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

Project Id: 2022-024

B.Sc. (Hons) Degree in Information Technology
(Specialization in Software Engineering)

Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Sri Lanka

October 2022

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

Project Id: 2022-024

The dissertation was submitted in partial fulfilment of the requirements
for the B.Sc. Special Honors degree in Information Technology (Specialization in
Software Engineering)

Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Sri Lanka

October

DECLARATION

We declare that this is our own work, and this report 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

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor Date

………………………….. ………….…………

(Dr. Anuradha Jayakody)

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 them 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, Vision impairment, Computer vision, Image processing, Document image analysis

ACKNOWLEDGEMENT

We'd like to begin by acknowledging our supervisor, Dr. Anuradha Jayakody, for his guidance, encouragement, and commitment during the research endeavour. Second, we want to thank our co-supervisor, Ms. Shashika Lokuliyana, for her devotion to us, her encouragement, and the information she has provided to us. Without them, this research project will not be fruitful.

We also want to sincerely thank our families, who have always supported us and inspired us to persevere in the face of difficulty. Without their aid, none of these will be successful.

Finally, we’d like to thank the members of our group for their assistance throughout the study project and the four years we spent in college. They gave us the strength to face any challenge as though they were our own brothers and sisters. Also, thank you for bringing life to our university experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
CNN	Convolutional Neural Network
OCR	Optical character recognition
SVM	Support Vector Machines
CRAFT	Character Region Awareness For Text Detection
API	Application Programming Interface

1. INTRODUCTION

1.1. Background & Literature Survey

Vision, one of the most important of our senses, is essential in every aspect of our lives. Without that, it is difficult for us to interact with the rest of the world on a daily basis. Even for individuals without any vision impairments, there is a chance of getting a vision condition when they grow old. According to current statistics, WHO estimates there are at least 2.2 billion people with near or distance vision impairments [1]. 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.

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, the authors 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.

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 lack functions like table interpretation and mathematical equation interpretation [6],[7]. Furthermore, most of these solutions do not 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 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.

In the following section, the literature reviews on each major component of the system will be discussed.

1.2.1. Document capturing and segmentation

This research component which is the document zone and content classification also already been researched and many papers can be found on the topic. However, most of them lack the ability to detect and classify all the necessary content types that could be included in a printed material [8],[9].

Andrea Corbelli et al. [10](Research B) 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. [8](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 the [9](Research C) document layout analysis can even be done by using one-dimensional convolutional neural networks which are 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 [9] 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.

Table 1.2.1 - 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

1.2.2. 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 [11] (Research P).

As demonstrated in [11], 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 lack interpretation of multiple charts [12] (Research Q) and even though they supported multiple chart interpretations they lack the text description which is needed for this research [13] (Research R).

Table 1.2.2 - Research gap for chart classification and interpretation with existing systems

	Research P	Research	Research	Proposing

		Q	R	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
Match extracted values into chart labels	No	No	No	Yes
Provide a textual description of chart data in plain English	Yes	No	No	Yes

1.2.3. 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 gives 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. [14] 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. [15] 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. [16] published an analysis of table recognition in document images with deep neural networks. It is supposed to recognize the most frequent table formats and it should be able to recognize them in the text as well as audio, video and audio format.

1.2.4. 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, has also 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.

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 [17]. Some tools don't even describe the basic colors in it [18], [19]. Since the "Image captioning algorithm based on multi-branched CNN and Bi-LSTM" paper [18] 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 a sufficient explanation for a blind person to understand an image in a printed document.

1.2.5. Math interpretation

Many mobile applications are available to blind people, which did not give perfect output to the user. Therefore, blind people face many difficulties. When students use the app and they have to study math and the app gives an incorrect output, they need to get the help of non-blind people to solve the problem. Because in math, if one letter is unrecognized or missing, the user is given a completely different formula.

Many researchers have converted mathematics equation into LATEX output. But the problem is that another separate application is needed to read the LATEX output. And many applications are computer based. So blind people cannot do it alone or have to get help from non-blind people.

The purpose of the research is to upload or capture the equation image into a mobile application so that blind or visually impaired individuals can understand it in mathematical form.

Rouhan Noor [20] research paper shows Bengali numerals recognized by handwritten letters. In research, the captured image is converted to RGB and then grayscaled to reshape the image. Then the developers use two separate models of CNN to get the best model performance. Those two models use 7 and 5 convolutional layers. And this research does not go beyond the identified numerical figures. Sidney Bender [21] research paper They developed Fine-Grained Feature Extractor (FGFE) model to recognize formula image and convert it to LATEX formula. FGFE's CNN block consists of 2-4 convolution layers and maximum polling. They used the token to match the exact equation and used BLEU to estimate the distance between sentences and they used the beam search option to increase the match to the correct token. This research has found that formulas cannot be clearly recognized in parentheses and the recognition accuracy of long formulas is low. Xiaohang Bian [22] improved the identified handwritten mathematical expression and proposed the Attention aggregation based Bi-Directional Mutual Learning Network (ABM) model based on attention aggregation. The model was trained on both opposite sides (L2R, R2L) image LATEX sequences. Then it learns both branches of the model together. The problem they faced in this research was that the R2L branch did not provide perfect output for learning. So they get less accuracy.

1.3. Research Problem

According to the current statistics, approximately more than 285 million of the world population can be identified as visually impaired or blind [1]. Not only the blind or vision impaired people but also those who have a lower literacy might have problems with reading paper documents as well. Also, there are many other reasons from which an individual can become print disabled [2]. So, considering the huge percentage of the population which has print disability there must be more accessible tools to aid them in their day-to-day life. Over decades people have been trying to come up with new techniques and solutions to help such people in need. Even though visually disabled people can read material in the braille method, not every document or text can be found in braille form. It can be difficult for them to read printed material by themselves. There are existing methods to identify the pictures in the printed document and some other paragraphs and sentences. But these methods are still research level and there is no application which is accessible for the public. And there are some complex items such as tables, graphs and mathematical equations which cannot be identified by existing solutions [3].

On many occasions these people have to get the assistance of another person in reading a printed material. But there can be many disadvantages in handing over documents like legal or confidential documents to another party for the assistance which can be unreliable. Until now there is no proper digital method that included all the technologies such as table reading, equations reading, image describing within one tool. Most of the tools do not even contain user friendly guidance catered to disabled users to guide the user. As people with print disabilities may have a hard time using a mobile interface or directing a camera at a selected text, the solution must come with proper guidance.

2. RESEARCH OBJECTIVES

2.1. Main Objective

Designing an assistive application to aid individuals with print disabilities to interpret printed materials.

2.2. Specific Objectives

Sub Objective 1: Scan and classify the content in the printed material using a proper algorithm and direct the relevant portions of the image to suitable interpretation algorithms.

- The printed material should be divided by its content in order to interpret using suitable algorithms.
- The scanning process of the material should be assisted properly for every user.

Sub Objective 2: Develop proper table comprehension and diagram description model within the system.

- The solution should be able to properly interpret the tables, diagrams within the scanned document.
- Interpretations of the tables and diagrams should be easy to understand and should be in plain simple English.

Sub Objective 3: Develop an algorithm to explain text and complex mathematical expressions to the user.

- The solution should be able to properly explain the mathematical equations included in the scanned material.
- Texts and paragraphs should be properly explained as well.

Sub Objective 4: Develop the system with proper image interpretation model.

- There should be a system to describe images in the printed material in plain English in an understandable manner.
- For users who don’t know the meaning of the colour names, the colours will be explained in another emotional manner. (People who have never seen colours).

3. METHODOLOGY

3.1. Methodology

This digital assistant could be a solution for the individuals with print disabilities to interpret with printed materials. A normal person can use their eyes to interpret the materials. But for a print disabled person, without others' assistance, they cannot see or identify anything by themselves. Other than receiving help from a third party, the only way to overcome the above difficulties is by listening to audio assistance.

When developing an assistive application, especially designing the user interface accessible for a vision impaired person, is a very serious issue and in this application, it has a proper interface for every type of user with audio guidance. This assistive application can provide vision disabled people with distinguishing images and describing them, identifying paragraphs, texts, identifying tables, diagrams and recognizing mathematical equations would be an uncomplicated task using this assistant. In this solution, the user has to point the camera into the material which he wants to interpret, and this process will also be voice assisted. Then the image will be automatically captured when the printed material is within the proper focus range. After successfully capturing the material, the system will automatically identify the different parts of the material like text, equations, tables, and images. Then the solution will be able to interpret the different regions of the printed material using appropriate algorithms. There are already a lot of apps to describe images on a basic level, but this digital assistant can provide better solutions by giving the information more accurately. Other than that, in image identification, the digital assistant will also use surrounding text to provide accurate information.

Reading table data and reading mathematical equations are also the functionalities that this digital assistant can provide. It can analyze the table data and identify the equations and provide precise information to the print disabled individual via voice assistance. The vision impaired person is guided by the audio in the app. It provides guidance by voice commands to capture the material precisely in a correct angle.

The assistive application is embedded with the above-mentioned functionalities which is going to be developed as a solution for the vision impairment of the print disabled people.

By this, people who have visual disabilities and unable to interpret printed materials will be able to coexist in the society without feeling dependent on others too much. This digital assistance will provide them a relief as well as a better understanding about the surroundings they live in.

