# Digital Assistant to Aid Individuals with Print Disabilities to Interpret Printed Materials

Priyashan Sandunhetti S. H. S.
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
it19187242@my.sliit.lk

Sanduni Madara P. G.
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
it19392172@my.sliit.lk

Dilitha Ranjuna G.P.
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
it19156484@my.sliit.lk

Prabhash K. V. A. S
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
it19117492@my.sliit.lk

J.A.D.C.A.Jayakody
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
anuradha.j@sliit.lk

Shashika Lokuliyana
*Faculty of Computing*
*Sri Lanka Institute of Information Technology*
Malabe, Sri Lanka
shashika.l@sliit.lk

***Abstract*— Print disability is the difficulty or inability to read printed material due to a perceptual, physical, or visual disability. An individual can be classified as a print-disabled individual if the person requires alternative access or accessible formats (Braille, Audio) to gain information from printed materials. Print disability can be caused by vision impairments, blindness, physical dexterity problems, learning disabilities, brain injuries, cognitive impairments, and literacy difficulties. There are millions of people around the world who cannot interpret printed materials due to the above difficulties. This can affect the individual's day-to-day life as well as their studies. Even though there are tools to interpret printed materials into text, most of the tools are not sufficient to aid print-disabled individuals and lack accessibility options within the tools. Therefore, we propose to develop a mobile-based application to aid print disabled individuals to interpret printed materials which they cannot access without any assistance otherwise. This application will be based on machine learning and image processing and will be able to interpret printed materials including text and paragraphs, mathematical formulas, tables, charts, and images. Also, the application will be designed with accessibility features which enable print disabled individuals to use the application without any third-party assistance.**

***Keywords—Print disability, Computer vision, Image processing, Document image analysis***

## I. INTRODUCTION

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.

Even though there are traditional solutions like braille [4] to aid these individuals to interpret printed materials, braille literacy of print-disabled individuals is 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, and 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. This discriminates against most basic human rights like the right to education, right to work and even political and justice rights. This causes a large gap between print-

disabled individuals and the general population when it comes to reading rights and equality.

To address these issues, 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 five main modules.

- Document zone segmentation and content classification
- Chart Interpretation
- Image captioning
- Texts and mathematics interpretation
- Tabular data interpretation

By successfully implementing the mentioned algorithms and developing a proper mobile-based interface with many accessibility options people who have visual disabilities and are unable to interpret printed materials will be able to coexist in society without feeling left out and it will make their day-to-day life much easier.

## II. LITERATURE REVIEW

### A. *Document capturing and segmentation*

This 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] address 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] 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 the [15] 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 clarity.

### B. *Chart Interpretation*

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]. 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, there are already proposed solutions for chart data extractions but most of them lacks interpretation of multiple charts [17] and even they supported multiple chart interpretation they lack the text description which is needed for this research [18].

### C. *Table comprehension*

Technical difficulties" is a term used to describe a range of conditions that make it difficult or impossible for a person to read printed materials. Braille is a well-known tool for visually impaired readers, but many argue that it can't be used in all circumstances. In order to recognize tables, we are providing a computerized design tool to help designers create Braille-friendly table-top designs. This program can be used to do tasks by people who have vision problems. It allows users with vision impairments and print challenges to complete their tasks. By utilizing this software, the user can sort things out and have an idea of what and how the table would look like. This program analyzes the data in the tables and speaks the precise information to users who are unable to print. The user-friendly instructions plus the fact that everything you need is in one accessible place make this program really easy to use. Users can utilize this technology and complete their tasks independently without any hesitation. S.S. Paliwal et al. (Research A) proposed the table detection and extraction model called “TableNet” which is a deep learning model for end-to-end table detection and tabular data extraction from scanned document images. . Namysl, M. et al.(Research B) conducted a table recognition and semantic interpretation system and it was supposed to recognize the most frequent table formats. and Hashmi, K. A. et al. (Research C) published an analysis of table recognition in document images with deep neural networks. It is supposed to recognize most frequent table formats and it should be able to recognize them in text as well as audio, video and audio format.

### D. *Image captioning*

Image captioning can be done in many ways, but when it comes to blindness, images should be described in such a way that a person who has been blind since birth can understand. This component of the research which is image captioning is also have already been researched by much expertise. Various tools have been built to interpret images in printed documents, and several assistive tools have also been implemented for

print disabled people. However, the majority of them are missing some critical factors that should be improved for use by print-disabled people. In most of the existing tools, even those that are great and are available on the Play Store, do not provide an explanation for this factor [8]. Some tools don't even describe the basic colors in it [7], [9]. Since "Image captioning algorithm based on multi-branched CNN and Bi-LSTM" paper [7] is a great paper which uses an attention mechanism to get the key features of the image to describe it, is not done for print disabled people. So, it also lacks the sufficient explanation for a blind person to understand an image in a printed document.

