which is the best OCR engine available. As the final output of the function, a JSON file will be generated using the read content. Then the content will be concatenated with output files from other graphical contents to make a final description of the original document image. The final description will be converted to an audio file for the user to listen by a Text-to-Speech API.

## Commercialization aspects

### Cloud based revenue model

This revenue model is used if the client agrees to install their application setup in a amazon cloud environment. Amazon will charge for monthly resource usage. Clients are charged with an additional service charge which will be a multiple of the entire usage bill from cloud service.

Total usage bill = Amazon cloud bill x billing factor

### On premises revenue model

If the client does not want to use a 3rd party service to host their application and wants to setup the system on premises, this revenue model will be used. Here client will be billed monthly based on the user accounts created in the system.

## Testing and Implementation

# References

[1] Who.int. (2018). GLOBAL DATA ON VISUAL IMPAIRMENTS 2010. [online] Available at: https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf [Accessed: 06-Aug-2019].

[2] World Health Organization, 15-Jan-2019. [Online]. Available: https://www.who.int/blindness/en/. [Accessed: 06-Aug-2019].

[3] "Accessible Mobile Apps | American Foundation for the Blind", Afb.org, 2019. [Online]. Available: https://www.afb.org/blindness-and-low-vision/using-technology/assistive-technology-products/mobile-apps. [Accessed: 20- Aug- 2019].

[4] Play.google.com. [Online]. Available: https://play.google.com/store/apps/details?id =clarosoftware.com.magx&hl=en. [Accessed: 22- Aug- 2019].

[5] "An In-depth Evaluation of the BARD Mobile App from the National Library Service for the Blind and Physically Handicapped | Access World | American Foundation for the Blind", Afb.org. [Online]. Available: https://www.afb.org/aw/15/2/15607. [Accessed: 20- Aug- 2019].

[6] Play.google.com. [Online]. Available: https://play.google.com/store. [Accessed: 21- Aug- 2019].

[7] "Read text aloud. PDF reading books speak text to voice, read it out loud", Captivoice.com.15 [Online]. Available: https://www.captivoice.com/capti-site/public/entry/explore. [Accessed: 22- Aug- 2019].

[8] N. Gamage, K. Jayadewa, S. Senanayake, K. Udeshitha and J. Jayakody, "Audio Assistance for Vision Impaired Individual To Recognize Graphical Content on Print Disable Documents", Annual conference 2017 - IET - Sri Lanka, 2017.

[9]: “How WhatsApp Uses End to End Encryption?”: Ashish Bijawat [online] Available at: https://www.linkedin.com/pulse/how-whatsapp-uses-end-encryption-ashish-bijawat/ [Accessed: 22- Aug- 2019].

[10]: “15 FREE MOBILE ENCRYPTION APPS TO PROTECT YOUR DIGITAL PRIVACY”: Nate Lord [online] Available at: https://digitalguardian.com/blog/15-free-mobile-encryption-apps-protect-your-digital-privacy [Accessed: 22- Aug- 2019].

[11] Cocodataset.org. (2020). COCO - Common Objects in Context. [online] Available at: http://cocodataset.org/ [Accessed 1 Mar. 2020].