3.1.1. System Architecture

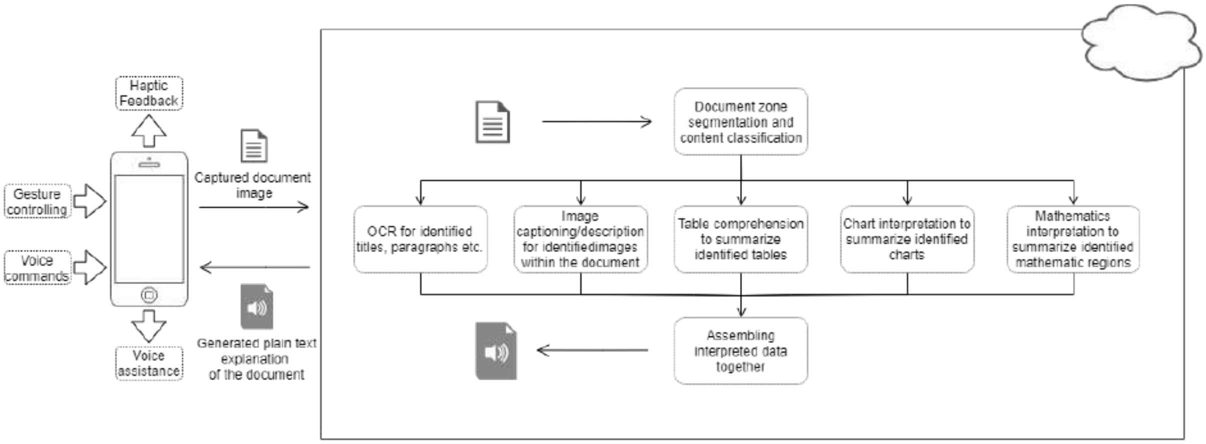

Haptic Feedback
Gesture controlling
Voice commands
Voice assistance
Captured document image
Generated plain text explanation of the document
Document zone segmentation and content classification
OCR for identified titles, paragraphs etc.
Image captioning/description for identifiedimages within the document
Table comprehension to summarize identified tables
Chart interpretation to summarize identified charts
Mathematics interpretation to summarize identified mathematic regions
Assembling interpreted data together

Figure 3.1.1 - High level system architecture

The overall system resides in two main layers as a mobile application and a backend Django server. Document capturing and voice assistance is implemented in mobile application and computing heavy image processing components resides in the backend server since the mobile processors cannot handle heavy workloads like CNNs and image processing algorithms.

3.1.2. Data collection methods

Collecting and analyzing the requirements for the study is a very important task throughout the whole research. So, we collected information by conducting 4 surveys and distributed them among people who have and who are communicating with blind people. Each survey was under the four main categories of this research to get how their needs vary when interpreting those images, equations, charts and tables. We must be especially concerned about the information gathered because it must be relevant to our research area. Prior to beginning implementation, it is important to gather the necessary requirements in order to determine whether the proposed functionality will provide the appropriate solutions. So, to gather needs, we use the approaches outlined below.

- Read research papers and reports that are related to the research we do.
- Do research on the existing and proposed systems similar to ours
- Talk with the supervisor and other expertise who have knowledge in our research area
- Conducted a survey to collect information related to the 4 components as well as for the whole system
- Research more when choosing the correct algorithms and dataset before starting the implementation

Since this research is mainly under image processing, we have to choose the correct datasets that are required to train the algorithms. For the chart type classification, there was a need for a dataset to be found to train the classifier. However, because there was no dataset that could be found, we collected numerous chart images from web sources, totaling approximately 7000 different chart images. To preserve consistency, 4000 of those images are utilized to train the model, with about 1000 images in the training dataset for each chart style (Horizontal bar, Vertical bar, Pie, Line). For image interpretation, after researching and going through the past studies we chose MSCOCO and the CelebFaces Attributes (CelebA) as the best dataset to train the models. We selected the Kaggle's CelebA dataset to recognize the facial attributes of humans. It consists of 202,599 face images of various celebrities, 10,177 unique identities and 40 binary attribute annotations per image.

3.1.3. Tools and Technologies

- Flutter/Dart
- Python
- TensorFlow/Keras
- OpenCV
- PyTesseract
- InceptionV3
- LayoutParser
- Detectron2
- YOLO
- Django

3.2. Commercialization aspects of the product

In this research project application, our main goal of this application is to avoid the reading dependency of people who are with print disabilities. From that we expect every blind person or person who has print disability issues to use this application and they will have a pleasant experience with using this application. As a result of this application, blind people will not find any third person to understand those printed documents. At present people will be facing some trouble when they cannot even read a physical document. When a person receives any private or legal document, they cannot keep their privacy. With these issues this is the right product to use.

Introducing this application to the community this application can subscribe to any person all around the world and especially since we are focusing on the disabled community. With this aspect, we are hoping to offer this product for free to use for selected groups who are with disabilities. And monetization will be collected by founding partnerships for the applications. Also, in the current time, Non-Government Organizations also provide funds to further improve this application.

In different disabled communities, there are a lot of projects running for these disabled people. Some of them are Kurzweil Education platforms, Refreshable Braille Displays, and Ultra Cane etc. With these platforms, we can promote this product as a new trending application especially for blind people. In the current period, people are obsessed with social media platforms and from that level, we can promote the application and we can get partnerships by demanding these applications by promoting the usefulness of this application to people who are with disabilities in our society.

The entire cost of installing the respiration rate measurement component is as follows:

Table 3.2.1 - Budget of the study

Item	Cost (LKR)
App Publishing cost on google play	12500.00
Backend Hosting Cost	14800.00
Paper Publishing Cost	47500.00
Total	74800.00

3.3. Testing and Implementation

3.3.1. Implementation

This section discusses the methods that the authors used to implement the solution from the mobile user interface into backend processing functions. The 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 the document image. Then the summarized data will be sent back to the mobile device to output the data as audio using a text-to-speech engine.

To assist every possible user, the mobile application is developed with assisted document focusing and automated document capturing. These assistive functionalities are developed using OpenCV native libraries for edge detection and using a separate TensorFlow Lite in-device model for object detection to identify documents.

The UI of the application is designed with OpenDyslexic font and features bigger buttons suited for vision-impaired users. Users will get audio assistance on where to touch or do a certain action in order to perform certain tasks. Furthermore, the UI is designed with colorblind users in mind.

In the following section, the methodology for each major component will be discussed separately.

A. Assistive document capturing

This module resides in the frontend part of the system within the mobile application. Mobile application is mainly developed with the flutter framework. To identify the document when the user opens the camera author had to use a real-time image processing method. Due to this reason, the image processing component couldn't be moved into the backend because of the network latencies the output will not be real-time. So, the author was restricted to using something lightweight that the mobile phone processing units can handle. TensorFlow lite is a really good solution for this, and it works well with the flutter framework however the technology is not matured yet and running object detection models like YOLO or SSD as TensorFlow lite models will still be heavy for the majority of mobile processors in the current mobile phone market. Ultimately, the decision was to implement document detection with edge detection methods. This can be done with the use of OpenCV native libraries for android. Since with flutter, there is the ability to manipulate native source code to implement and interact with native APIs like camera APIs, the processing will be faster.

The process of edge detection consists of image enhancement, contour detection and deciding the corners using the detected contours. All this processing is applied to the camera image in real-time.

B. Document segmentation and content classification

For the implementation of document segmentation, there is a Document Image Analysis (DIA) library called LayoutParser [23] which uses the state-of-the-art object detection algorithm Detectron2 [24] 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.

C. 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.

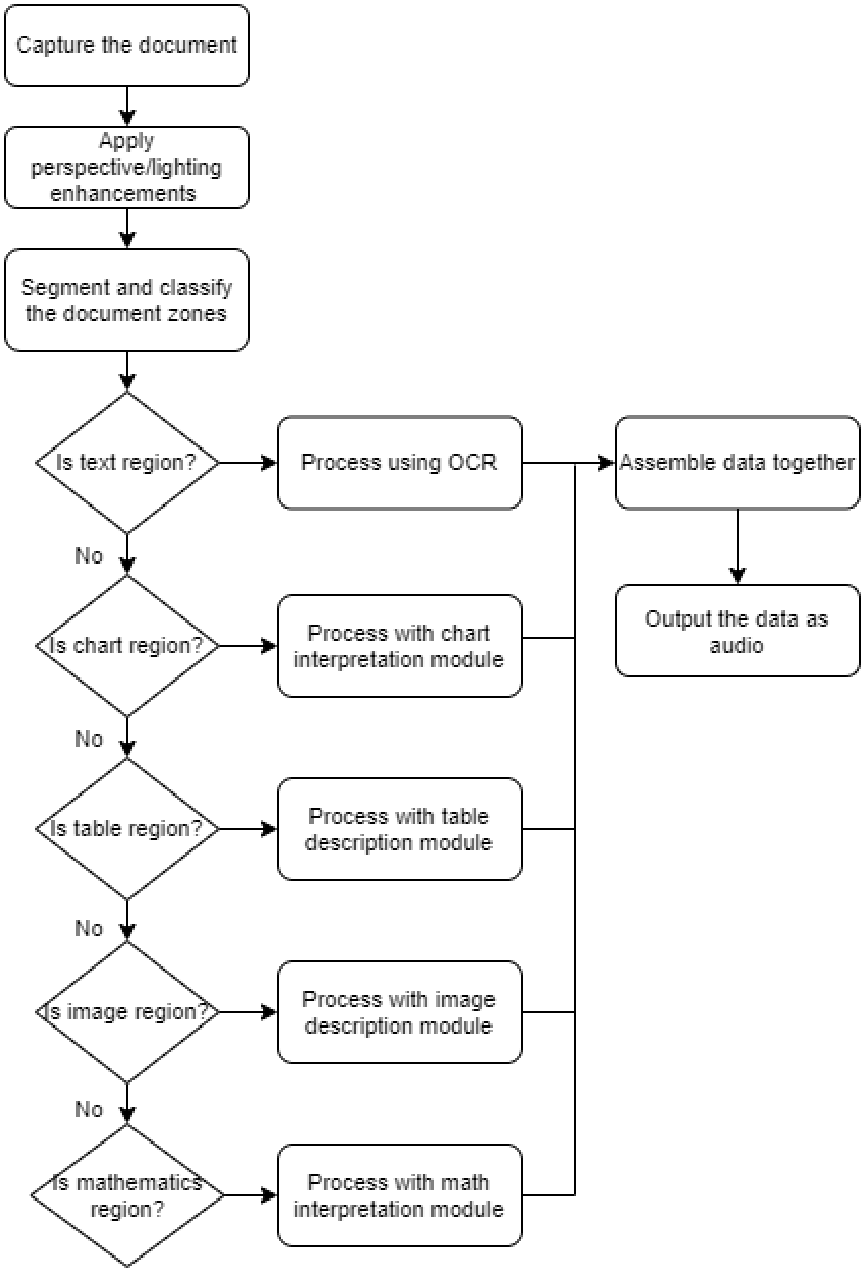
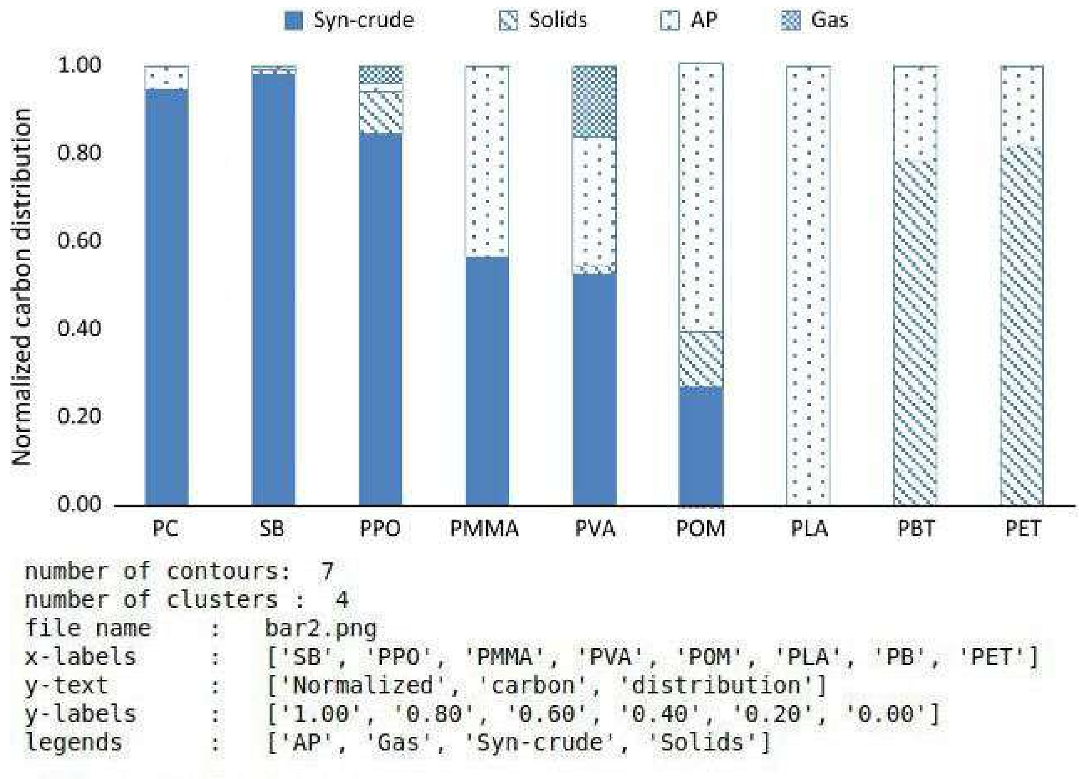