### *E. Math interpretation*

There are numerous ways to identify the math equation, but when it comes to blindness, neither the reading component nor the identified math equation were included in the mobile app. The majority of researchers focus on reading or identifying mathematical equations. The issue is that both components must work together for visually challenged students to read math. Most projects are completed using equation images that have been transformed to the LaTex language for math, with most study focusing on different languages to detect the math component. The Latex formula must then be transformed into text that can be read aloud and sent to the text-to-speech system. As a result, using mobile apps consumes a lot of computing power.

Rouhan Noor [10] created the identified numerical data in Bangla. The model was created using a Convolutional Neural Network (CNN) and only focused on converting Bengali language numerical data to number format. The neural network was built with 5 convolutional layers. It was also designed to recognize handwritten Bengali numbers.

The mathematical expression discovered using an aggregation-based bi-directional neural network is being studied by Xiaohang Bian [11]. It was created in order to recognize handwritten mathematical expressions and translate them into Latex formulas. Additionally, a bi-directional mutual learning neural network (L2R and R2L) was used. That one-to-one learning method for decoding allows for knowledge transfer. But the two decoders do not learn in the same way. As a result, learning progresses slowly.

In this study, Sidney Bender [12] successfully captured the mathematical equation by utilizing the extremely strict syntax rule. The method that was used was to take the image after it had been taken, then to split the image math equation into a distinct image, and then use the fine-grained feature with that image to convert it to the Latex formula.

## III. METHODOLOGY

This section discusses the methods that authors used to implement the solution from mobile user interface into backend processing functions. User will capture the document with the aid of assistance from the mobile application and the captured document will be sent to the processing backend to extract and summarize the data from document image. Then the summarized data will be sent back to the mobile device to output the data as audio using text-to-speech engine.

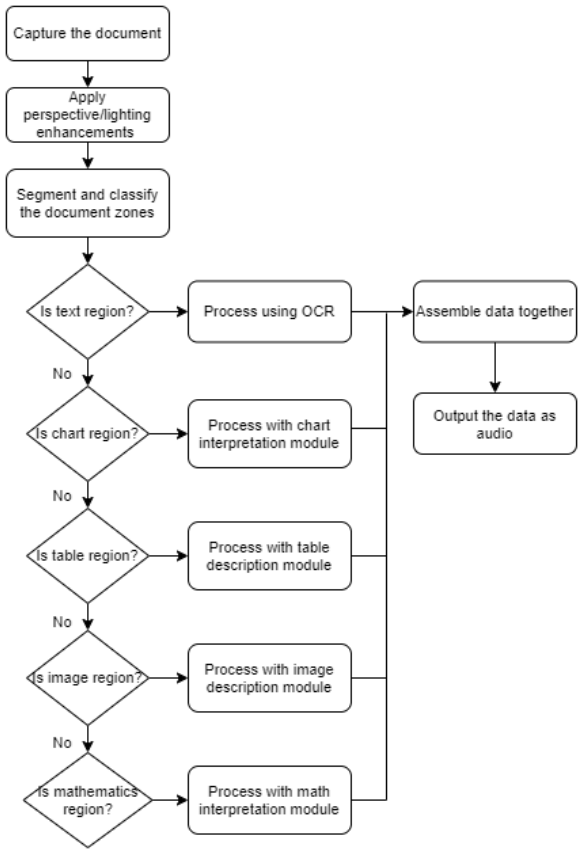


*Fig. 1. Flow of the document interpretation process*

### *A. Document capturing and segmentation*

For the implementation of document segmentation, there is a Document Image Analysis (DIA) library called LayoutParser[] which uses the state-of-the-art object detection algorithm Detectron2[] as the foundation for analyzing document images. LayoutParser has separate pre-trained deep learning models with different datasets designed for document layout analysis. In this implementation, the layout analysis is done by the ensemble learning method with multiple pre-trained LayoutParser models. Specifically, the models that were trained with PubLayNet and PRImA layout analysis datasets. The model trained with the PRImA dataset is capable of separately identifying mathematical regions and image regions while the PubLayNet dataset trains models to

identify general document areas like texts, titles and tables more accurately.

### *B. Chart Interpretation*

In this module, the first task is to classify and label the input chart. For that function, a Convolutional Neural Network following VGG-16 architecture is used. The optimizer used for the model is SGD with a 0.001 learning rate. The images are pre-processed to set a fixed size of 224 x 224 x 3 before sending through the network. The classified chart will be sent to chart-specific image processing algorithms to process.

