DOCUMENT READING SYSTEM FOR BLIND PEOPLE ("READING EYE")

19-20-J 17

Project Proposal Report

E.M.D.D. Ekanayake

I.N. Kalansooriya

P.S.N. Kularathne

W.R.P. Fernando

B.Sc. Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2019

DOCUMENT READING SYSTEM FOR BLIND PEOPLE ("READING EYE")

19-20-J 17

Project Proposal Report

E.M.D.D. Ekanayake

I.N. Kalansooriya

P.S.N. Kularathne

W.R.P. Fernando

Ms. Suranjini Silva

B.Sc. Special (Honors) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

August 2019

DECLARATION OF THE CANDIDATE & SUPERVISOR

We declare that this is our own work and this proposal 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 our 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.

Name	Student ID	Signature	Date
IT16102156	E.M.D.D. Ekanayake.		
IT16176348	I.N. Kalansooriya.		
IT16165762	P.S.N. Kularathne.		
IT16079328	W.R.P. Fernando.		

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

……... ……

Signature of the supervisor Date
(Ms. Suranjini Silva)

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.

Keywords

Virtual Reader, Assistive Technology, Computer Vision, Deep Learning, Natural Language Processing, Artificial Intelligence

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

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 has to 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.

1.2. Literature Survey

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].

According to the major the development of the technology, there are some developed products to avoid the difficulties of the reding documents which help visually impaired people to improve their reading ability [3],[12].

However, focusing on the improvement of reading ability and gain knowledge for academic purposes for the visually impaired people there is a major impediment because they have limited resources for access to most of the specific types of documents.

1.2.1 Reading and detecting text-based contents and Region Identification

Due to the major development of modern technology, there are many existing document reading apps are available for visually impaired people.

Some of these are Amazon Kindle, BARD Mobile, Capti Voice, KNBF Reader, etc. [15]. 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 [19] that is built into or installed on the device for focusing visually impaired people [15],[16],[17],[18].

However, the market acceptance is less as useful details gathered from the developed apps are not remarkably more than that from the white cane and replies received from the tools are not many users 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 [11] 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 [11]. Some of the applications have done pre-processing and OCR classification and detect the texts [34].

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 [35].

1.2.2 Reading and detecting graphical Images-based contents and Table-based contents

As mentioned in the literature survey previously, some applications can identify and detect graphical images-based contents as below.

Schmoozer used to identify the graphical image-based contents separately and creates a text file with the meaning of the image. This process achieved by comparing the image with a trained database inside the server. The dataset in the database trained using the SVM algorithm. HOG feature extraction technique used to identify unique features in the image and convert into a decimal value to store in a database [11]. According to that the main purpose of the "Graphical image identification function" is separately recognizes and labels the images available in print disable document. Authors implemented a strong dataset, which contains a huge number of images of living beings and many single objects with significant behavioral changes. Then applied HOG feature extraction to identify unique features in the graphical image and convert into a decimal value to store in a database. The database was trained using SVM algorithm to convert the dataset into the trained dataset [11].

Moreover, Azati software indicates that the use of neural networks with deep learning is one of best the option to detect, analyze and read the graphical images because of the development of new technologies in the world [10]. Therefore, that is one of the optimal methods that we plan to implement our solution in this research.

Table based content identification function in Schmoozer is to recognize the type of data table (whether it is 2columns or 3columns) and convert table data into meaningful digitized text. It implemented with a strong dataset with a huge number of 2 column and 3 column tables are the foremost step in the table identification process. Then applied HOG feature extraction to identify unique features of 2 column and 3 column tables and convert into a decimal value to store in a database the database was trained using SVM algorithm method to convert the dataset into the trained dataset. When the function" obtain a cropped and identified table segment from graphical region identification function it extracts HOG features and converts into a decimal value. The decimal value was compared in with the trained data set to find the type of the table

whether the table has 2 or 3 columns. The data inside the image of the table was read using in build the OCR function of MATLAB and wrote the data into a text file [11].

1.2.3 Reading and detecting charts based-contents and Equations based contents.

The Most software applications developed to detect and analyze text contents but most of them cannot detect chart-based content and equations-based content [15],[16],[17],[18].

Detecting equation regions from scanned books has received attention in the document image research community in the past few years. Compared with regular text blocks, equation regions have more complicated layouts so we cannot simply use text lines to model them. On the other hand, these regions consist of text symbols that can be reflowed, so that the OCR engines should parse them instead of rasterizing those like image regions.

A Simple Equation Region Detector for Printed Document Images in Tesseract is one of the research that has been conducted for one of the equation detector, which use to detect equations using Tesseract OCR engine. The equation detector is built into the layout analysis stage. Specifically, it works after the text line detection and before the paragraph/block partition, so that the detected regions won't be mistakenly grouped into a text paragraph. In addition, we skip these regions during the recognition stage, because no special parser has been built yet in the Tesseract engine [14]. This will basically classify the text symbols then identify the text regions and expand the text regions to read and detect the equations-based contents.

Another research called Schmoozer mainly focuses to detect mathematical equations based on the separately recognize characters and symbols in a mathematical equation that are printed on documents and convert them into digitized text. The first step was designing of mathematical symbols and numbers using Photoshop in bitmap format. Bitmap files are easy to create and pixel data stored in a bitmap file could be accomplished by using a set of coordinates that allow the data to be conceptualized as a grid. Therefore, all the major mathematical characters were created in bitmap format and stored inside a folder called "letters_numbers". Then created a MATLAB template and match each character /mathematical symbol/number into created bitmap images [11].

Moreover, according to the Schmoozer use to identify the equations by obtained input images from the graphical regions identification function Read image, Convert to grayscale and converted into a binary value and Finally match every character of the input image with the binary value obtained from the created template and regenerate the equation as digitized text. Function's output is a set of single digitized values for a single character/symbol in the equation and converts to DDL file for use to web service [11].

When considering the literature survey for chart-based content detection, we were unable to find comparatively numerous researches since this feature compromise of new methodologies and techniques.

One of the major challenges in chart image classification is the variability of the structure and visual appearance of each chart type found from and conducted research called Convolutional Neural Network Based Chart Image Class [33].

However, the different approaches have been proposed in the literature to deal with the problem of chart image classification, the variation in chart style, size, color, resolution, and the content are not yet resolved.

1.2.4 Enhance the Security and Data Encryption

There are many of reading applications for visually impaired people in the app stores. but none of them has any encryption algorithm to secure their communications with servers. Therefore, we focused on another famous mobile applications that use encryption algorithm to secure the back-end communication. WhatsApp is a very famous communication app. It uses a protocol named 'signal' to encrypt data. This protocol uses asymmetric and symmetric key cryptographic algorithms [34]. Only one shared key use for encryption and decryption processes in symmetric key cryptographic algorithms.