Capture the document
Apply perspective/lighting enhancements
Segment and classify the document zones
Is text region?
Process using OCR
Assemble data together
Output the data as audio
No
Is chart region?
Process with chart interpretation module
No
Is table region?
Process with table description module
No
Is image region?
Process with image description module
No
Is mathematics region?
Process with math interpretation module

Figure 3.3.1 - Overall data flow of the system

All of the chart interpretation models are developed on top of the CRAFT (Character Region Awareness for Text Detection) [25] 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 [26] engine and the colors of the legend correspond to the slices.

After extracting the data, the plain English sentences are generated using a general template-based sentence generation method.



```
number of contours:  7
number of clusters :  4
file name     :   bar2.png
x-labels      :   ['SB', 'PPO', 'PMMA', 'PVA', 'POM', 'PLA', 'PB', 'PET']
y-text        :   ['Normalized', 'carbon', 'distribution']
y-labels      :   ['1.00', '0.80', '0.60', '0.40', '0.20', '0.00']
legends       :   ['AP', 'Gas', 'Syn-crude', 'Solids']
```

Figure 3.3.2 - Extracting labels from a bar chart

D. Table comprehension

This tool examines a table in a printed document to discover how the x-y axes are ordered and the contents of the table. The data in the table's rows and columns are then recognized. It then digitally stores the discovered data in a CSV file and utilizes a summarizing model to inform the user of the table's creation date. This technique mostly uses open CV libraries to split the tables, as well as evolution-based techniques to transform a digital picture into a binary representation and employ filters to attempt to precisely identify the table. In addition, X-Y coordinates are used to examine the table's threshold values, and the Pytesseact [26] package is used to read the image's text. The data were clustered using the SK-learn data clustering process. After completing the entire process at the conclusion of the procedure, the result is summarized and sent to the user as an audio file.

In essence, this tool examines a printed table to identify how the x-y axes are set up and what is on the table. The information in the table's rows and columns is then recognized. Then, it digitally stores the recognized data in a CSV file and utilizes a summarizing model to inform the user of the table's date.

In this technique, the tables are mostly separated using open CV libraries, and the conversion of a digital picture to a binary format and usage of filters to attempt to clearly identify the table are more evolution-based approaches. Additionally, the Pytesseact [26] library is used to read the text from the picture, and the threshold values of the table are validated using X-Y coordination. The data were clustered using the SK-learn data clustering algorithm, after all. After completing the entire procedure, the process is concluded, and the user is provided with the output in the form of an audio file.

E. Image captioning

In the image captioning section, it is necessary to extract the image’s visual features with more detailed content and then generate captions which should be more catered towards vision impaired users by adding more descriptions in a way they can understand. So, this component is built with the use of various models and is mainly divided into two models like caption model and the object detection model.

First, the caption model does the caption generation after getting an input image. 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. To train the model, the MS-COCO [27] dataset is used. It is large-scale object detection, segmentation, and captioning dataset.

Then secondly, by using YOLO object detection, a model has been developed for object recognition. It will identify each object with the respective sub-objects in the image separately and then divide them into two groups as human and non-human attributes. After identifying humans, again it is gone through one multi-model which is a combination of five different models that gives five outputs to identify the main attributes of a human which will output the emotions and attributes of the human subject. Finally, through a template model, it generates a detailed description of the image in a way a blind person can understand.

F. Math interpretation

- Image preprocessing

Utilizing the generated dataset, transform photos using the albumenation tool in Python. Change the image's brightness, contrast, background color, and Gaussian noise. It increases image recognition speed while using less computing power.

- Model

Created model has neural network architecture it shown in Figure 4.2. it has the encoder decoder architecture and it used to ResNet encode the image and transform the encoded image to text. In here use the image-to-sequence and sequence-to-sequence models.

- Encoder

This encoder makes use of CNN in order to extract a 2D feature map from the picture that is fed into it. The image is encoded through the use of the Resnet architecture. After that, the feature-map is projected such that it is consistent with the transformer; after that, a 2D spatial encoding is applied; and finally, the sequence is flattened to be 1D. The 2D positional coding is a sinusoidal coding that is fixed.

- Decoder

The decoder is a Transformer stack with non-casual attention to the encoder output and causal self-attention. As is standard, training is done with teacher forcing, which means that the ground truth text input is shifted one off from the output.

It is common knowledge that utilizing input vectors is preferred when using 1D function coding. In addition to this, we link the line number encoding (LNE), which is the scalar text line number where the token is located, to it. In order to address issues with line-level errors, we incorporated line variety coding into the model. We have not but formally concluded what impact it has had on the performance of the version, and we only mention it here for the sake of completeness.

Combining Vision and Language styles is one of our organization's advantages, and it's also one of its strengths. When it comes to the processing of photographs, CNNs like ResNet are widely regarded as among the best available options. And transformers are ideally suited for language modeling and natural language processing tasks because they have properties that are exceptionally helpful in the management of noisy and imperfect characters, the likes of which are frequently found in published materials. If I had both the visible characteristic map and the linguistic version of the version, I would be able to do a better job with it than if I had just relied on the visual features by themselves.

It utilizes a decoding method that is simple to grasp and chooses the most promising possibility token at each stage. The fact that beam search decoding does not presently provide an improvement in accuracy indicates that the version is actually biased or dependable.

Make up your own symbolic vocabulary with the help of latex formulas. The clause can be reworded so that it corresponds with the token. After the model has been executed and the Latex output has been returned, it is then possible to make use of the built-in expressions in order to transform the Latex output into the mathematical path.

After the completing the model, it add to python based framework to the connect with mobile application. Use the Django framework to the backend and connect. After the connection to the backend if the user upload or the capture the image then image will send to the backend and run through the backend and after getting the latex output then it converting to the mathematical readable text. Finally readable mathematical pass to the application and read it to the user.

3.3.2. Testing

Testing is an essential component both before and after the application is released to its users. The importance of a well-structured testing plan cannot be overstated. Unit testing, integration testing, benchmark testing, and end-user testing all directly address issues with the program.

The testing technique for the implemented app addressed all functional and non-functional requirements. Initially, the system components were tested separately using unit testing. Following the completion of the solution, implementation and end-user testing are performed.

In this section, we will briefly discuss the results on the overall integrations tests as the testing is covered in detail in the individual reports.

Table 3.3.1 - Test results of the application

Test Case #	**Test case**	**Result**
001	Mobile app launches from the android menu	Pass
002	All the buttons and widgets are visible	Pass
003	App navigates through pages with the buttons	Pass
004	App navigates through pages with the gestures	Pass
005	Camera launches when prompted	Pass
006	Document is captured with the shutter command	Pass
007	Captured document is sent to the server	Pass
008	Text-to-speech output is generated for the document data	Pass

4. RESULTS AND DISCUSSION

4.1. Results

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 features a smart assistant which provides audio assistance to navigate through the application.

The proposed approach can capture a printed document or a book and automatically segment it into five main components as text, charts, images, tables and mathematical equations. Which takes off the issue where the print disabled individual has to identify the regions within the document and scan only the identified regions to feed relevant algorithms like image captioning or chart interpretation. If there is a table in the captured document, relevant algorithms can extract data from the table and read it by rows and columns, as well as describe the data inside the table descriptively.

- Results of assistive document capturing

  The main functionality of this component is to detect the document using real-time edge detection and assist the user with capturing and cropping process. This component is implemented in the mobile application and consists of UI's which are catered to be used by visually impaired users. Furthermore, the edge detection model which resides within the camera application successfully detects the document edges given that there is enough lighting within the scene.

