DIGITAL ASSISTANT TO AID INDIVIDUALS WITH PRINT DISABILITIES TO INTERPRET PRINTED MATERIALS

Project Id: 2022-024

Project Proposal Report

Priyashan Sandunhetti S. H. S.

B.Sc. (Hons) Degree in Information Technology

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

March 2022

DIGITAL ASSISTANT TO AID INDIVIDUALS WITH PRINT DISABILITIES TO INTERPRET PRINTED MATERIALS

Project Id: 2022-024

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology

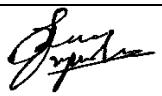
Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

March 2022

A declaration, copyright statement, and the statement of the supervisor.

We declare that this is our work, and this proposal does not incorporate without acknowledgment 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 acknowledgment is made in the text.

Name	Student ID	Signature
Priyashan Sandunhetti S. H. S.	IT19187242	

The above candidates are researching the undergraduate Dissertation under my supervision.

Name of supervisor: Dr Anuradha Jayakodi

Name of co-supervisor: Ms Shashika Lokuliyana

Signature of supervisor: Date:

Signature of co-supervisor: Date:

ABSRACT

There is a huge percentage of the world population who cannot interpret printed materials due to various reasons. This inaccessibility can be caused by vision impairments, physical dexterity problems, learning and literacy difficulties. This inability to read printed material due to visual, physical or perceptual can be identified as print disabilities. These disabilities can affect individuals' day to day life as well as their education and literacy. Even though there are accessible mediums for print disabled individuals to comprehend printed/written materials by methods like braille, ELIA Frames and audiobooks vast majority of necessary documents, books are not available in most of the countries. Also, most of the written materials cannot be converted into accessible mediums due to copyright and other various laws. This causes a global reading inequality and there are millions of people affected by this issue [1]. Even when using the available accessible mediums to interpret the printed materials there are still many barriers that cannot be overcome by these mediums. For instance, images, mathematics, tables and graphs are hard to interpret using methods like braille. Also, when it comes to materials like legal and personal documents, they are rarely available in accessible mediums, and it is harder to get the assistance of a third party to interpret such documents. For these issues, we propose a solution to overcome the barriers of the print disabled to interpret printed materials. This solution will be implemented to run on users' mobile devices and it will be able to scan and interpret printed material in a vocal form. Furthermore, this solution will be implemented with several accessibility options to cater for every possible user. This system will consist of a mobile application and cloud backend where suitable image processing and computer vision algorithms are stored. The application will also be voice-assisted for better accessibility. With the use of this solution, we hope to lower reading inequality and make the lives of individuals with print disabilities much easier.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF APPENDICES

1. Introduction

1.1.Background and Literature Survey

According to current statistics, WHO estimates there are at least 2.2 billion people with near or distance vision impairments [2]. Most of these individuals have difficulties when trying to interpret printed materials. Other than vision-impaired people, there are many forms of disabilities that cause an individual to be print disabled [3]. Overall, we can conclude that a considerable portion of the global population suffers from one or many types of print disabilities [1]-[3]. Reading is closely related to humans' everyday life and interacts with everyday elements like education, literacy, work, healthcare, justice, political participation and cultural belongings. Also, with reading being the main format of gathering information and communication, it is a necessary skill to survive in most modern societies. As shown in figure 1.1 our survey shows that the vast majority of participants considered reading to be very important in individuals' day-to-day life.

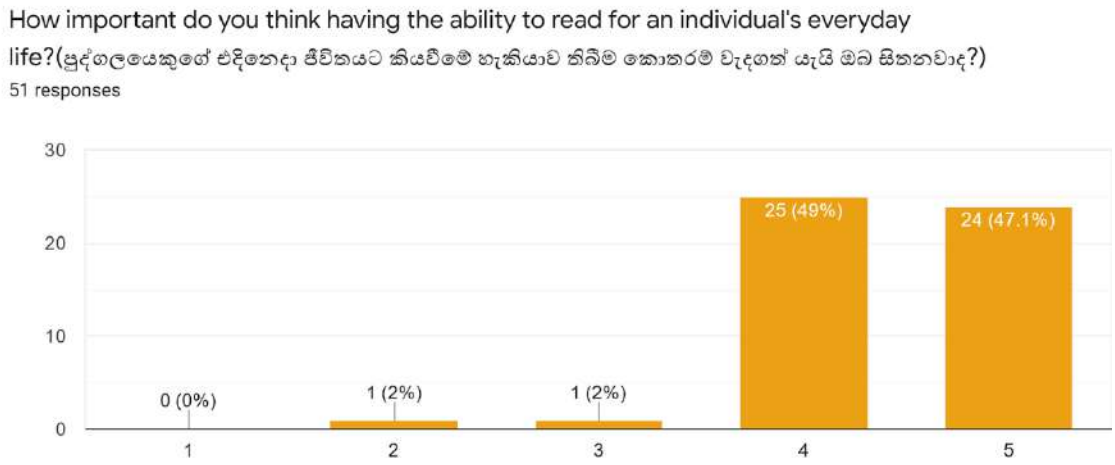


Figure 1.1 - Survey responses to the importance of reading ability

Even though there are traditional solutions like braille [4] to aid these individuals to interpret printed materials, braille literacy of the print disabled individuals are as low as 10% [5]. Also due to the average cost of a brail book being higher than the normal issue and because of the low availability of braille books, braille cannot be considered as the best solution for print disability. For materials that are not available in accessible formats like braille, print disabled individuals must have to rely on a third party. This third party can be a human or an assistive tool [6]. When considering another human who can access

normal printed material to interpret the printed documents on print disabled individuals' behalf there can be issues like privacy and mistrust. For personal, legal and confidential documents, a print disabled individual cannot solely rely on another human being to assist.