However, in asymmetric key cryptographic algorithms, here are two keys to encrypt and decrypt named as public and private. When data encrypt with public key the encrypted data must be decrypted by using private key. WhatsApp uses an algorithm based on Curve25519[34].

LastPass is an encrypted password manager. It can be used IOS, Android, windows and as an integrated browser extension for desktop [35]. Its use as a secure passwords storage. All passwords will be encrypted with device-level AES-256-bit encryption

with PBKDF2 SHA-256[36].and it also uses salted hashes to store passwords securely. AES algorithm with 256-bit long key is not very efficiency. But however, there is no back-end communication. Therefore, efficiency is not necessarily important fact. Viber is another famous communication app. It encrypts the message before sending it to anyone.it uses an ephemeral one-time 128-bit symmetric key for encryption [36] Salsa20 encryption algorithm being used for encryption [36].

Signal is another famous app for the message encryption. It uses signal protocol for the encryption. The protocol based on the combination of the double Ratchet Algorithm, prekeys, and a Triple Diffie-Hellman (3XDH) handshake [37]. It uses HMAC-SHA256, Curve25519 and AES-256 as primitives [37].

1.3. 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.

However, after completion of the background and the literature survey, there were certain identified limitations that laid the foundation for our research. The proposed solution is based on to reduce those limitations and perform an improved better solution for visually impaired people and at the same allow the visually impaired people to reach the same level of reading level as a normally sighted person without a barrier by providing the descriptive idea and best understanding of texts, digital images, charts, tables and mathematical equations.

Most of the solutions the ones had been carried out in advance had a similarly faced when compared to the proposed system solution, furthermore developed features and integrated functionalities were minimal when compared with the proposed solution.

When reviewing the research gap, the technological advancements that have taken place over the restrictions and the limitations in the technologies at past times of development should not be disregarded, as the client-server data security and cloud storage advancements present as per today, did not exist back then.

The use of advanced machine learning techniques and image processing libraries, conventional neural networks, Deep Learning Frameworks and techniques, cloud hosting with data security are the sum of technologies that improve the usability and the widening factors which are the most trending technologies in the present that our proposed system wish to implement. With the use of new technologies, the result of the proposed system will become an accurate, more efficient, effective and optimal solution for the targeted assistive technology area for visually impaired people.

Below is the comparison of the proposed system features and the previous systems features identified by the literature review.

Feature	Amazon Kindle	BARD Mobile	Capti Voice	KNFB Reader	Schmoozer	**Reading Eye (Proposed System)**
Text Identification and Reading	✓	✓	✓	✓	✓	✓
Graphical Image /Pictures Identification and Reading	X	X	X	X	✓	✓
Equations Identification and Reading	X	X	X	X	✓	✓
Table content Identification and Reading	X	X	X	X	✓	✓
Charts Identification and Reading	X	X	X	X	X	✓
Voice output	✓	X	✓	✓	✓	✓
Voice input (Voice as command inputs)	X	X	X	X	X	✓
Cloud Storage	X	X	X	X	✓	✓
Client-server Communication data security	X	X	X	X	X	✓
Cross Platform Mobile Application (Android/ IOS)	✓	✓	X	✓	X	✓
Images' size reduction for efficient Communication	X	X	X	X	X	✓

Table 1.2.1: Comparison

1.4. 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.

2. OBJECTIVES

2.1. Main Objective

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.

2.2. Specific Objectives

The application of key research areas in the project and application of techniques in the relevant key areas are discussed as follows.

2.2.1. The key objective aimed (1)

- Reading and detecting text-based contents and Region Identification and generate audio output.

2.2.1.1. The research areas/pillars used as optimal

- 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.

2.2.1.2. Method of achieving the stated specific objective

- The first task is region identification. When the camera takes a photo, we cannot guarantee the picture will only contain text, image, equation, chart or table, as it will only contain text. It can contain images or charts as well. Maybe the picture will contain a combination of text, images, and charts. First of all, before start analyzing we need to identify the contents in the captured photo. We need to identify text areas because majority of the document will anyway be text. After that other contents should be identified and separated. We intend to do this separation by cropping the photo.

- The image segmentation task is do by semantic or instance partitioning to partitioning an image into multiple segments.

- The second task is the reading text-based content. The module will be implemented to read text in a document image. Text content should extract using the Mask R-CNN. After all, read word by word of that extracted chunk using Optical Character Recognition techniques.

- The third task is the final audio output generation. Train out the meaningful descriptive finalized and concatenated JSON files and sent those to the smartphone as well - structured detailed wise, meaningfully described JSON files.

2.2.2. The key objective aimed (2)

- Enhance the security of the collected data, application, and client-server communication and reduce the size of images for highly efficient communication.

2.2.2.1. The research areas/pillars used as optimal

- Implement an encryption algorithm unique to the application for securing the communication between the cloud and the application.

- Implement an algorithm (Hybrid algorithm using DES and AES for better efficiency and security) to reduce the size of the images before transferring to the cloud for efficient communication between the cloud and the application.

- Enforce cloud policies to safeguard and handle the process according to the standards and best practices.

2.2.2.2. Method of achieving the stated specific objective

- Use the Diffie-Hellman algorithm for the secure key exchange to make key exchange much more secure and reliable.

- Use randomly generated 256-bit long secret key and embed it into the application to make key storage much more secure.

- Create an algorithm to identifying unwanted area of an image and removing it.

2.2.3. The key objective aimed (3)

- Reading and detecting graphical Images-based contents and Table-based contents.

2.2.3.1. The research areas/pillars used as optimal

- Training the pre-processed data sets by using machine learning algorithms, neural network techniques, Optical context of character recognition (OCR) to identify and separate the graphical images and the table-based content.

.

2.2.3.2. Method of achieving the stated specific objective

- To identify individual objects and their characteristics in the image, we will train a Mask R-CNN with relevant input datasets. The first input of this neural network will the cropped image as usual. Then other datasets such as animals, vehicles, clothes will be used to train the CNN. ImageNet will be referred to get these datasets when training and testing the models.

- The identification of the tables is done by using Optical Character Recognition techniques to extract that content. A neural network will be implemented to detect characters and numbers in the table region accurately. When reading tables, it is important to identify table columns and rows.

- Finally make a JSON file for made meaningful descriptive detected images such as texts, graphical images, tables and combine the JSON files as a meaningful finalized JSON file as including the data as the captured photograph in the initial stage by the visually impaired person.

2.2.4. The key objective aimed (4)

- Reading and detecting charts based-contents and Equations based contents.

2.2.4.1. The research areas/pillars used as optimal

- Training the pre-processed data sets by using machine learning algorithms, neural network techniques, Optical context of character recognition (OCR) to identify and separate the charts based-contents and Equations content.

2.2.4.2. Method of achieving the stated specific objective

- To detect the charts and equations should need to break chart based and equations-based contents module into two submodules as Charts and Equations separately.