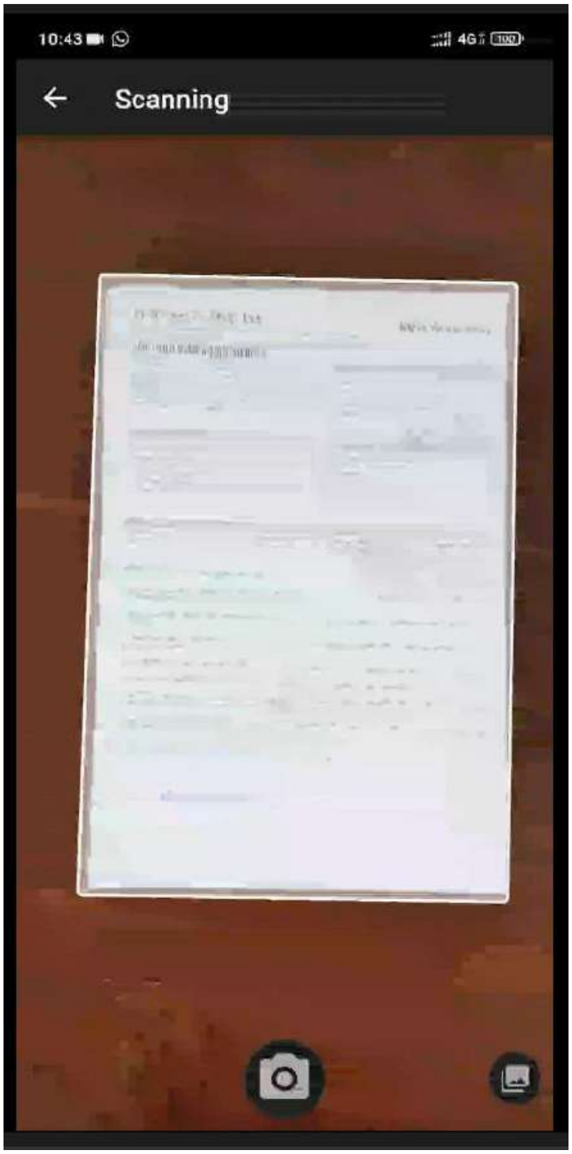


Figure 4.1.1 - Real-time edge detection results shown as overlay

The detected corners of the document are automatically selected for cropping in the crop window of the application. Users can do smaller adjustments for the crop or users can make the cropping process to be automatic because the application will automatically detect the exact points to crop the document from the overall image.

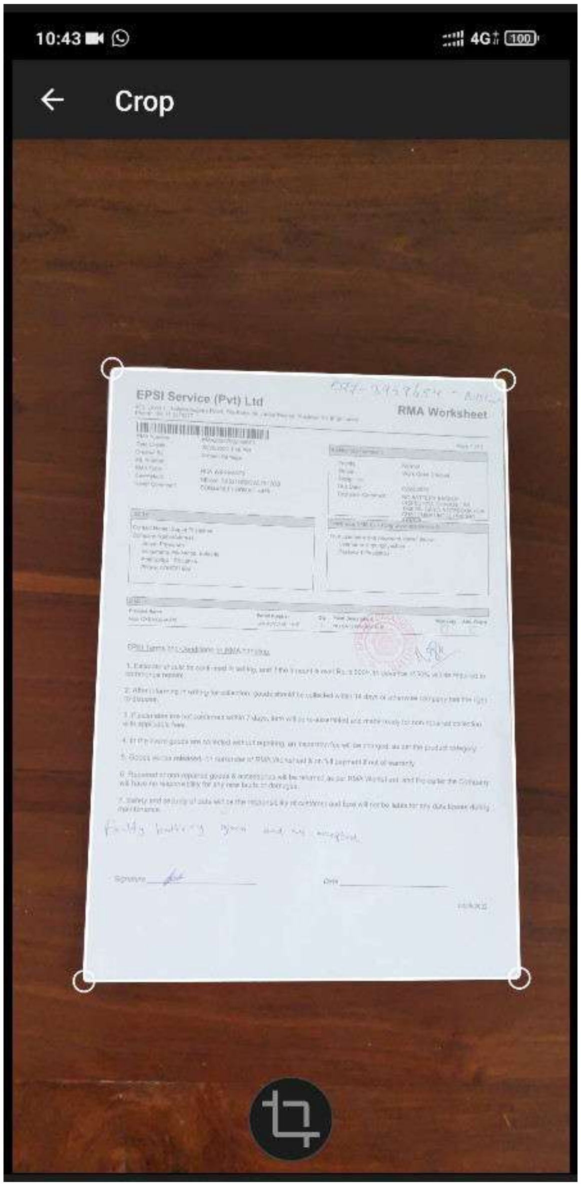
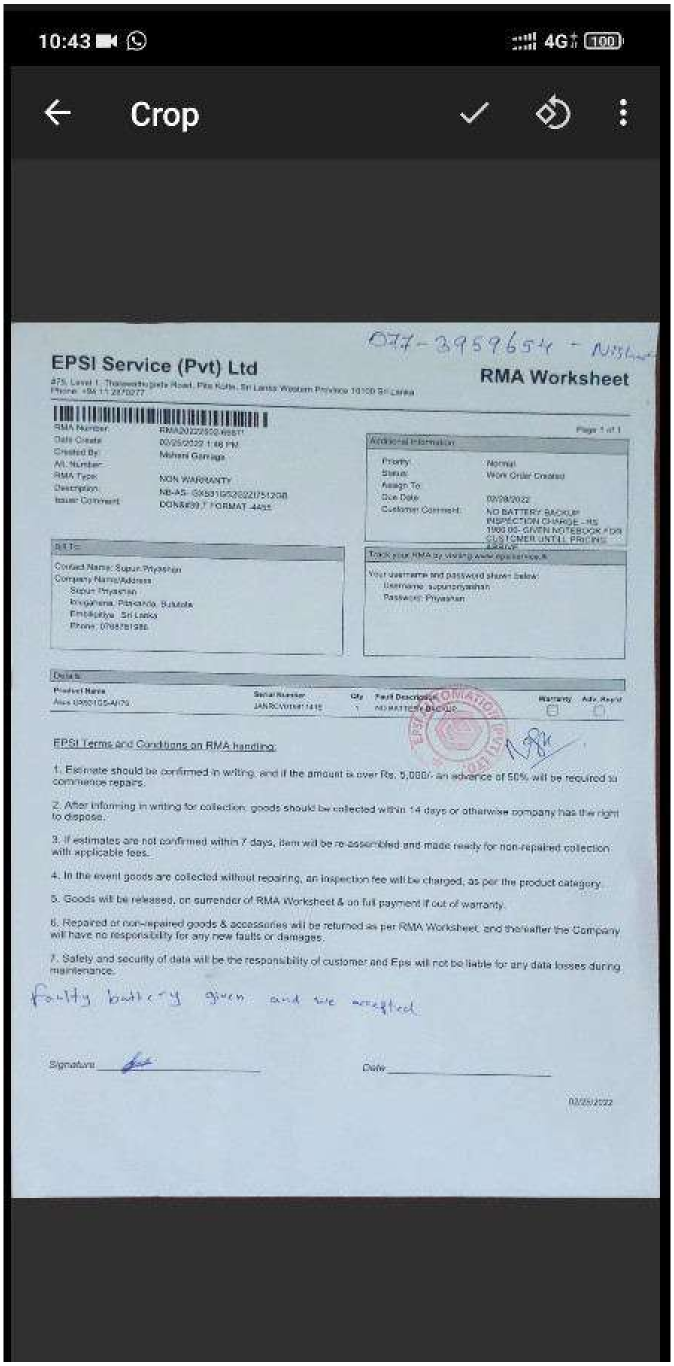


Figure 4.1.2 - Detected edges are automatically selected for cropping

After the cropping process, the final image can be shown to the user and will be sent to the backend server for interpretation or the user has the option to re-adjust the cropping of the image.



077-3959654 - Nish

EPSI Service (Pvt) Ltd
475, Level 1, Thalawathugoda Road, Pita Kotte, Sri Lanka Western Province 10100 Sri Lanka
Phone: +94 11 2870277

RMA Worksheet

Page 1 of 1

RMA Number:	RMA20222502-46871
Date Create:	02/25/2022 1:48 PM
Created By:	Mihani Gamage
Alt. Number:	
RMA Type:	NON WARRANTY
Description:	NB-AS- GX501GS2022I7512GB
Issuer Comment:	DON&#39;T FORMAT -4495

Additional Information

Priority:	Normal
Status:	Work Order Created
Assign To:	
Due Date:	02/28/2022
Customer Comment:	NO BATTERY BACKUP INSPECTION CHARGE - RS 1990.00- GIVEN NOTEBOOK FOR CUSTOMER UNTILL PRICING APPRV

Bill To:

Contact Name: Supun Priyashan
Company Name/Address:
Supun Priyashan
Kirigahena, Pitakanda, Bulutota
Embilipitiya, Sri Lanka
Phone: 0768781986

Track your RMA by visiting www.epsservice.lk

Your username and password shown below:
Username: supunpriyashan
Password: Priyashan

Details

Product Name	Serial Number	Qty	Fault Description	Warranty	Adv. Req'd
Asus GX501GS-AI76	JANRCV01[illegible]1415	1	NO BATTERY BACKUP	□	□

EPSI Terms and Conditions on RMA handling:

1. Estimate should be confirmed in writing, and if the amount is over Rs. 5,000/- an advance of 50% will be required to commence repairs.

2. After informing in writing for collection, goods should be collected within 14 days or otherwise company has the right to dispose.

3. If estimates are not confirmed within 7 days, item will be re-assembled and made ready for non-repaired collection with applicable fees.

4. In the event goods are collected without repairing, an inspection fee will be charged, as per the product category.

5. Goods will be released, on surrender of RMA Worksheet & on full payment if out of warranty.

6. Repaired or non-repaired goods & accessories will be returned as per RMA Worksheet, and thereafter the Company will have no responsibility for any new faults or damages.

7. Safety and security of data will be the responsibility of customer and Eps will not be liable for any data losses during maintenance.

Faulty battery given and we accepted

Signature ______________ Date ______________

02/25/2022

Figure 4.1.3 - Final cropped result

- Results of the document segmentation module

  Since this component is developed on a state-of-the-art object detection model (Detectron2) and pre-trained with huge datasets with hundreds of thousands of images the accuracy of the component is very high. To be specific, the model trained with the PubLayNet dataset includes 360 hundred thousand images and the size is about 102GB. So, the results are highly accurate and the model is very reliable.