With these issues, print disability has a huge impact on an individual's everyday life in many aspects [figure 1.2]. This discriminates against most basic human rights like the right to education, right to work and even political and justice rights.

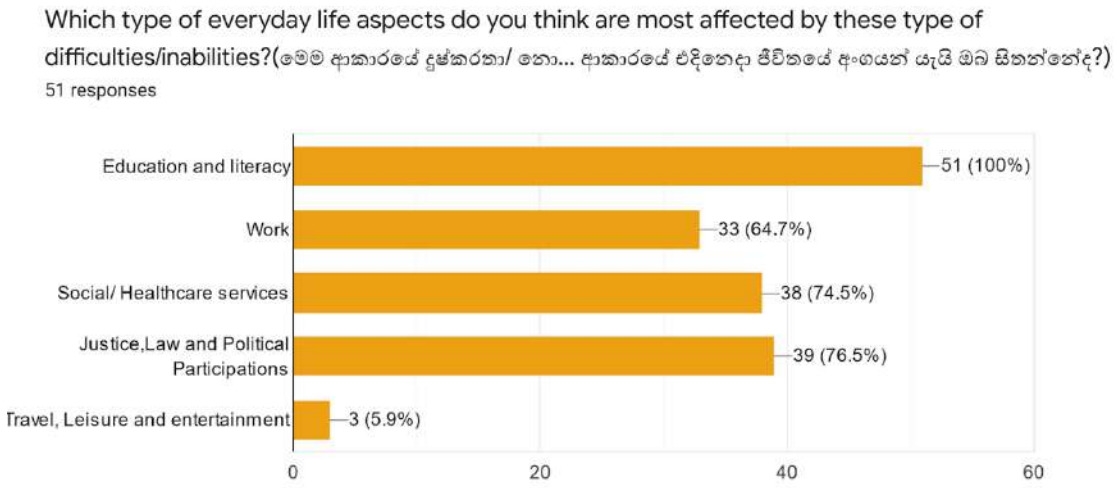


Figure 1.2 - Survey results for life aspects affected by reading inability

This causes a large gap between print disabled individuals and the general population when it comes to reading rights and equality. As shown in figure 1.3 most of the participants in our survey agrees that reading should be more accessible and equal among everyone and everyone must have equal rights to access printed materials.

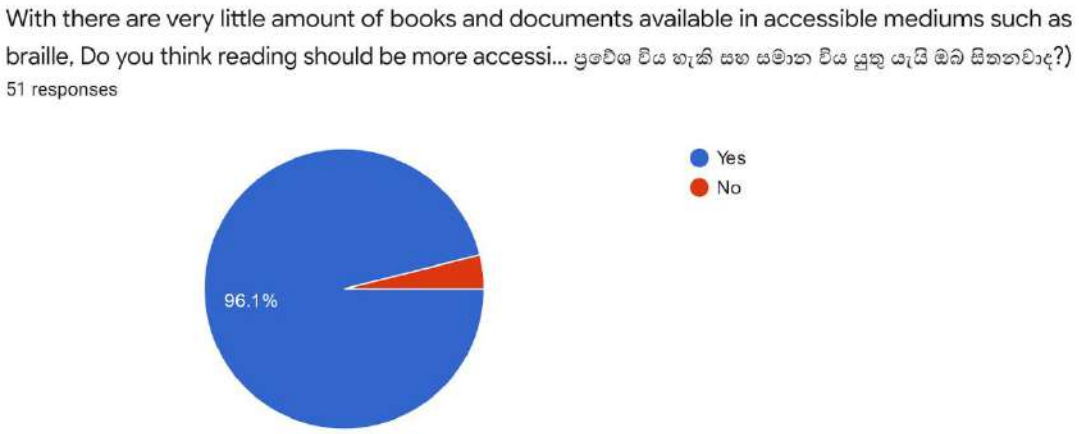


Figure 1.3 - Survey results on the reading equality and rights

With these issues in mind, we propose a solution to aid print disabled individuals to interpret printed materials with better accessibility options to enable the application to be used by themselves. This solution mainly focuses on the Assistive Technology research area and consists of four main modules.

- Document capturing and document zone content classification module
- Chart interpretation module
- Image interpretation module
- Texts and mathematics interpretation module
- Tables interpretation module

This individual proposal mainly focuses on the first module which is the document capturing and document zone content classification module and the graph interpretation module. These modules fall under the computer vision research area and are crucial to the system because all of the other algorithms' accuracies and work depends on the classification and it is a necessary and most important module of the system.