- After the break down can separate the cropped image of charts and equations using Mask R-CNN as mentioned in the region identification section.

- Due to the variation in chart style, size, color, resolution and content apply the convolutional neural networks to classify chart images to identify several common charts types like pie, bar and histograms.

- Neural network model will be trained to analyze common charts and to the following identification use CNN based model inspired from the LeNet architecture.

- At the end of the model, the fully connected layer with a SoftMax activation function makes use of the learned high-level features to classify the input images into predefined classes like pie or bar chart.

- Use OCR techniques to detect the texts inside the charts.

- Equation detection and identification done as follows,
    - OCR will be used to recognize characters, symbols and numbers in the image in a mathematical equation.
    - NLP can be used to understand the meaning of these symbols and then generate meaningful sentences.
    - That equation will be further developed to read quite complex equations later.
    - When implementing above tasks Keras and python library will be used as usual.
    - Finally make a JSON file for made meaningful descriptive detected images such as texts, graphical images, tables and combine the JSON files as a meaningful finalized JSON file as including the data as the captured photograph in the initial stage by the visually impaired person.

3. METHODOLOGY

This is the illustration of the methodology for the proposed “Reading-Eye” system and this is the methodical process for the requirements gathering, project planning, system designing and implementation to generate an effective solution for visually impaired people to read any type of document and understand the content without hassle. The proposed project has very significant research areas such as Computer Vision, Deep Learning, Natural Language Processing (NLP), Artificial Intelligence, Cloud Computing and Mobile platform development.

3.1. System Overview

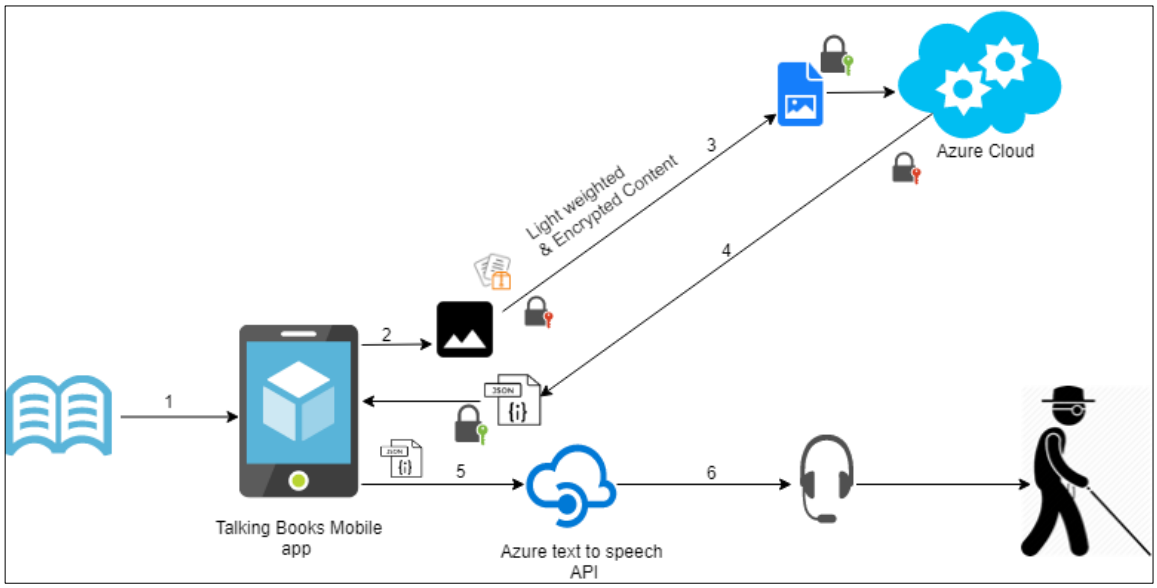

Figure 3.1.1: High Level Architecture Diagram

System overview brings up the most appropriate tools, technologies and software solutions for the implementation phase and flow of the proposed solution. We proposed to develop a cross-platform mobile application that is capable of understanding the contents of various documents (ex: newspapers, reference books, research papers) and reading out aloud for a blind user to understand its content. The application can read text, images, tables, equations and charts.

3.2. Proposed Technologies

- Machine Learning Techniques – Keras, TensorFlow will be used to create DL models.
- Cloud Hosting Service – MS Azure will be used for hosting and backend processing.
- Text-to-Speech API – MS Azure service
- Mobile App – React Native will be used for implementation of the cross-platform mobile application.

As a solution for the above-mentioned research problem, we are proposing to develop a cross-platform mobile application. The basic functionality of the application is reading text, images, charts, tables and equations of any kind of document. Moreover, security of the application will be enhanced as well.

When using the application, a user can point the camera of his device to the printed document which he wants to read. Then the camera will capture the photo and the app will set it to correct orientation. If the photo capturing went wrong, the user will be notified by the application (via an audio message). After that the captured image file will be compressed and encrypted using a specific algorithm to ensure the security of the file before uploading it to the cloud server. We are planning to host the application in a cloud server (MS Azure) to do the back-end processing because a mobile phone or tablet do not have such processing capabilities. Thus, in the cloud server all the processing will be handled and only the generated audio file will be sent to the mobile application. Then the mobile application will play the audio file for the user to listen.

Mainly focus in this application to read contents of various document types and books which contains texts, graphical images, charts, mathematical equations and tables. The organizing of the text content varies in each document type. Although we often see paragraphs of text in books and booklets, bills are not like that. There organization can be totally different as they contain line by lines instead of paragraphs.

In the proposed solution there are six significant tasks to accomplish.

1. Region Identification, Reading of text-based content & audio generation
2. Reading and detecting images-based contents and tables-based contents.
3. Reading and detecting charts-based contents and equations-based contents.
4. Enhance the security and Encryption.

Detailed description of the mentioned tasks and how we are going to achieve them are discussed in the below section.

3.2.1. Region Identification, Reading of text-based content & Audio generation

The first task is region identification. When the camera takes a photo, we cannot guarantee the picture will only contain one thing, like it will only contain text. It can contain images or charts as well. Maybe the picture will contain a combination of text, images and charts. First of all, before start analyzing we need to identify the contents in the captured photo. Identifying text areas because majority of the document will anyway be text. After that other contents should be identified and separated. We intend to do this separation by cropping the photo. For example, if a photo contains two paragraphs of text and one chart, the chart area will be cropped and saved and the same will happen to text as well. This is what will happen in the region identification.

After studying of ways to accomplish the mentioned task when implementing the system, we found that deep learning is the most suitable technology that there is. Deep learning techniques provide more accuracy when detecting objects and areas and currently considered the best solution for computer vision problems in the industry.

Image segmentation is the task of partitioning an image into multiple segments. This is quite similar to grouping pixels together on the basis of specific characteristic(s). There are two types of image segmentation.