All of the chart interpretation models are developed on top of the CRAFT(Character Region Awareness for Text Detection) [] model. For bar charts and line charts first, we identify the axes by detecting the longest horizontal and vertical lines. Then for the vertical and horizontal bar charts, we used connected component analysis to fit rectangles for each bar in the chart and calculate the values. Then we can identify corresponding labels using the CRAFT model.

For pie charts, we perform gradient analysis to identify the boundaries of the chart. After detecting the boundaries a calculation is done to calculate the volume of each slice. Then corresponding labels are identified using the Tesseract-OCR engine and the colours of the legend correspond to the slices.

### *C. Table comprehension*

There is no suitable arrangement that can read a table and convert it to an audio file for use by a customer who is blind when other technologies are taken into account.

This tool was developed using the VGG19 architecture after looking for an algorithm that trains on a dataset during the development process. We also went with that architecture because of how quickly it can be upgraded. There is a possibility that this will occur because the convolutional neural network of Vgg19 contains 19 layers and a version of the network that has already been trained on more than a million photos that can be loaded from the ImageNet database. Additionally, due to the size of the parameters being a huge number, there is greater precision when compared to other archaeologists.

This section's goals are to identify the table and assess its complexity. The data is then extracted from each table after that. In addition to VGG-19, we also employed other technologies like TensorFlow, which makes it simple to implement while building mobile apps. We applied it due to that. It is quite simple to identify the table both column- and row-wise because we used technologies like mathplot and boxplot to visualize the positioning of the position. Instinctually, it is simple to create an image that is incredibly accurate because we use an xml file to locate the image's table placement points and train the algorithm to recognize and correctly predict a typical image. Since the pytesseract library has more features and is a prime library, it makes it simple to analyze tabular data and can provide the output as a digital format of a table data CSV file.

The ICDAR-2017 POD dataset was produced as a result of the competition concerning detecting graphical Page Object Detection (POD) that was held in 2017. For tables, formulas, and figures in the dataset, bounding box information is provided. 1600 photos are used to fine-tune our network out of the 2417 images in the dataset, and 817 images are used as a test set. We offer our results with an IOU threshold value ranging from 0.5-0.9 in order to directly compare them with earlier approaches because the previous methods have reported results on various IoU thresholds.

We will utilize the Marmot and Marmot Extended datasets for Table Recognition to train our model. The authors of this paper made the data publicly available. With the Marmot dataset, we want to get the coordinates for the bounding boxes of the tables, and with the extended version of the same dataset, we want to gather the coordinates for the bounding boxes of the columns.

### *D. Image captioning*

It's difficult for persons who are blind or visually challenged to caption images in a printed document. It is necessary to extract the image's visual features with more detailed content and then need to generate captions which should be more catered towards vision impaired users by adding more descriptions in a way they can understand. So, the captured image of the printed document should be sent to a model to extract the features of the image. Then it will identify the main features with the respective sub-objects to describe the image efficiently. In this component, Convolutional Neural Network (CNN) is used to extract features of the image while a Recurrent Neural Network is converting data into natural language. So, a neural network with an encoder-decoder model is used in image extraction and captioning. Bahdanau Attention Mechanism consists of 3 Dense layers that detect the important features without any human supervision to make the captioning more accurate. CNN Encoder consists of a Dense layer, a single fully connected layer and the Recurrent Neural Network (RNN) Decoder consists of 4 convolutional layers.

To train the model, the MS-COCO dataset which is one of the most commonly used datasets for image captioning tasks are used. It is large-scale object detection, segmentation, and captioning dataset. The dataset includes over 82,000 images, each with at least five different caption annotations. When it comes to object detection and captioning, the MSCOCO dataset built by Microsoft has the purpose of providing the best

results possible. This dataset provides a collection of daily activities and their accompanying captions.

### *E. Math interpretation*

The user's input is read aloud by the flutter Text-to-Speech system in the mobile app, which then displays the user's response. Students with visual impairments all around the world have learned the ability to use programs, however, the majority of apps are unable to extract mathematical equations in order to interpret data. As a consequence of this, this tactic is helpful in the process of developing a mobile app to interpret mathematical problems.

The primary effort should be placed on recognizing the perfect math equation region and extracting the component in order to implement document math equation detection and reading of the discovered math equation. This will allow for the implementation of both of these features.