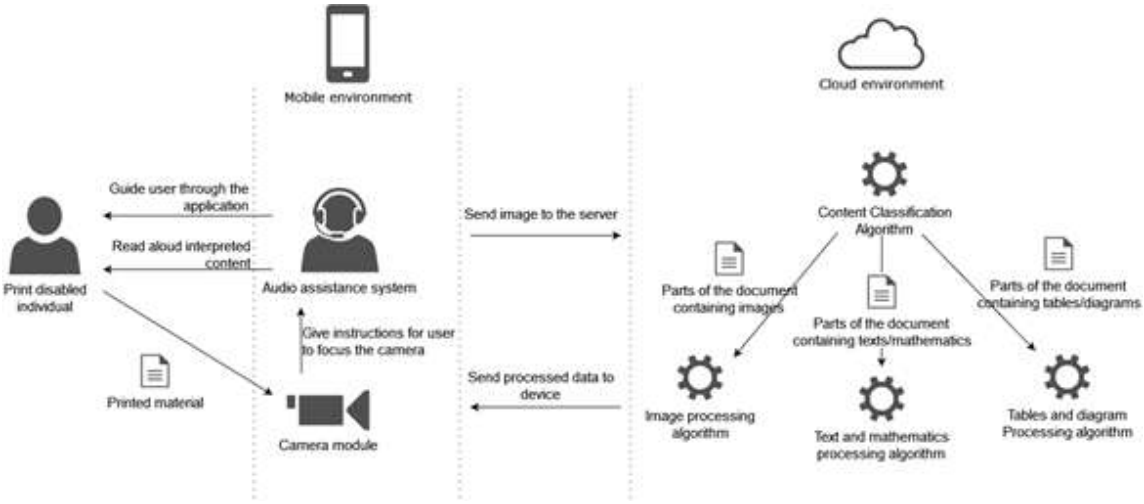


Figure 1.4 - Modules of the system

Furthermore, the chart interpretation module is necessary for the system because otherwise, the print disabled individual won't be able to access and interpret graphs in a general printed material. Even though there are methods like tactile graphics to aid blind and vision-impaired people to read charts, the materials which are available in tactile graphics [10] are rare and useless in day-to-day life.

1.2.Research Gap

When considering the whole system, there are already implemented mobile tools to assist print disabilities. But most of them are implemented for a limited audience and lacks functions like table interpretation and mathematical equation interpretation [6],[7]. Furthermore, most of these solutions doesn't have the options to guide the user in the material capturing process and to automatically capture the document when in the focus range. Also, there are promising solutions like wearable devices [8] to aid reading, this type of solution also lacks practicality when used by blind people because tracing in printed lines can be difficult for them. When looking for publicly available solutions in general app stores there are really good applications to interpret captured documents and images, but they lack the functions like interpreting every content (Text, images, tables) within a document at once and also they include general UI and doesn't have accessibility options for vision-impaired users.

Furthermore, the results of the survey indicate that most of the participants do not know much about available assistive tools [Figure 1.5] and most of them agrees that there are no sufficient accessibility options in available tools [Figure 1.6].

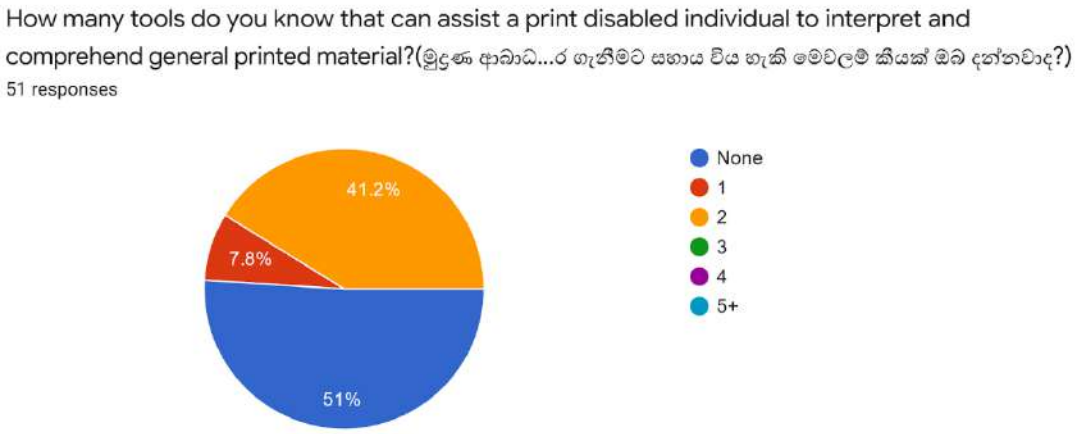


Figure 1.5 - Popularity of available assistive tools

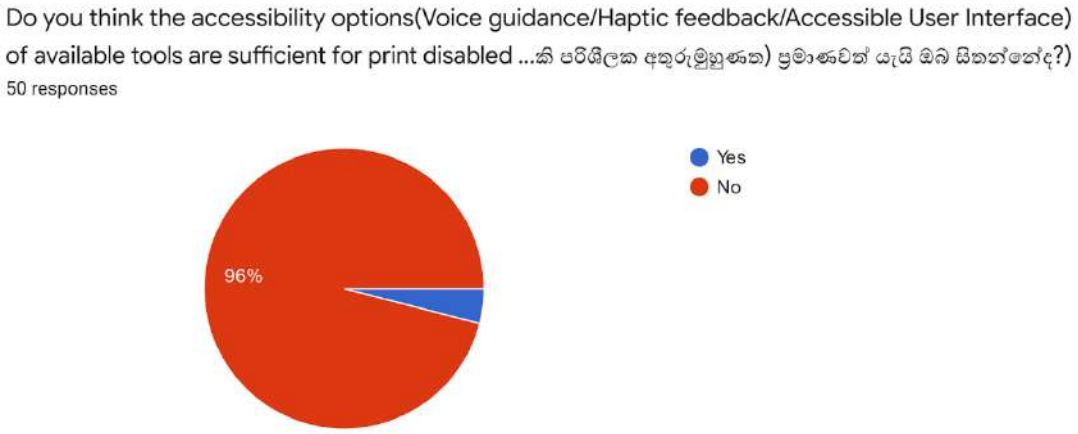


Figure 1.6 - Survey on accessibility options of available tools

The individual research component which is the document zone and content classification also already have been researched and many papers can be found on the topic [9]. However, most of them lack the ability to detect and classify all the necessary content types that could be included in a printed material [11],[12]. Andrea Corbelli et al. [13] (Research B) addresses this issue using the XY-cut algorithm to segment the document and classify the segmented document using heuristic methods for table detection and SVM classifier for other classes. This method is able to classify most of the available but lacks