- Semantic Segmentation
- Instance Segmentation

Semantic segmentation describes the process of associating each pixel of an image with a class label, (such as flower, person, road, sky, ocean, or car). We can use this semantic segmentation method to identify the area of text, image, chart, equation and table. Many convolutional neural network models have been introduced for semantic segmentation of an image like Fast-RCNNs, Faster-RCNNs. In here, for region identification we plan to use Mask R-CNN. It is an extension of Faster-RCNN. Mask R-CNN is efficient when detecting objects, classifying objects and segmentation which is exactly what we are trying to achieve. [Figure 3.2.1.1] [Figure 3.2.2.1]

Implementation of the Mask R-CNN which receives the photo captured by mobile camera as the input and the neural network will then segment the photo into correct regions as expected. (Since to be processed by the neural network, the image must be converted to a binary array. Next, the binary arrays will be used to develop datasets for both training and testing the networks. This will happen in other features throughout the project).

Then the segmented areas will be cropped from the original image into several copies. Like text area as one image and chart as another image and if there is another type, it should be taken as another image. This way we can get the inputs for the analyzing feature. When implementing the neural networks, we use Keras and TensorFlow. For cropping the images, we use OpenCv library.

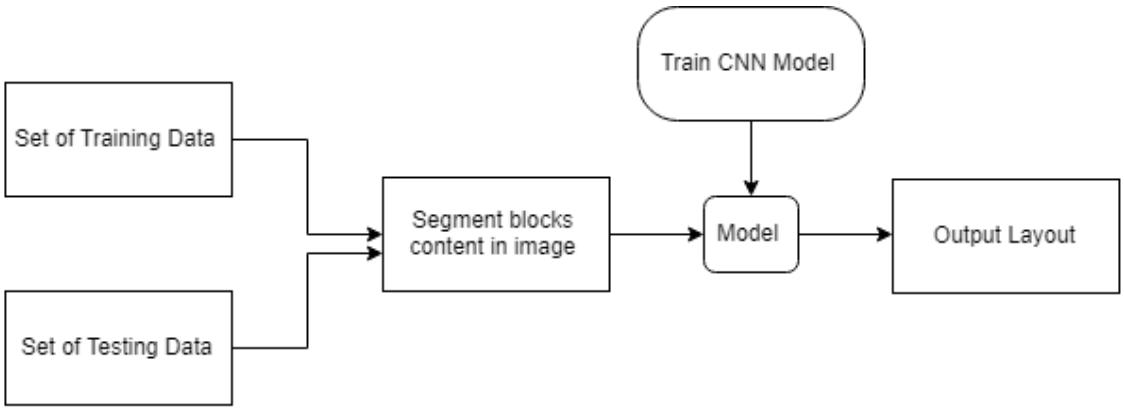

Figure 3.2.1.1: Segmentation Methhodology

Second task is the reading text-based contend. The module will be implemented to read text in a document image. As we propose in the previous section, we extract the text content using a Mask R-CNN. Now we have to read word by word of that extracted chunk. As this task, is to use Optical Character Recognition techniques. It can be achieved implementing another type of neural network which consists of input layer, one hidden and output layer. The image of the text area will be the input of the neural network. Neural network will detect characters in that image accurately and gives the output as a sequence of characters. Then this output will be used to generate the audio file.

Third task is the final audio output generation. Train out the meaningful descriptive finalized and concatenated JSON files and sent those to the smartphone as well - structured detailed wise, meaningfully described JSON files. [Figure 3.2.1.2]

Accessing Azure text to speech API for real-time reading of generated JSON file.

Using the headset visually impaired person will be getting the access to listen detailed well-described single audio of the document generates by the Azure text to speech API.

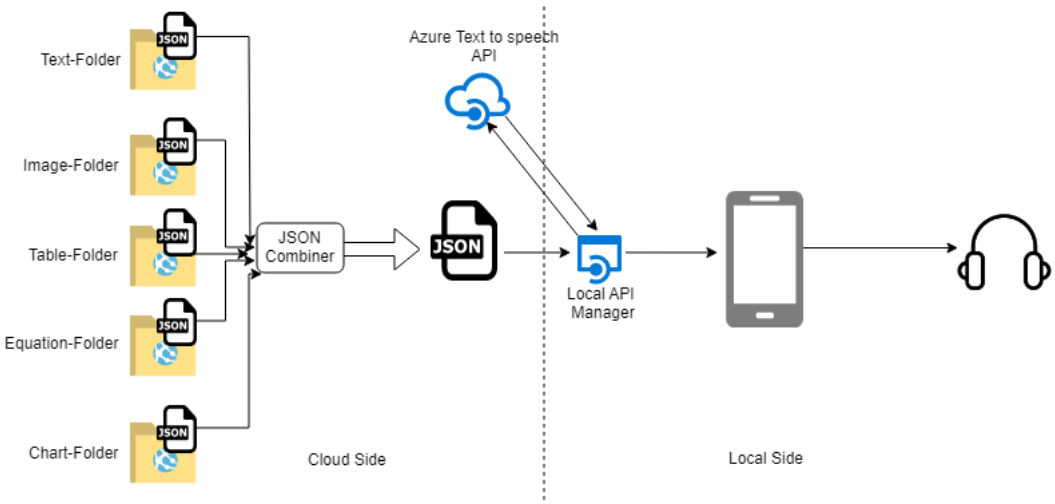

Figure 3.2.1.2: Audio Genaration

3.2.2. Reading and detecting graphical Images-based contents and Table-based contents

The motivation behind this module is to read images and to read tabular data in each document. Reading images should be defined precisely here. When reading an image, we are trying to detect main objects in the image as the first step. For example, if the image is of a white dog playing with a ball in a ground. The main objects can be identified as a dog and a ball. Other than that, there is the background. So, the outcome of this reading should be something like this sentence: “White dog is playing with a ball in green field”.

Using semantic segmentation, we can identify the background as green field. To identify individual objects and their characteristics in the image, we will train a Mask R-CNN with relevant input datasets. The first input of this neural network will the cropped image as usual. Then other datasets such as animals, vehicles, clothes will be

used to train the CNN. ImageNet will be referred to get these datasets when training and testing the models. [Figure 3.2.2.1]

Tables are important when presenting data in a meaningful way. In this module, a separate part will be implemented to read tabular data in a document. This is particularly useful when a user is reading a document like an electricity bill or a telephone bill. Because this type of document has tables with data in it. Since tables have characters and numbers, we need to use Optical Character Recognition techniques to extract that content. A neural network will be implemented to detect characters and numbers in the table region accurately. When reading tables, it is important to identify table columns and rows. The lines between one column and another should be clearly recognized because when reading tabular data, we need to know what the columns are with their headers. Just like that, knowing the margins of the rows is important. The method proposed can be explained using an example of a simple data table.