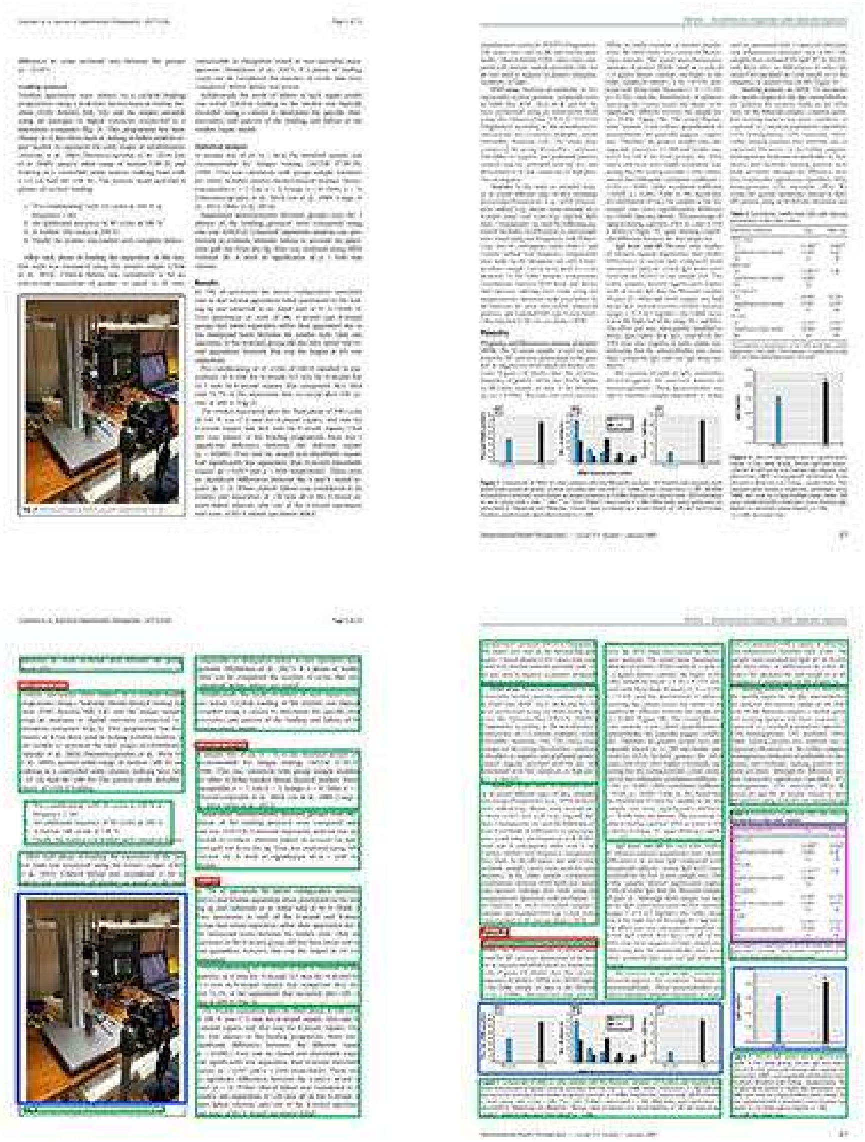


Figure 4.1.4 - Results of the document segmentation algorithm

- Results of the chart classification model

  This component consists of a custom-defined modified version of a VGG-16-like convolutional neural network for the classification of the images. The dataset is also collected by the author and the model is trained from scratch. The model has shown ~99% accuracy within the training dataset as well as the validation dataset.

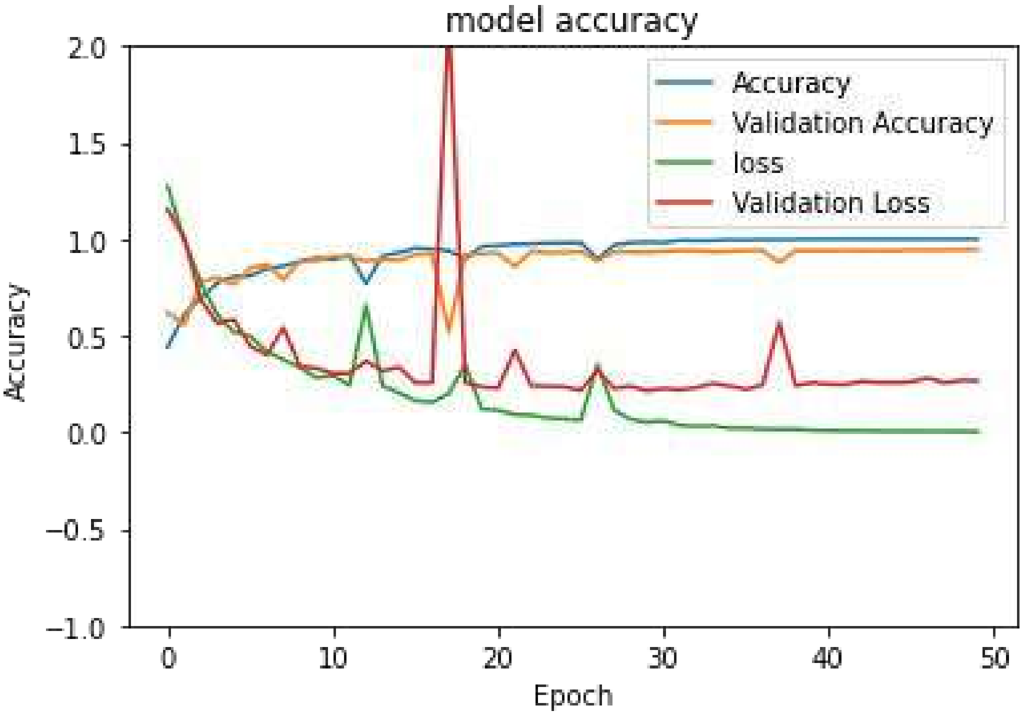


Figure 4.1.5 - Training and validation accuracy of chart classification model

Furthermore, the confusion matrix for the model results for the testing (consists of 1000 chart images) dataset shows that pie charts are classified with ~99% accuracy while horizontal bar charts are classified with ~100% accuracy. Also, the vertical bar charts and line charts can be classified with ~96% accuracy with the trained model.

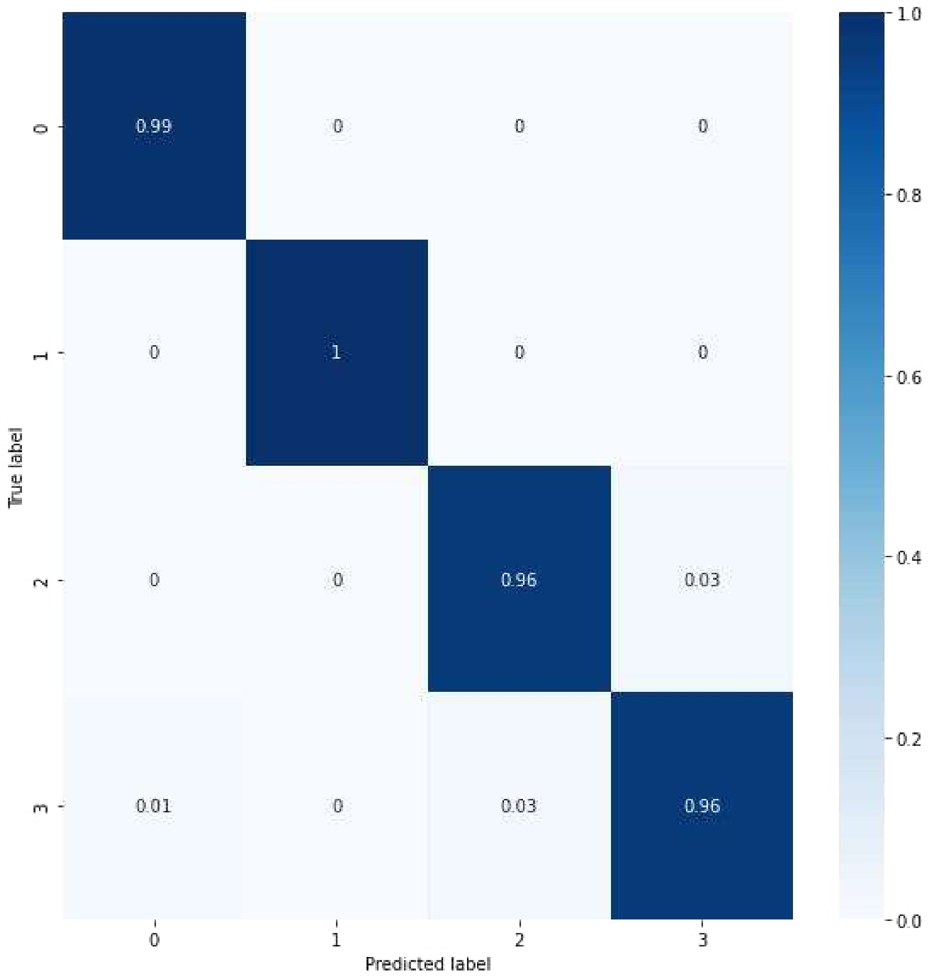
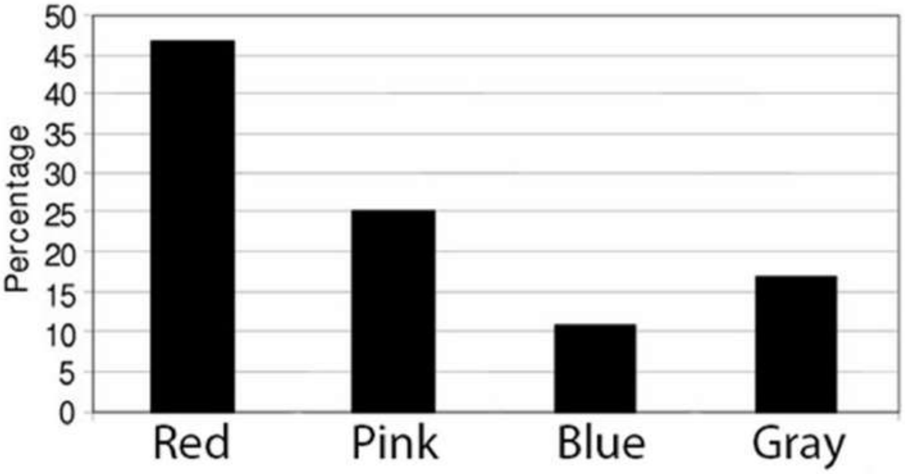
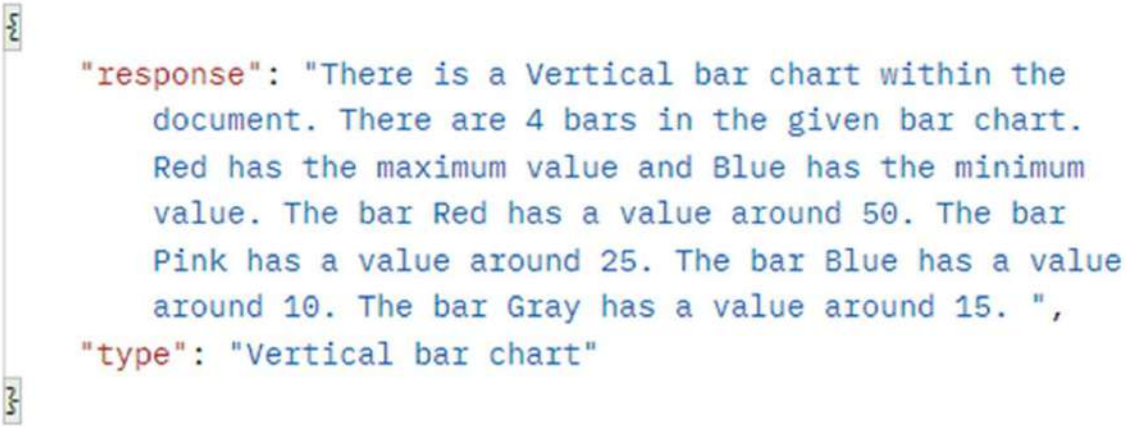