different chart classifications. Ranajit Saha et al. [14] (Research A) have developed a graphical object detection framework that can segment and classify tables, figures and equations separately but lacks the ability to separate charts and graphs from images. Furthermore, as shown in [15] (Research C) document layout analysis can even be done by using one-dimensional convolutional neural networks which is fast and economic in data usage that suits the performance capabilities of mobile devices. Also, this low computational cost means this approach can be implemented in even cloud environments without much cost. But this approach [15] also classifies charts and images under the same class as figures which is not feasible for our kind of document interpretation model. Also, because of the nature of the users that we are implementing the system for, there will be a need to crop out the document from the overall captured image and enhance the document by fixing the perspective issues and adjusting noise and lighting issues to increase the clarity.

Table 1.1- Research gap for document segmentation and classification with existing systems

	Research A	Research B	Research C	Proposing solution
Detect and enhance the document from the overall scanned image (Distortions/Lighting)	No	No	No	Yes
Detect images and graphs separately	No	Yes	No	Yes
Classify text and mathematic expressions separately	No	Yes	No	Yes
Tabular structure detection	Yes	Yes	Yes	Yes
Optimized for mobile and cloud usage	No	No	Yes	Yes

The other part of this research component is the chart decoding module. This part of the research also requires a classification method for different types of charts. Then each chart will be decoded using proper methods and the decoded data will be turned into simple plain English in order to read aloud for print disabled users as seen in [16] (Research P). As demonstrated in [16], it is possible to classify multiple types of charts with greater accuracy and generate alt-text for each type of chart using suitable algorithms. Furthermore, I'll be comparing this proposed solution in contrast to Research Q [17] and Research R [18].

Table 1.2- Research gap for chart classification and interpretation with existing systems

	Research P	Research Q	Research R	Proposing solution
Classify different types of charts	Yes	Yes	Yes	Yes
Extract data from bar charts	Yes (Vertical/ Horizontal /Stacked)	Yes (Vertical)	Yes (Vertical/ Horizontal /Stacked)	Yes (Vertical/ Horizontal /Stacked)
Extract data from pie charts	Yes	No	Yes	Yes
Extract data from line charts	No	No	Yes	Yes
Provide a textual description of chart data in plain English	Yes	No	No	Yes

By reviewing these available approaches, proposing solution can be implemented with many more novel and creative options which will be a hybrid solution for captured document classification and decoding the identified charts within the document.

1.3. Research Problem

According to the world health organization [19], there are 2.2 billion of the global population have near or distant vision impairment. Out of those vision-impaired populations, 300 million individuals can be considered as print disabled [20]. Print disability is a difficulty or inability to read printed material due to a perceptual, physical or visual disability [21]. These disabilities vary from literacy difficulties to vision impairments and blindness. Learning disabilities like dyslexia can also be considered as print disabilities. Most of these people access documents via methods like braille or in the forms of audiobooks. The issue is, there are 1%-7% of books are available in braille format globally [20]. Also, the majority of the documents, books that an individual has to use in day-to-day life are not available in accessible formats. As shown in figure 1.7, the majority of the participants that participated in the survey agree that there is a need for better accessible tool aid print disabled individuals.

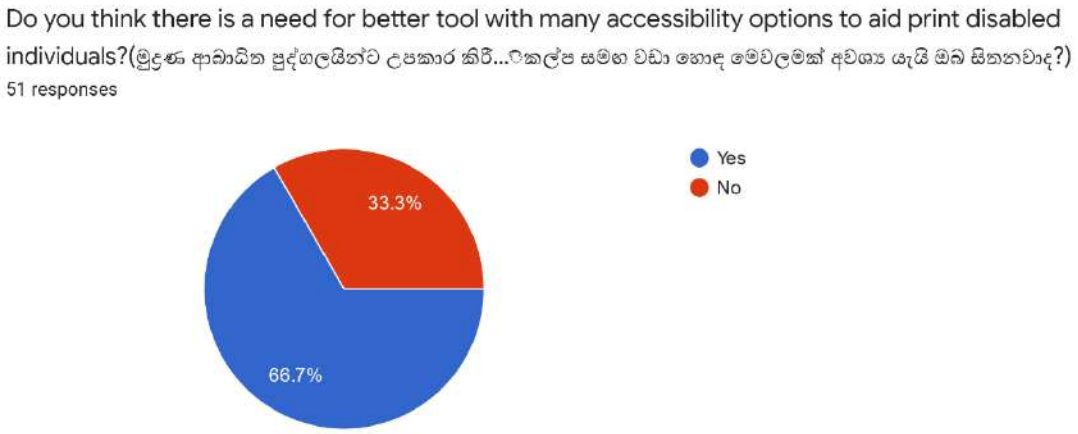


Figure 1.7 - Survey results on the need for better tools for print individuals

There are already available mobile tools to assist the reading of the materials but most of them either lacks the accessibility options for the vision-impaired users and all of them doesn't convert complex features that can be found in a document such as tabular structures, charts, images and mathematical equations [6]. When it comes to charts, the only way for a print disabled individual to access a chart in printed material is the tactile methods which are very rare and costly to produce. Even the available tactile charts can be hard to grasp for most individuals.