The table contains marks of three students. Usually in documents tables have a title. Usually in documents, tables have a title. In such cases, the table title will be read to the user to get an idea about the data contained in the table. [Table 3.2.2.1]

Student Name	Marks
Anne	70
Marc	80
John	85

Table 3.2.2.1: Example data table

The outcome of reading this table will be like “Anne’s mark is 70”. First column’s row data will be read along with the second column’s row data. Therefore, identifying column margins and row margins is important in this process. A convolutional neural network will be implemented to identify the columns, rows and data in it.

In both features after detecting objects and texts the system needs to create a detailed description of the analyzed data. This is called text generation because machine itself has to learn the features in the image and form cohesive sentences for the user to understand.

Natural Language Processing, or NLP, is the sub-field of AI that is focused on enabling computers to understand and process human languages. NLP helps to accomplish many exciting tasks. Text generation is one of such tasks which can be achieved using deep learning models, particularly LSTM Recurrent Neural Networks. It's a special type of RNN. We create a language model for generating natural language text by implement and training state-of-the-art Recurrent Neural Network. The specialty of Recurrent Neural Network is its activation outputs are propagated towards both directions. This creates loops in the neural network architecture which acts as a 'memory state' of the neurons. This state allows the neurons an ability to remember what have been learned so far which will be helpful when creating meaningful descriptions for both tasks. For this module also will be implemented using Keras neural network library.

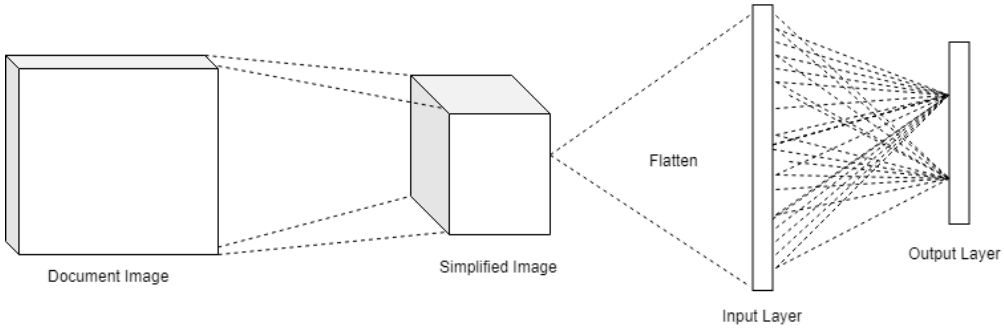

Figure 3.2.2.1: CNN Model

3.2.3. Reading and detecting graphical Charts -based contents and Equations-based contents.

The motivation behind this module is to automatically classify the chart image types and equations from an image dataset. Therefore, we break this module into two submodules as Charts and Equations separately. In this module also, we get a cropped image of charts and equations using Mask R-CNN as mentioned in the region identification section.

Chart image classification is an important step in chart recognition and understanding. Different approaches have been introduced to deal with the problem of chart image classification. As it is quite difficult when classifying charts due to the variation in chart style, size, color, resolution and content are not yet resolved. In our research, we propose applying of convolutional neural networks to classify chart images to several common chart types like pie, bar and histograms. Firstly, a neural network model will be trained to analyze common charts, after that it will be expanded to identify more

types later. We design a CNN based model [Figure 3.2.2.1] inspired from the LeNet architecture. It has a total of eight layers including one initial input layer, five hidden layers and fully connected layer and ending with the output classifier layer. The five hidden layers are convolutional and pooling layers which do the feature extraction of the input images while the fully connected layer does the classifying.

At the end of the model, the fully connected layer with a SoftMax activation function makes use of the learned high-level features to classify the input images into

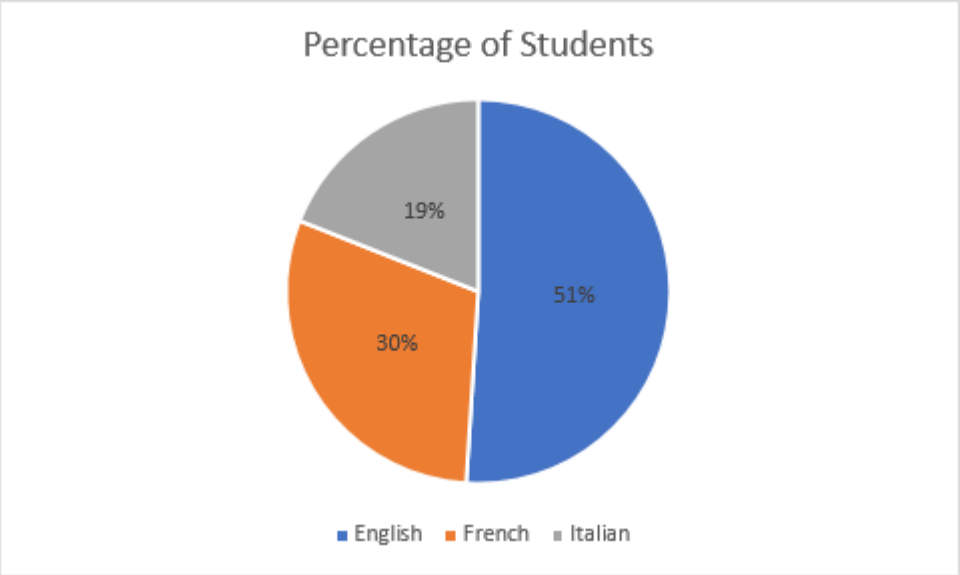

Figure 3.2.3.1: Example Pie Chart

predefined classes like pie or bar chart. This model is composed of two main parts. First part is the feature extraction model. And the second part is the classification model. In feature extraction model, neural network learns to detect different high-level features from the input images. It consists of a sequence of convolution and pooling layers. The goal of convolution is to extract features from the input image. Using all the learned features, network can decide the type of the chart. This happens in the classification model. After detecting the type of the chart, then we have to read the content of that chart. Again, we have to perform a similar task. That is detecting text inside the chart. For this also we can use same OCR techniques using model as mentioned in previous section. The chart contains characters and regions of its own. Below mentioned is a sample pie chart we will be analyzing in the application.

When reading the pie chart, the system will give an output like this:

“Chart title is Percentage of students. English – 51%, French – 30% and Italian – 19%”.

If existence of characters were recognized above the chart, that will be taken as the chart title. If there were words inside and below the chart, that will be taken as the data input. These characters (e.g.: English and %) and numbers will be read using optical character recognition. Then the read data will be the input for generating text. NLP will be used for generating cohesive sentences in a way user can comprehend.