Figure 4.1.6 - Confusion matrix of trained CNN chart classifier

- Results of interpreting charts

  This component has four separate interpretation methods for each chart type. For horizontal and vertical bar charts, it detects the bars and detects the values and labels as well. The accuracy of the OCR results of the images are vary and depends on the OCR module we used called CRAFT-OCR. However, the used OCR engine is highly accurate and reliable for this use case.

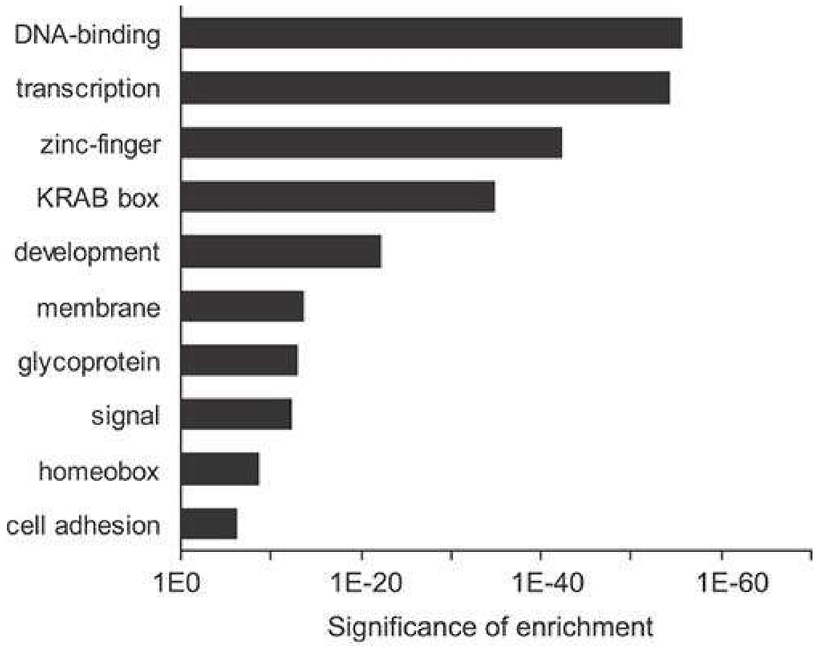


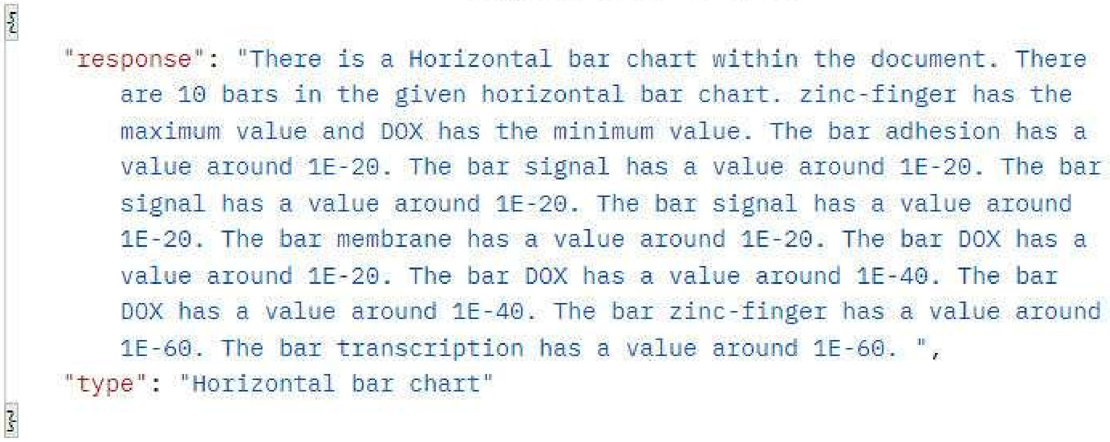


```
{
    "response": "There is a Vertical bar chart within the
        document. There are 4 bars in the given bar chart.
        Red has the maximum value and Blue has the minimum
        value. The bar Red has a value around 50. The bar
        Pink has a value around 25. The bar Blue has a value
        around 10. The bar Gray has a value around 15. ",
    "type": "Vertical bar chart"
}
```

Figure 4.1.7 - Result of vertical bar chart interpretation

For all the bar charts, the bar count detection is highly accurate, and the output is given as a range value rather than an exact value to maintain consistency.

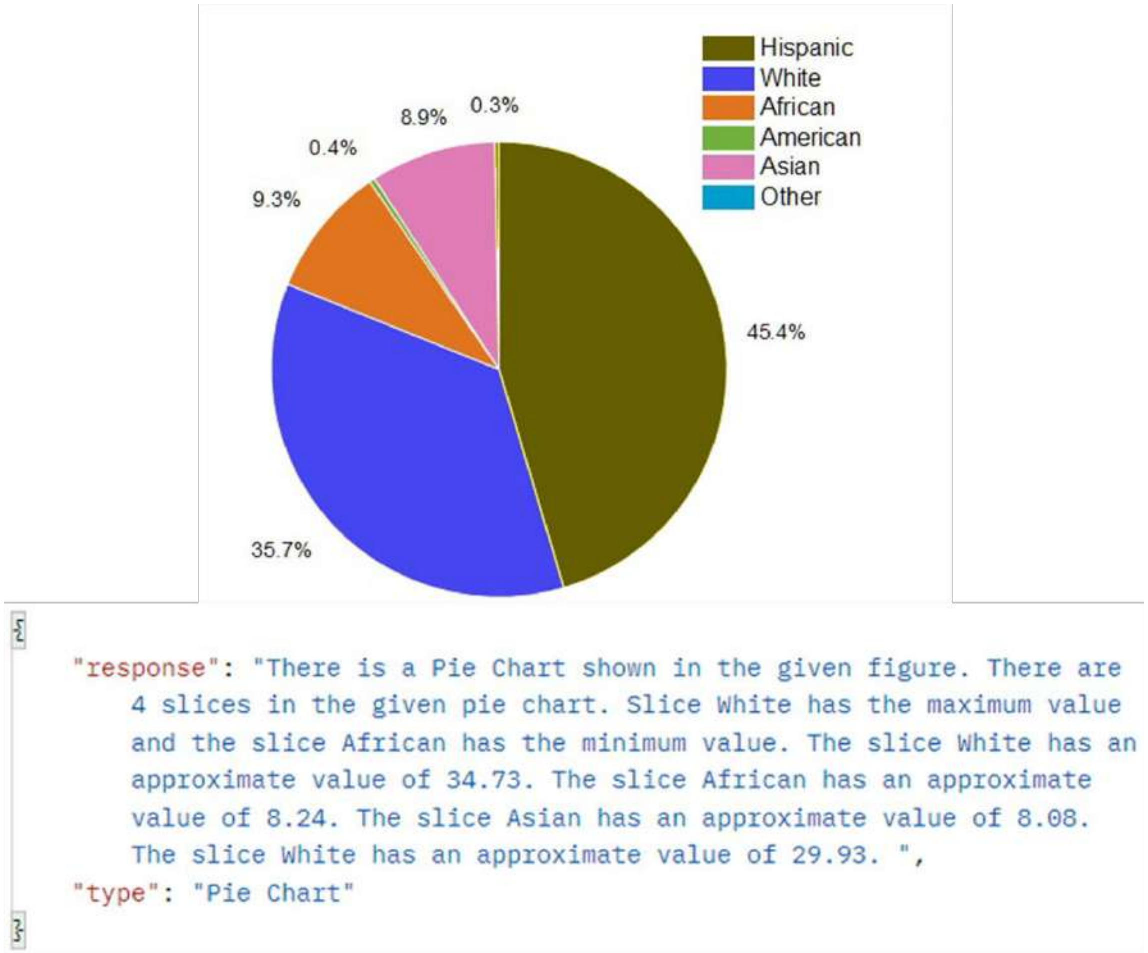




```
{
    "response": "There is a Horizontal bar chart within the document. There
        are 10 bars in the given horizontal bar chart. zinc-finger has the
        maximum value and DOX has the minimum value. The bar adhesion has a
        value around 1E-20. The bar signal has a value around 1E-20. The bar
        signal has a value around 1E-20. The bar signal has a value around
        1E-20. The bar membrane has a value around 1E-20. The bar DOX has a
        value around 1E-20. The bar DOX has a value around 1E-40. The bar
        DOX has a value around 1E-40. The bar zinc-finger has a value around
        1E-60. The bar transcription has a value around 1E-60. ",
    "type": "Horizontal bar chart"
}
```