If we are to implement a mobile smartphone-based solution to address the print disability issue, there must be proper accessibility options in place [Figure 1.8]. In the solution, there is a need to scan the document automatically and the scanning process should also be assisted because most of the print disabled individuals cannot properly capture a physical document using a smartphone.

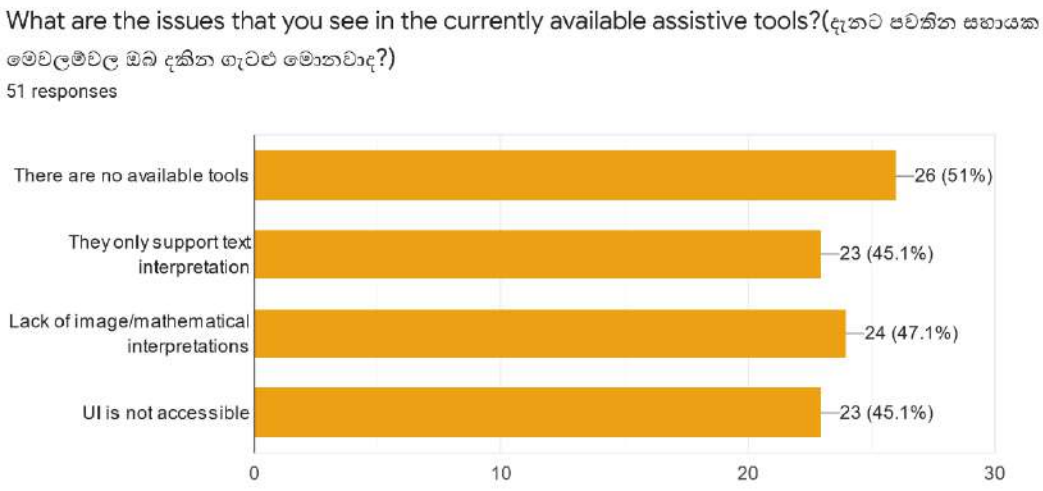


Figure 1.8 - Survey results on issues in the current accessibility tools for print disabled

In this proposal, the addressing issues are the assisted document capturing, enhancing, segmenting and classification of the segments and interpreting the identified chart into plain English.

2. Objectives

2.1. Main Objective

The main objective of this research is to develop the assisted document capturing methods, document content segmentation and classification of the segmented parts and the interpretation of the identified charts within the document.

2.2. Specific Objectives

- To assist the user with the document capturing process
  The document capturing process has to be assisted because most of the users that will be using the system will not be able to capture a physical document using their mobile phone camera properly.
- To crop out and enhance the captured document
  The captured image will include an outside area other than the document and the document will have perspective errors and lighting, noise issues which will affect the systems' accuracy greatly. So, the document has to be cropped out from the overall image and enhanced.
- To segment and classify the regions of the document
  The content in the document has to be segmented and classified in order to interpret the parts of the document using suitable methods.
- To classify the detected charts furthermore into different types of charts
  For each type of chart, there are different methods to extract data. Before extracting the data from the charts, there is the need to classify charts by their type.
- To extract the data from the charts and turn the extracted data into plain English sentences in a meaningful way.
  Finally, the identified charts in the captured document must be interpreted in simple English in order to read aloud to print disabled users.

3. Methodology

This research component of the project includes several computer vision algorithms and image processing methods. The whole component will be distributed in two environments as client-side in the mobile phone and the classification and interpretation process will be done in the server-side cloud environment. In the mobile environment, there will be a pre-trained deep learning computer vision model to aid the capturing process which will be developed using TensorFlow lite because it supports on-device machine learning and performs way better in mobile devices.

Then the captured image will be sent to the backend cloud environment to process. First of all, the image will be enhanced in the backend. The image enhancements can be done by using python libraries like OpenCV and Python imaging library (PIL). Then the document segmentation can be done by using methods like top-down, bottom-up and heuristic methods. Most of these methods are used before the development of convolutional neural networks (CNN) and the CNNs will be able to segment the document more accurately in less time. Another advantage of using CNN is it can be developed to deal with many issues in a single model. So there is a capability to implement both the segmentation and classification process in a single model. Also, for the different types of chart classifications, a trained CNN model will be the best choice because CNNs are specifically developed to deal with image data and excels at image processing [16]. For the classification processes, trained SVM (Support vector machines) models are widely used as well but for image processing purposes CNNs tend to be more accurate [22].

For the data extraction from the charts, rule-based methods and deep neural networks can be used but for better efficiency and results, hybrid methods combining advantages of rule-based methods and deep neural network-based methods proves to be more useful [18]. Then to convert extracted data into meaningful sentences, predetermined templates can be used for each chart. Also, there is the option to use LSTM (Long Short-Term Memory) networks to generate meaningful text sequences.

As per the datasets, for most of the classification algorithms, there are already available datasets like ExcelChart400K [18], GO-IIIT-5K [11] and several ICDAR datasets [11].