Another feature of this module will be reading equations. Below contains an example of equation that will be read using our application.

$$\mathbf{Y = a * (x^2 + z^2) + c}$$

The above equation will be read like,

“Y equals “a” into (*) x squared plus (+) z squared plus (+) c”. Terms like into and plus will be generated using the symbols in the equation. Same as before OCR will be used to recognize characters, symbols and numbers in the image. NLP can be used to understand the meaning of these symbols and then generate meaningful sentences. First, we plan on implementing the system to read simple equations as above. It will be further developed to read quite complex equations later. When implementing above tasks Keras and python library will be used as usual.

3.2.4. Enhance the Security and data Encryption

Cryptography is an essential part of the data security in modern digital world. Encryption algorithms were invented to protect integrity, confidentiality and authenticity of data. There are lot of encryption algorithms in use nowadays. These algorithms can be varied according to their functionality. They can be divided into two types, asymmetric and symmetric. Asymmetric algorithm uses two keys called private key and public key to encrypt and decrypt. These type of encryption algorithms are not efficient at encrypting data. Symmetric algorithms use one shared key to encrypt and decrypt. These types of algorithms never share their keys through the internet which makes them secure than the asymmetric algorithms. We use Diffie-Hellman symmetric encryption algorithm for key sharing to make the key sharing much more secure. AES and 3DES are the most used encryption algorithms. But these algorithms have pros and cons as well. 3DES is more vulnerable to brute force attacks. AES is more secured than the 3DES, but it will consume more resources specially when the key length is 256 bit long. We cannot expect to have more processing power for mobiles than the desktop computers. Therefore, 3DES and AES are not an option. Because of that we are developing unique hybrid encryption algorithm using AES and DES that has efficiency in encrypting and decrypting and also suitable for mobile applications, higher security and must be immune to Differential cryptanalysis [26] and Linear cryptanalysis [26] for our mobile application to secure communication between the cloud and the app.

When mobile captures an image of a document with camera, the image will go through compressor algorithm to be light weighted and then it will be encrypted using implemented encryption algorithm, after that it will be sent to the cloud.

When mobile captures an image of a document with camera, the image will go through compressor algorithm to be light weighted and then it will be encrypted using implemented encryption algorithm, after that it will be sent to the cloud.

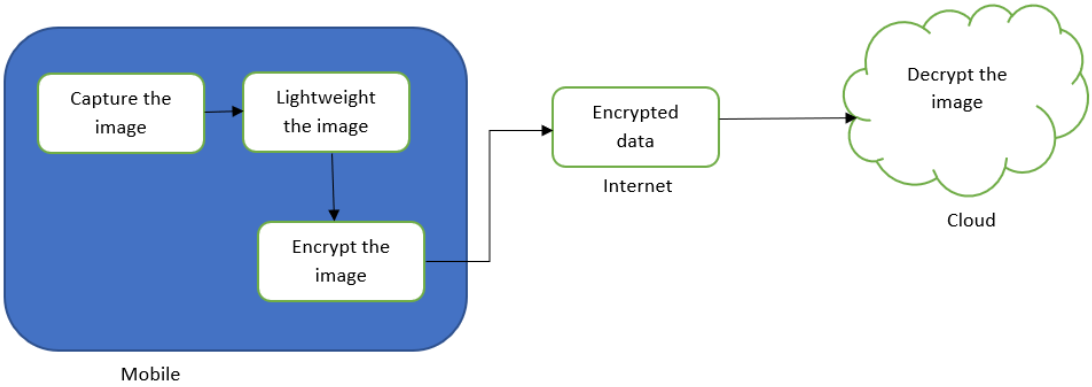

Figure 3.2.5.1: Flow of image compressing and encryption when uploading

3.3. Hardware / Software Requirements

- Client-Side Mobile Phone

Resource	Minimum Requirement
CPU	1.3 GHz above
Storage	1GB or above
RAM	2GB or above
Operating System	IOS or Android