Figure 4.1.8 - Result of horizontal bar chart interpretation

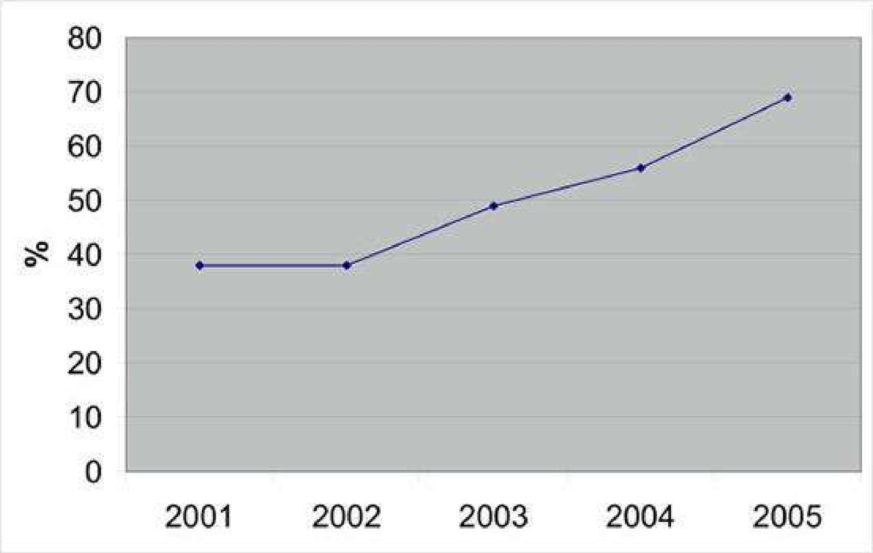
Furthermore, for the pie charts, we use a python dictionary consisting of every possible RGB value to cross-check against the chart legend and the pie chart. We take the values by calculating the area of the slices rather than just reading the text in the chart. So even pie charts without any numerical values are supported. This method also achieved great results. However, the most complicated part of this type of chart image processing is to match the corresponding label to the calculated value. So, there are some minor issues that were seen when we were testing the algorithm where values matched the wrong labels.

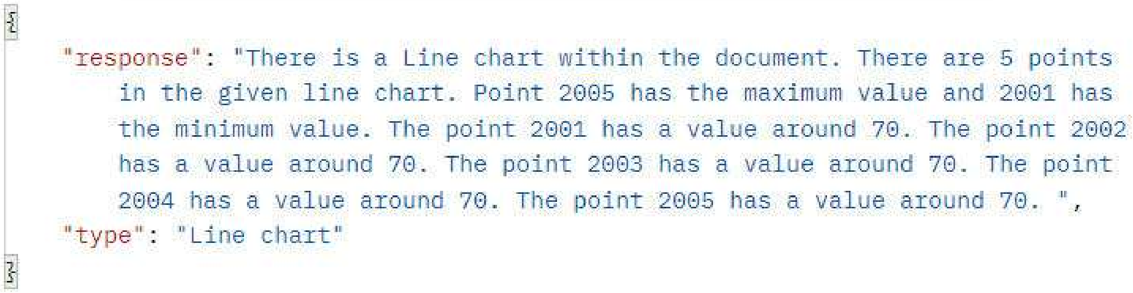


```
{
    "response": "There is a Pie Chart shown in the given figure. There are
        4 slices in the given pie chart. Slice White has the maximum value
        and the slice African has the minimum value. The slice White has an
        approximate value of 34.73. The slice African has an approximate
        value of 8.24. The slice Asian has an approximate value of 8.08.
        The slice White has an approximate value of 29.93. ",
    "type": "Pie Chart"
}
```

Figure 4.1.9 - Result of pie chart interpretation

Since line charts do not have easily recognizable features like bars or slices it is hard to detect the exact values of the charts. The smaller points and thin lines in the line charts are hard to detect and hard to differentiate from other features (axes, legend) within the chart. Out of all four of the chart types of line chart has the least reliable output but the output is accurate enough to generate a meaningful sentence for our use case.





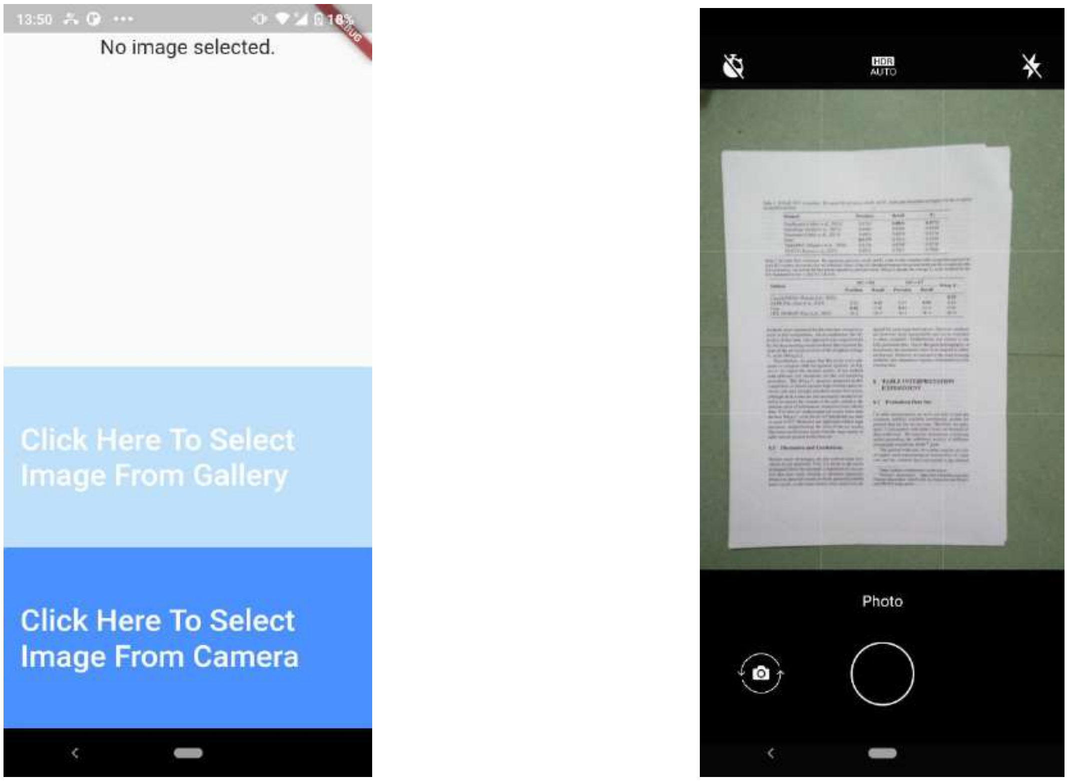
```
{
    "response": "There is a Line chart within the document. There are 5 points
        in the given line chart. Point 2005 has the maximum value and 2001 has
        the minimum value. The point 2001 has a value around 70. The point 2002
        has a value around 70. The point 2003 has a value around 70. The point
        2004 has a value around 70. The point 2005 has a value around 70. ",
    "type": "Line chart"
}
```

Figure 4.1.10 - Result of line chart interpretation

Overall, the chart interpretation module gives highly accurate results compared to currently available studies and applications. Since this is done with mostly rule-based methods, the performance of the component is also high. The final generated explanations of the charts are insightful and useful for print-disabled users.

- Results of table interpretation

The application were identified some several significant results by table extracting and summarizing methodology. The identifying the table is initially done by the classification model and from that document will divided into some sub sections that are tables, charts, diagrams, images, and text likewise. At first system will assist the user by voice and it says about the application and main tasks to the user. As shown in the below figure that system will asking to select a surface in the mobile phone to scan the document or otherwise if there’s a vision impaired or person with print disability, he or she can upload the gallery image as preference.

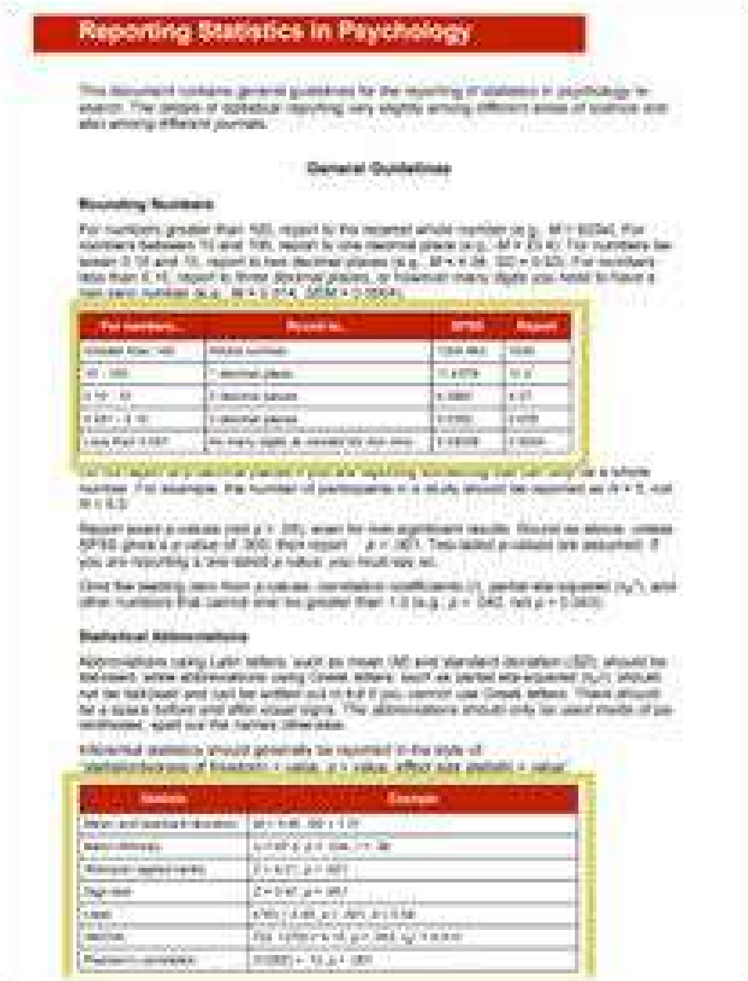


No image selected.