3.1. System Architecture

3.1.1. Software Solution

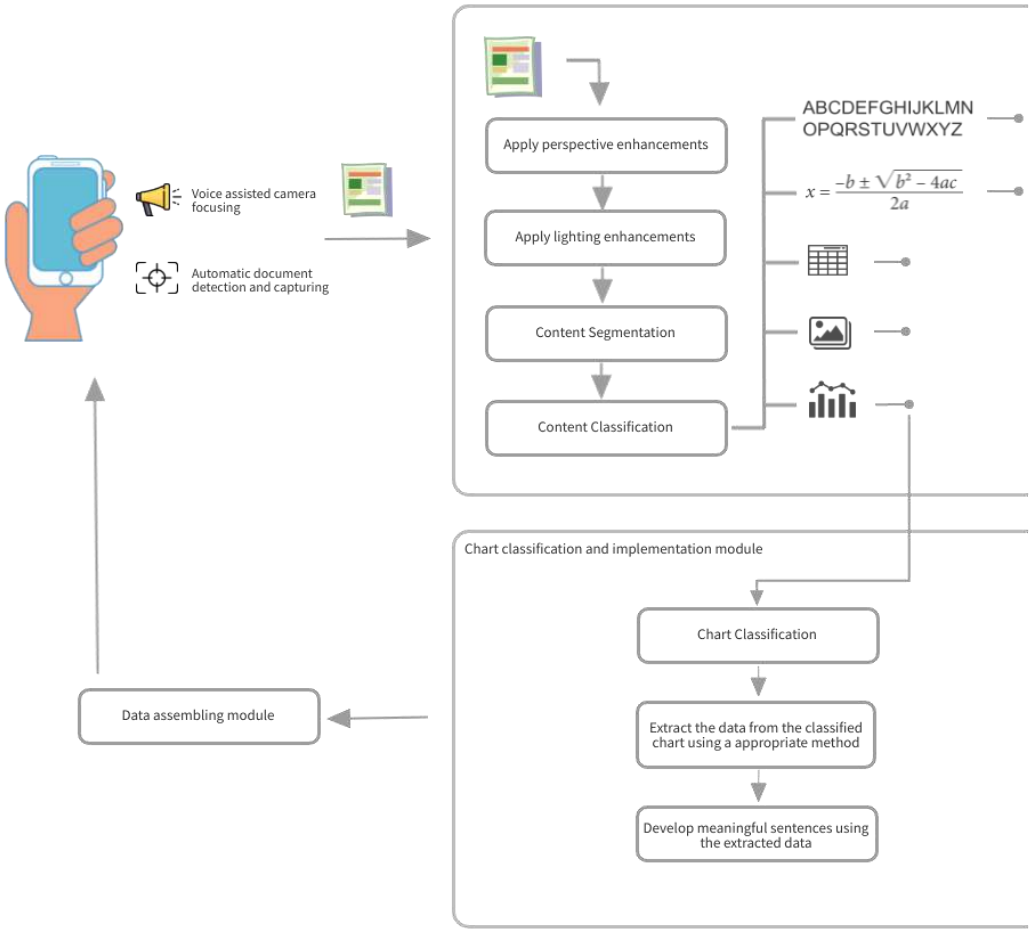


Figure 3.1 – System overview diagram

3.2. The flow of the project

3.2.1. Requirement gathering and analysis

After reviewing relevant research papers and review papers the requirements of the proposing system was decided. Furthermore, a survey was done to gather information regarding the matter [Appendix C]. The features and the requirements can also be highlighted using the results of the survey. Also, for the analysis phase, already publicly available mobile applications are tested for usability regarding the target user audience and the accuracies of the system.

3.2.2. Feasibility study

The proposed system will be developed within the allocated time period with completing each phase within the time limit shown in the Gantt chart [Appendix A]. There will be a need to learn new technologies and frameworks in order to gain proper knowledge and implement the system. Because most of the models and algorithms will be developed with optimization in mind, there will not be a need for high-end devices to run the system so the proposed research component also will be economically feasible.

3.2.3. Implementation

The system will be implemented with the proposed features and accessibility in mind. Proper software architectures will be used for easier component interconnectivity. All of the models will be developed with proper practices and will be trained with the best available datasets

3.2.4. Testing

This system will be tested using standard methodologies. Firstly, each component will be tested separately by mostly testing the trained models using testing datasets. Then after the proper integration testing, the system will be released for usability testing. The system will also be tested with the help of print disabled/ vision impaired individuals to ensure the accessibility of the system. This will assure a software system with fewer bugs and issues and a better user experience.

3.3. Project requirements

3.3.1. Functional requirement

- A proper document scanning module should be in place
- The scanning process should be assistive
- Document zone segmentation and classification model
- Proper chart classification model
- Proper chart interpretation and language processing model

3.3.2. Non-functional requirement

- Faster processing
- Better accuracy
- Well optimized for cloud/mobile use

3.3.3. User requirements

- User must have a smartphone with working camera
- User must be able to control the device with touch inputs or voice commands
- User must have basic English knowledge
- User must not have hearing disabilities

4. Budget and justification

For this research component, we do not aim to commercialize our individual components and the whole research will be commercialized as a mobile application. Following is the budget justification for the whole research system.

Table 4.1 - Budget and budget justification

Item	Cost (Rs)
App publishing costs on app stores	5000.00
Backend hosting cost	10000.00
Research paper publishing cost	5000.00
Total	20000.00

Reference List

[1] W.-J. Chang, L.-B. Chen, C.-H. Hsu, J.-H. Chen, T.-C. Yang, and C.-P. Lin, "MedGlasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients," IEEE Access, vol. 8, pp. 17013–17024, 2020.

[2] "Blindness and vision impairment," Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 11-Feb-2022].

[3] Southern Cross University, "Students with a print disability - Southern Cross University," Edu.au. [Online]. Available: https://www.scu.edu.au/copyright/for-students/students-with-a-print-disability/. [Accessed: 11-Feb-2022].

[4] K. Smelyakov, A. Chupryna, D. Yeremenko, A. Sakhon, and V. Polezhai, "Braille character recognition based on neural networks," in 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018.

[5] A. Graves, "Braille literacy statistics research study: History and politics of the 'braille reader statistic': A summary of AFB leadership conference session on education," J. Vis. Impair. Blind, vol. 112, no. 3, pp. 328–331, 2018.