Table 3.3.1: Client-Side Hardware Requirements

3.4. Gantt Chart

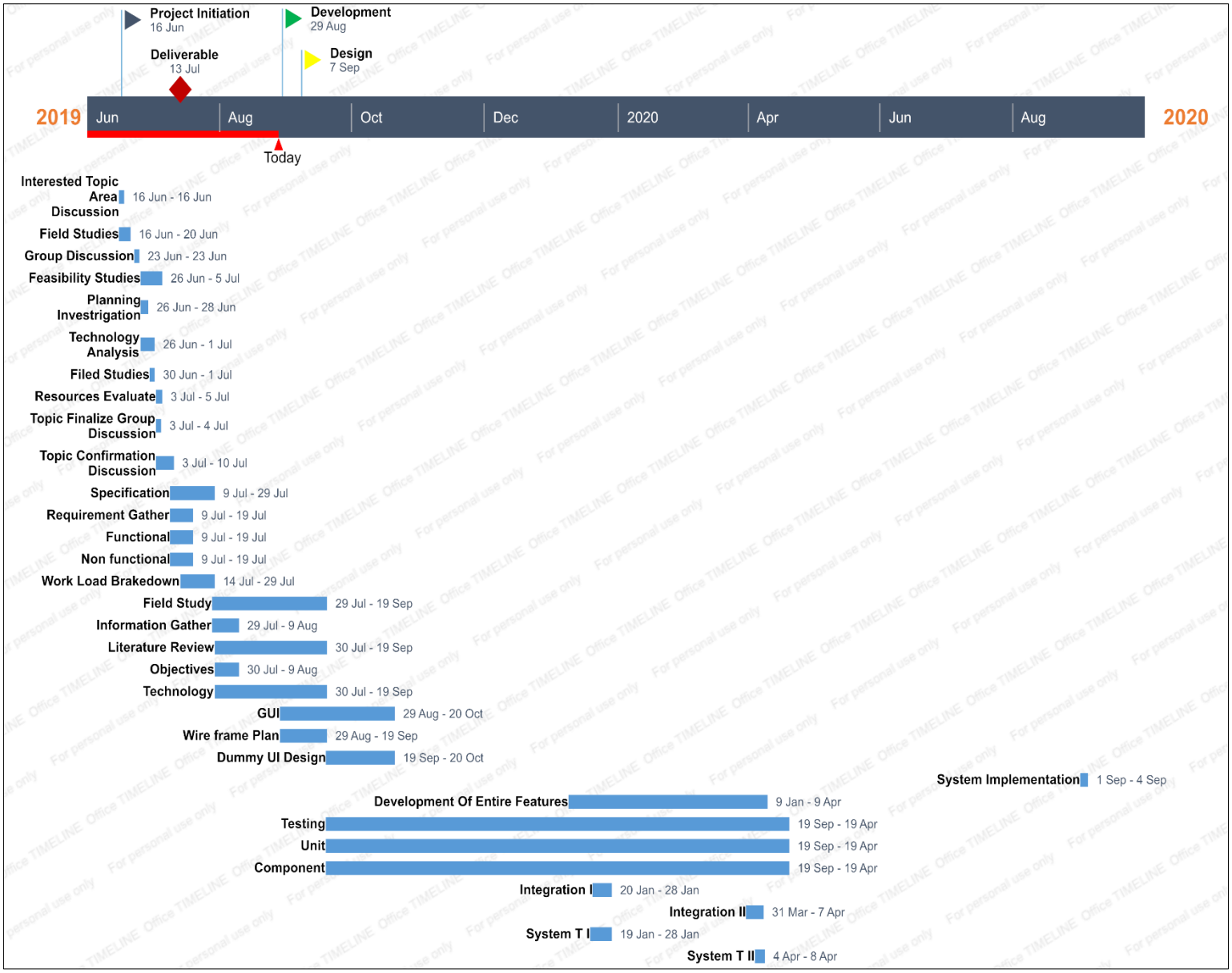

Figure 3.4.1: Gantt Chart

4. DESCRIPTION OF PERSONNEL AND FACILITIES

4.1. Personnel and Facilities

Student ID	Student Name	Work Allocation
IT16102156	E.M.D.D. Ekanayake **Analyzing and Describing Text with Creation of the Mobile Application.**	• Develop a user-friendly cross-platform mobile application for vision-impaired people. o Analyze captured photos and auto-rotate the image in the actual direction where the captured image should be in the actual position. o Make the single command to start analyzing. o Make the option after analyzing the document auto play starts and the user able to listen to the document. • Identify text, images, charts, tables, and the equations uniquely and create another digital image for different identical matches. • In text - Identify language patterns and create a detailed description. • Generate a JSON file of that detailed description for the user to listening purpose. Concatenate all JSON files and generates a single audio clip using open source API service for the user to listen.
IT16176348	I.N. Kalansooriya **Create an Algorithm to Encrypt and Reduce the Size of the Images.**	• Create an algorithm to reduce the size of the images, remove noises and preprocesses to get meaningful information.

		• Create an encryption algorithm to make the assurance of the communication. The encryption algorithm must include the following properties.
		  o High efficiency
		  o Resistant to brute force attacks
		
		• Use the Diffie-Hellman algorithm for the secure key exchange to make key exchange much more secure and reliable.
		
		• Use randomly generated 256-bit long secret key and embed it into the application to make key storage much more secure.
		
		Enforce cloud policies to safeguard and handle the process according to the standards and best practices.

IT16165762	P.S.N. Kularathne
	
	Analyzing and Describing Charts and the Equations in the Documents
		• Detect charts, equations in the uploaded digital photos of the document.
		
		• In charts - Identify the type of the charts as,
		  o Line charts
		  o Bar charts
		  o Pie charts etc.
		
		• Represent the relevant data of the charts accordingly by using natural language processing and deep learning techniques.
		
		• In equations - Identify numbers, variables, and operators in the complex equations accordingly using neural networks and machine learning techniques.

		• Create a detailed description of the identified content of charts and equations. Generate an audio file of that detailed description for the user to listening purpose.
IT16079328	W.R.P. Fernando **Analyzing and Describing Images and Tables in the Documents**	• Detect images, tables in the uploaded digital photos of the document. • In Images - Identify objects, people, places, actions and background regions in the image using deep learning techniques. • In tables - Identify columns and row data in tables accordingly and analyze the content. • Create a detailed description of the content that was identified for both images and tables. Generate an audio file of that description for the user to listening purpose.
IT16102156 IT16176348 IT16165762 IT16079328	E.M.D.D. Ekanayake. I.N. Kalansooriya. P.S.N. Kularathne. W.R.P. Fernando. **Design, Development and testing**	• System Designing • System Development • System Testing

Table 4.1: Work Breakdown Structure

4.2. Work Breakdown Diagram

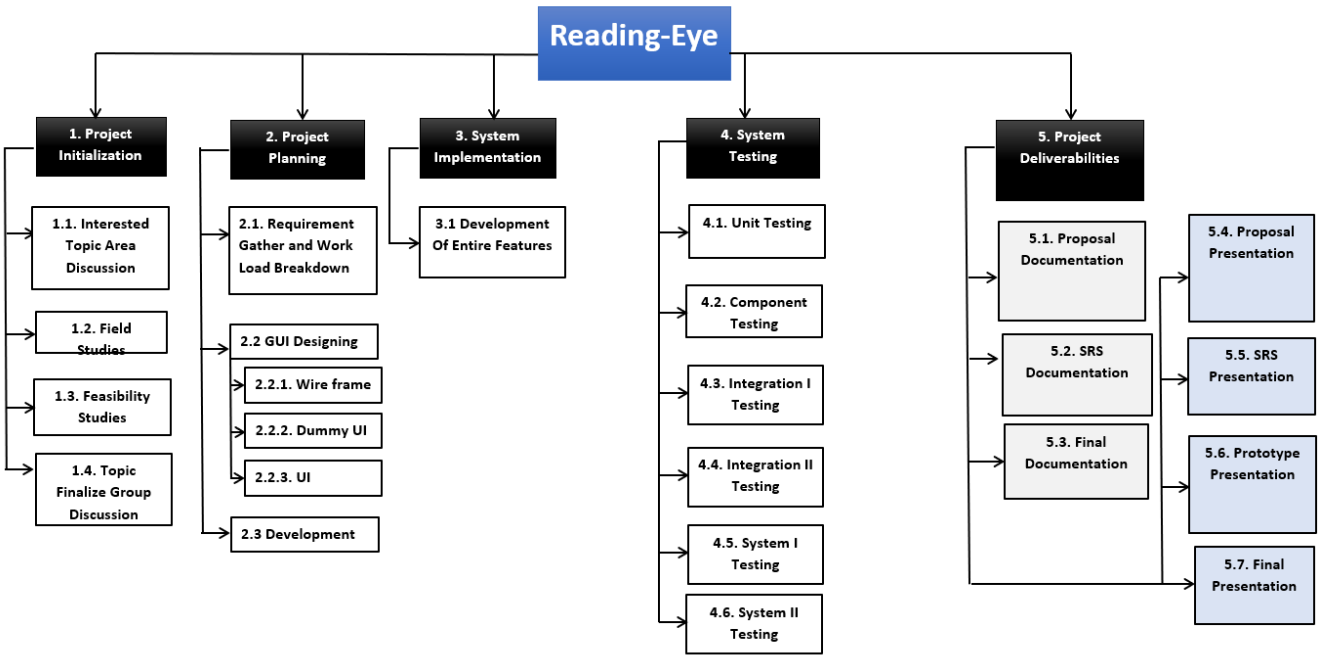

Figure 4.2.1: Work Breakdown Diagram

5. BUDGET PLAN

Budget for one year.