As a consequence of this, we made use of PubLayNet to segment the page and limits in order to extract features in order to isolate the mathematical component. It trained with neural networks called Faster RCNN and M-RCNN, and it pre-trained the model to use the segment to the document data. Identification of the text, title, table, figure, and list were the primary focuses of this activity. And in this instance, it is used to train for the recognition of equation components. After the equation has been bounded, the equation component's coordinates should be retrieved, and then the equation image should be cropped such that it is distinct from the math section. Following the use of a distinct image, the optical character recognition (OCR) module that is included in Pytesseract is utilized in order to extract the mathematical equation in the form of text.

The mobile app makes use of the flutter Text-to-Speech technology in order to read and output information to the user of the app.

## IV. RESULT AND DISCUSSION

This study is mainly focused on building an application to read document content for vision-impaired people. The application consists of 5 main components like Document capturing and segmentation, Chart interpretation, Table comprehension, Image captioning and Math interpretation. As a result of this application, the time it takes to capture an image is greatly reduced by using autofocusing and image capturing techniques and it is featured with a smart assistant which provides audio assistance to navigate through the application.

The proposed approach can capture a printed document or a book and segment it into four main components like charts, images, tables and mathematical equations. In the document segmentation, it detects the objects in the document and analyses the document images into the above components. This can label the charts as pie charts, bar charts, or line graphs and read the chart's informative content, which even a user with a print disability can comprehend. If the user captures a table using the mobile app, it can extract data from the table and read it by rows and columns, as well as describe the data inside the table descriptively. It can also even identify a complex mathematical equation and read it to the vision impaired person in a way they can understand. Moreover, this developed application can identify any image rather than a chart, table or an equation and extracts the main features of it and explain the image in the best way possible for the user to understand it. As a combination of the above-mentioned components, our solution gives the visually impaired individuals to have access to a reliable, autonomous and an accurate document reading service.

## V. CONCLUSION

Due to a variety of factors, a large percentage of the world's population is unable to interpret printed materials in a normal way. This inaccessibility can be caused by vision impairments, physical dexterity problems, and learning and literacy difficulties. Print disability prevents a person from obtaining information from printed material in the traditional manner and necessitates the use of alternative methods of accessing the information. These disabilities can have an impact on people's daily lives as well as their education and literacy. Materials such as legal and personal documents are rarely available in accessible mediums, and it is more difficult and not suitable to obtain the assistance of a third party to interpret such documents.

There are still many obstacles that cannot be addressed by using accessible mediums that are currently available for vision-impaired people. For instance, images, mathematics, tables, and graphs are hard to interpret using methods like braille and most of the available document scanners and OCR applications do not have accessibility options for vision impaired people. Braille is also not accessible to every single person due to braille documents being expensive and also it is not guaranteed to have a braille version for every document. So, there is a significant gap in reading rights and equality between print disabled people and the general population.

As a solution to this problem, we have developed a mobile application to scan and interpret the content inside the printed materials with multiple accessibility options to cater for every possible user. For better accessibility “ABCD Name” is implemented with a voice assistant to help the user to capture documents and to navigate through the application. The UIs are designed with suitable fonts and buttons which makes it easy for visually impaired users to use the application. This application is mainly focused on converting the captured image of the printed

documents or books into digitalized text and reading the content to the user. The developed solution supports text, mathematical equation, table data, chart data and image interpretation. For future work, this application can be further developed to read documents with extremely small fonts, hand-written documents and complex structured documents like newspapers.

## References


[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. He, Y. Lu, and S. Chen, "Image captioning algorithm based on multi-branch CNN and bi-LSTM," *IEICE Trans. Inf. Syst.*, vol. E104.D, no. 7, pp. 941–947, 2021.

[8] B. Makav and V. Kilic, "A new image captioning approach for visually impaired people," in *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)*, 2019.

[9] P. Shah, V. Bakrola, and S. Pati, "Image captioning using deep neural architectures," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1–4.

[10] R. Noor, K. Mejbaul Islam and M. J. Rahimi, "Handwritten Bangla Numeral Recognition Using Ensembling of Convolutional Neural Network," 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018, pp. 1-6, doi: 10.1109/ICCITECHN.2018.8631944.

[11] Bian, X., Qin, B., Xin, X., Li, J., Su, X., & Wang, Y. (2021). Handwritten mathematical expression recognition via attention aggregation based Bi-directional mutual learning. In arXiv [cs.CV]. http://arxiv.org/abs/2112.03603

[12] S. Bender, M. Haurilet, A. Roitberg and R. Stiefelhagen, "Learning Fine-Grained Image Representations for Mathematical Expression Recognition," 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), 2019, pp. 56-61, doi: 10.1109/ICDARW.2019.00015.