[6] N. D. U. Gamage, K. W. C. Jayadewa, and J. A. D. C. A. Jayakody, "Document reader for vision impaired elementary school children to identify printed images," in 2019 International Conference on Advancements in Computing (ICAC), 2019.

[7] S. Muralidharan, D. Venkatesh, J. Pritmen, R. Purushothaman, S. J. Anusuya, and V. Saravanaperumal, "Reading Aid for Visually Impaired People," International Journal of Advance Research, vol. 4, no. 2, 2018.

[8] R. Shilkrot, J. Huber, W. Meng Ee, P. Maes, and S. C. Nanayakkara, "FingerReader: A wearable device to explore printed text on the go," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, 2015.

[9] T. A. Tran, K. Oh, I.-S. Na, G.-S. Lee, H.-J. Yang, and S.-H. Kim, "A robust system for document layout analysis using multilevel homogeneity structure," Expert Syst. Appl., vol. 85, pp. 99–113, 2017.

[10] C. Jayant, M. Renzelmann, D. Wen, S. Krisnandi, R. Ladner, and D. Comden, "Automated tactile graphics translation: In the field," in Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility - Assets '07, 2007.

[11] R. Saha, A. Mondal, and C. V. Jawahar, “Graphical object detection in document images,” in 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 51–58.

[12] D. Augusto Borges Oliveira and M. Palhares Viana, “Fast CNN-Based Document Layout Analysis,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1173–1180.

[13] A. Corbelli, L. Baraldi, F. Balducci, C. Grana, and R. Cucchiara, “Layout analysis and content classification in digitized books,” in Communications in Computer and Information Science, Cham: Springer International Publishing, 2017, pp. 153–165.

[14] R. Saha, A. Mondal, and C. V. Jawahar, “Graphical object detection in document images,” in 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 51–58.

[15] D. Augusto Borges Oliveira and M. Palhares Viana, “Fast CNN-Based Document Layout Analysis,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1173–1180.

[16] A. Balaji, T. Ramanathan, and V. Sonathi, “Chart-Text: A fully automated chart image descriptor,” arXiv [cs.CV], 2018.

[17] W. Dai, M. Wang, Z. Niu, and J. Zhang, “Chart decoder: Generating textual and numeric information from chart images automatically,” J. Vis. Lang. Comput., vol. 48, pp. 101–109, 2018.

[18] J. Luo, Z. Li, J. Wang, and C.-Y. Lin, “ChartOCR: Data extraction from charts images via a deep hybrid framework,” in 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021.

[19] “Blindness and vision impairment,” Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 31-Jan-2022].

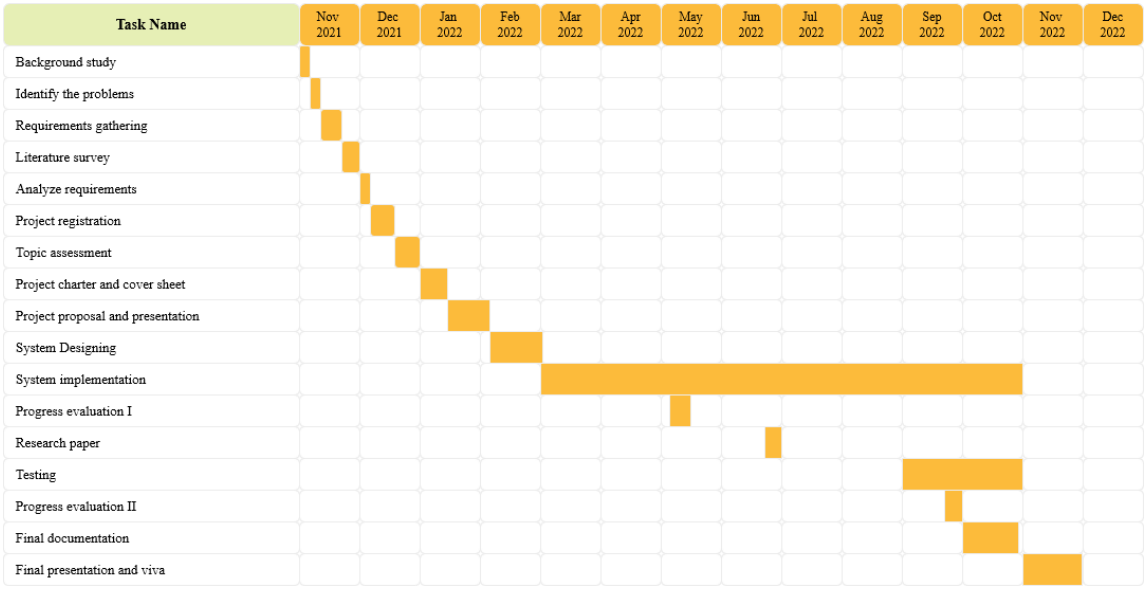
[20] B. Okpeh, “Towards ending the information poverty among persons with print disabilities: A proposed draft (amendment) bill for disability diversification of the national library of Nigeria,” SSRN Electron. J., 2021.

[21] “What is print disability,” Visionaustralia.org. [Online]. Available: https://www.visionaustralia.org/services/print-accessibility/what-is-print-disability. [Accessed: 31-Jan-2022].

[22] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. N. R. S. N. Prudhvith, and N. Kumar, “Image Classification using SVM and CNN,” in 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020.