Description	**Qty**	**Unit Cost (Rs.) (for 1 year)**	**Total Amount (Rs.)**
Hosting charges (Azure cloud)	1	76,000	76,000
Cost for devices (Smart phones for Testing NOT NECESSARY)	2	57,500	115,000
Traveling charges	4	10,000	40,000
Communication charges	4	10,000	10,000
Printing charges	-	4,500	4,500
Stationary charges	4	500	2,000
Internet Facility	4	12,000	48,000
Other charges	4	2000	8,000
Total Project Cost Estimate			**303,500**

Table 5.1: Budget Plan

REFERENCE LIST

[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]A. Hellem, “Resources for the Visually Impaired,” All About Vision. [Online]. Available: https://www.allaboutvision.com/lowvision/resources.htm. [Accessed: 06-Aug-2019].

[4]"Braille Facts", Royalblind.org, 2019. [Online]. Available: https://www.royalblind.org/national-braille-week/about-braille/ braille-facts. [Accessed: 07- Aug- 2019].

[5]"Braille", En.wikipedia.org, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Braille. [Accessed: 07- Aug- 2019].

[6]"The History of Braille [Your Braille Resource] - Braille Works", Braille Works, 2019. [Online]. Available: https://brailleworks.com/braille-resources/history-of-braille/. [Accessed: 08- Aug- 2019].

[7] D. Radojichikj, "STUDENTS WITH VISUAL IMPAIRMENTS: BRAILLE READING RATE", (IJCRSEE) International Journal of Cognitive Research in Science, Engineering and Education, vol. 3, 1, 2015, 2015. [Accessed 8 August 2019].

[8]"Braille versions of textbooks help blind college students succeed - Marketplace", Marketplace, 2019. [Online]. Available: https://www.marketplace.org/2017/10/12/braille-versions-textbooks-help-blind-college-students-succeed/. [Accessed: 09- Aug- 2019].

[9] A. Gilani, S. Qasim, I. Malik and F. Shafait, "Table Detection using Deep Learning." [Accessed 10 August 2019].

[10]"Image Detection, Recognition, and Classification with Machine Learning", Azati.ai, 2019. [Online]. Available: https://azati.ai/image-detection-recognition-and-classification-with-machine-learning/. [Accessed: 11- Aug- 2019].

[11] 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.

[12]"Canny edge detector", En.wikipedia.org, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Canny_edge_detector. [Accessed: 20- Aug- 2019].

[13] K. Maladkar, "6 Types of Artificial Neural Networks Currently Being Used in ML", Analytics India Magazine, 2019. [Online]. Available: https://www.analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/. [Accessed: 20- Aug- 2019].

[14]"A Simple Equation Region Detector for Printed Document Images in Tesseract - IEEE Conference Publication", Ieeexplore.ieee.org, 2019. [Online]. Available: https://ieeexplore.ieee.org/iel7/6627713/6628563/06628621.pdf. [Accessed: 22- Aug- 2019].

[15]"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].

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

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

[18] "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].

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

[20]A. Hellem, “Resources for the Visually Impaired,” All About Vision. [Online]. Available: https://www.allaboutvision.com/lowvision/resources.htm. [Accessed: 22-Aug-2019].

[21] J. Brownlee, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks", Machine Learning Mastery. [Online]. Available: https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/. [Accessed: 14-Aug-2019].

[22] C. Vision and C. Code), "Step-by-Step Implementation of Mask R-CNN for Image Segmentation", Analytics Vidhya, 2019. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/. [Accessed: 23- Aug- 2019].

[23] "How to do Semantic Segmentation using Deep learning", AI & Machine Learning Blog, 2019. [Online]. Available: https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning/. [Accessed: 23- Aug- 2019].

[24] "Natural Language Processing is Fun!", Medium, 2018. [Online]. Available: https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e. [Accessed: 19- Aug- 2019].

[25] "Learning to Read: Computer Vision Methods for Extracting Text from Images", Medium, 2019. [Online]. Available: https://medium.com/capital-one-tech/learning-to-read-computer-vision-methods-for-extracting-text-from-images-2ffcdae11594. [Accessed: 21- Aug- 2019].

[26] “A Study of Methods Used To Improve Encryption Algorithms Robustness”, Luminiţa Scripcariu, Faculty of Electronics, Telecommunications and Information Technology, Technical University “Gheorghe Asachi” of Iaşi, Iaşi, ROMANIA. [Accessed: 21- Aug- 2019].

[27] A. Consulting, "Deep Learning based Text Recognition (OCR) using Tesseract and OpenCV | Learn OpenCV", Learnopencv.com, 2018. [Online]. Available:

https://www.learnopencv.com/deep-learning-based-text-recognition-ocr-using-tesseract-and-opencv/. [Accessed: 23- Aug- 2019].

[28] J. Amara, P. Kaur, M. Owonibi and B. Bouaziz, "Convolutional Neural Network Based Chart ImageClassification.." [Accessed: 23- Aug- 2019].

[29] "Natural Language Processing is Fun!", Medium, 2018. [Online]. Available: https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e. [Accessed: 21- Aug- 2019].

[30] "Learning to Read: Computer Vision Methods for Extracting Text from Images", Medium, 2019. [Online]. Available: https://medium.com/capital-one-tech/learning-to-read-computer-vision-methods-for-extracting-text-from-images-2ffcdae11594. [Accessed: 20- Aug- 2019].

[31] “A Study of Methods Used To Improve Encryption Algorithms Robustness”, Luminiţa Scripcariu, Faculty of Electronics, Telecommunications and Information Technology, Technical University “Gheorghe Asachi” of Iaşi, Iaşi, ROMANIA[Accessed: 19- Aug- 2019].

[32] A. Consulting, "Deep Learning based Text Recognition (OCR) using Tesseract and OpenCV | Learn OpenCV", Learnopencv.com, 2018. [Online]. Available: https://www.learnopencv.com/deep-learning-based-text-recognition-ocr-using-tesseract-and-opencv/. [Accessed: 23- Aug- 2019].

[33] J. Amara, P. Kaur, M. Owonibi and B. Bouaziz, "Convolutional Neural Network Based Chart ImageClassification.."Pdfs.semanticscholar.org, 2019. [Online].Available:https://pdfs.semanticscholar.org/5fe7/c360e851f648095bb630eec742c7b809cf2b.pdf. [Accessed: 23- Aug- 2019].

[34]: “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].

[35]: “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].

[36]: " viber encryption overview" :Rakuten Viber [online] Available at :https://www.viber.com/app/uploads/viber-encryption-overview.pdf [Accessed: 22-Aug- 2019].

[37]: "Demystifying the Signal Protocol for End-to-End Encryption (E2EE)" Justino Mora, [online] Available at: https://medium.com/@justinomora/demystifying-the-signal-protocol-for-end-to-end-encryption-e2ee-ad6a567e6cb4 [Accessed: 22- Aug-2019].