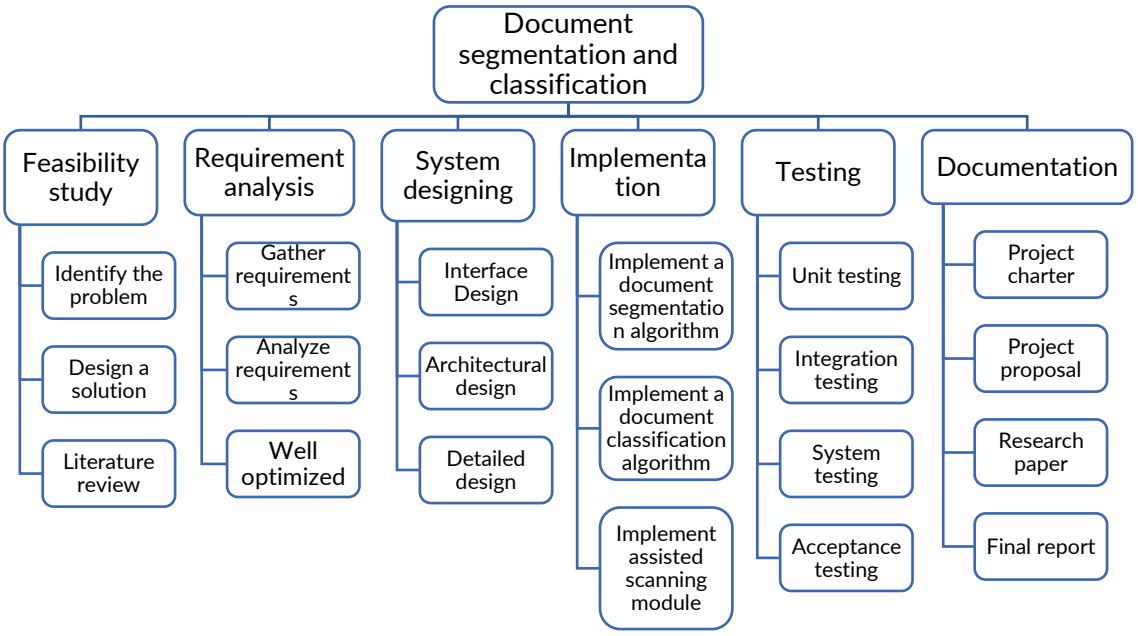
Appendices

A. Gantt chart



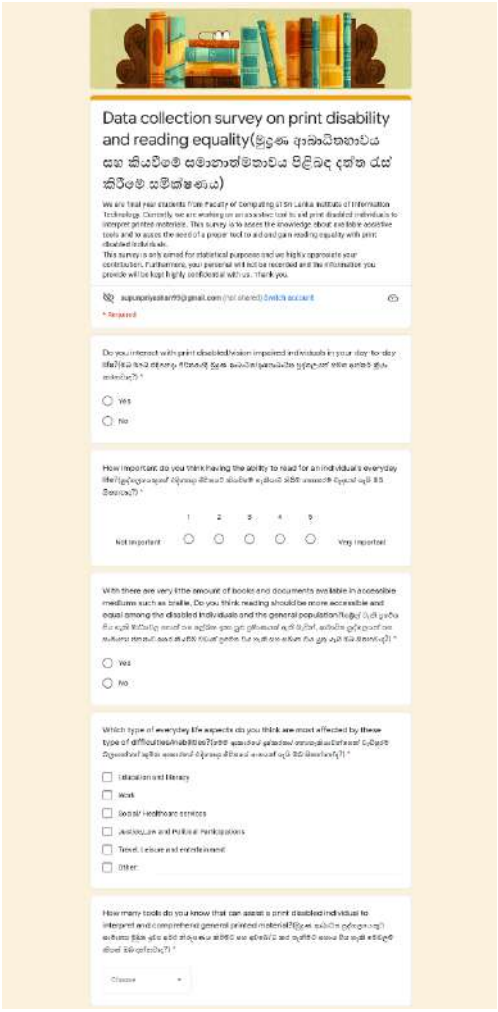
Appendix A

B. Work Breakdown Structure

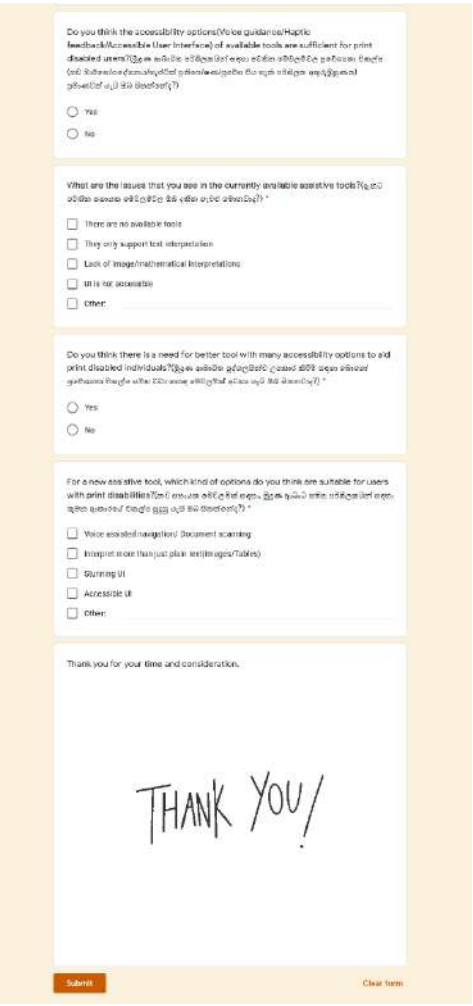


Appendix B

C. Online survey



Appendix C



Appendix D