13:50
Click Here To Select Image From Gallery
Click Here To Select Image From Camera
HDR AUTO
Photo

Figure 4.1.11 - Document scanning process

Next system will be getting the image and identifying the document edges because of that system can clearly process the document image. For this purpose, there should be enough light to capture and scan the document. After detecting scanning the document it will segmenting the document into separate section. From that all the table images will passing to the table extraction module. As shown in the below figures system identifies the tables and segmenting it.



Reporting Statistics in Psychology

This document contains general guidelines for the reporting of statistics in psychology research. The details of statistical reporting vary slightly among different areas of science and also among different journals.

General Guidelines

Rounding Numbers

For numbers greater than 100, report to the nearest whole number (e.g., M = [illegible]). For numbers between 10 and 100, report to one decimal place (e.g., M = [illegible]). For numbers between 0.10 and 10, report to two decimal places (e.g., M = [illegible], SD = [illegible]). For numbers less than 0.10, report to three decimal places, or however many digits you need to have a non-zero number (e.g., M = [illegible], SEM = [illegible]).

For numbers...	Round to...	SPSS	Report
Greater than 100	Whole number	[illegible]	[illegible]
10 - 100	1 decimal place	[illegible]	[illegible]
0.10 - 10	2 decimal places	[illegible]	[illegible]
0.001 - 0.10	3 decimal places	[illegible]	[illegible]
Less than 0.001	As many digits as needed for non-zero	[illegible]	[illegible]

Do not report any decimal places if you are reporting something that can only be a whole number. For example, the number of participants in a study should be reported as N = 5, not N = 5.0.

Report exact p-values (and p < .001), even for non-significant results. Round as above, unless SPSS gives a p-value of .000; then report p < .001. Two-tailed p-values are assumed. If you are reporting a one-tailed p-value, you must say so.

Omit the leading zero from p-values, correlation coefficients (r), partial eta-squared (η_p^2), and other numbers that cannot ever be greater than 1.0 (e.g., p = [illegible], not p = [illegible]).

Statistical Abbreviations

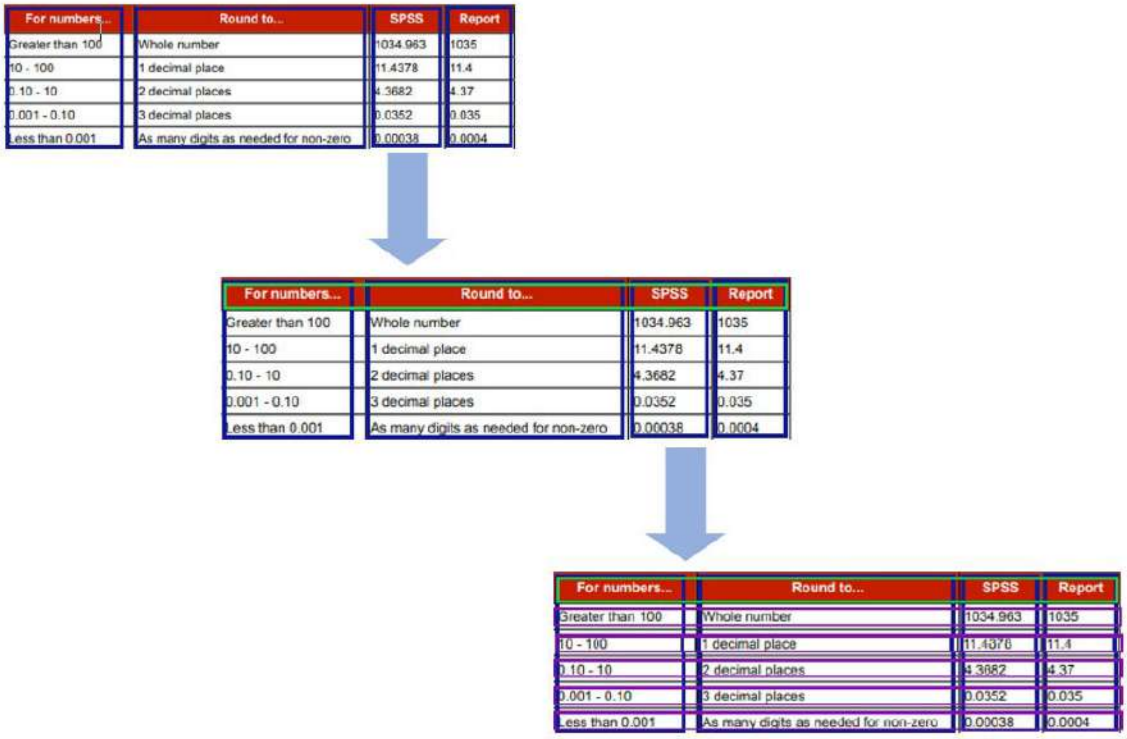
Abbreviations using Latin letters, such as mean (M) and standard deviation (SD), should be italicised, while abbreviations using Greek letters, such as partial eta-squared (η_p^2), should not be italicised and can be written out in full if you cannot use Greek letters. There should be a space before and after equal signs. The abbreviations should only be used inside of parentheses; spell out the names otherwise.

Inferential statistics should generally be reported in the style of: "statistic(degrees of freedom) = value, p = value, effect size statistic = value"

Statistic	Example
Mean and standard deviation	[illegible]
Mann-Whitney	[illegible]
Wilcoxon signed-ranks	[illegible]
Sign test	[illegible]
t-test	[illegible]
ANOVA	[illegible]
Pearson's correlation	[illegible]

Figure 4.1.12 - Identifying the tables within the document

Then after the table identification it will identify the rows and columns separately. For the purpose of analyzing system should identify all the column headers. After the identification of column headers system has the ability to identify the row wise details and in the model all the rows and columns identifies as follows.



For numbers...	Round to...	SPSS	Report
Greater than 100	Whole number	1034.963	1035
10 - 100	1 decimal place	11.4378	11.4
0.10 - 10	2 decimal places	4.3682	4.37
0.001 - 0.10	3 decimal places	0.0352	0.035
Less than 0.001	As many digits as needed for non-zero	0.00038	0.0004

Figure 4.1.13 - Identifying rows and columns

Then with summarizing module system will identifying the number of rows and number of columns. System will be identifying the data within a row and with each column system checking is there any numeric value columns and from that system will identifying the maximum and minimum values in each column.

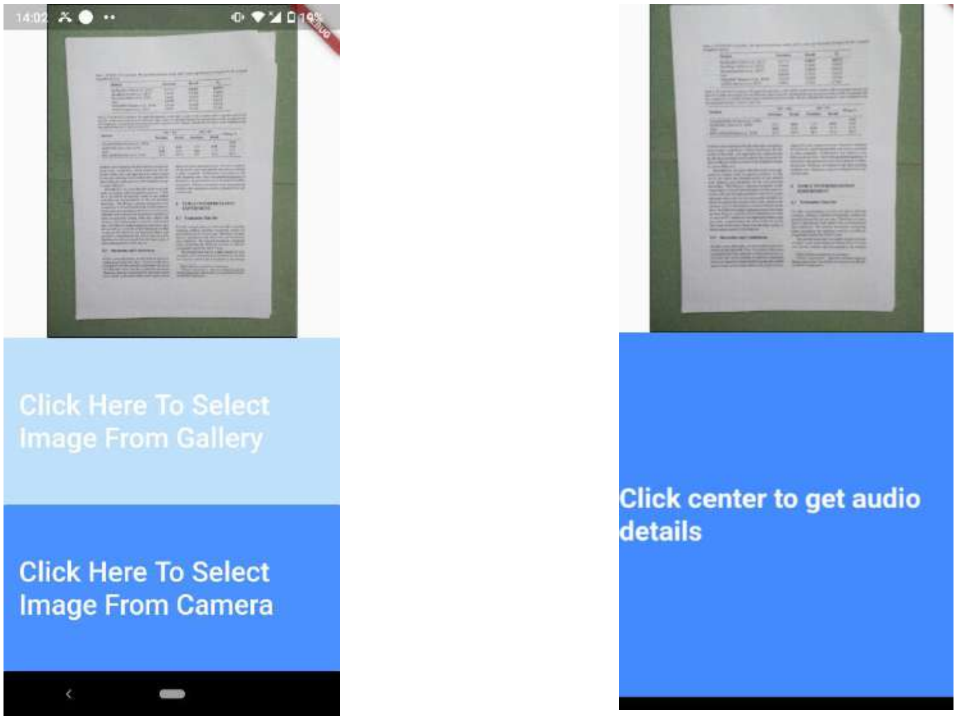
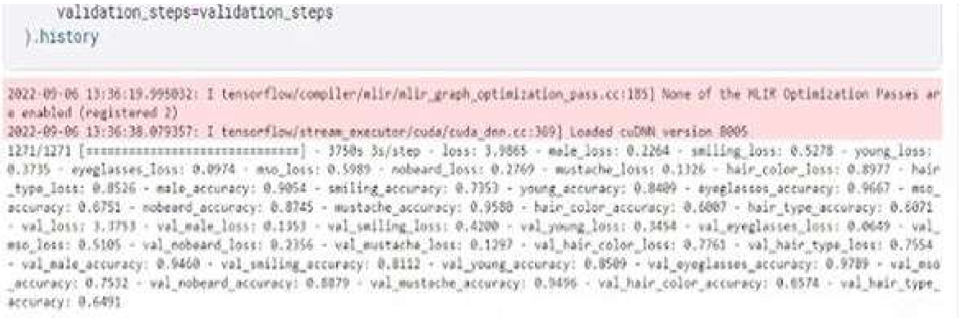

14:02
Click Here To Select Image From Gallery
Click Here To Select Image From Camera
Click center to get audio details

Figure 4.1.14 - UI of the table interpretation

After the analyzation done it will prompt an explanation in voice output. There are some steps to follow to get the voice output and system will assist the user from the voice commands and user can easily hear the results as expected.

- Results of image captioning and description

The image captioning algorithm works for almost any image and generates descriptions for images that are descriptive enough for print disabled users. It identifies the main objects in the image and describe in a way that a blind personal can understand. As the features of detected humans are extracted with efficientv2 implemented as a multi output model the accuracy of the generating captions are also very high as shown in the



```
    validation_steps=validation_steps
).history
```

```
2022-09-06 13:36:19.995032: I tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:185] None of the MLIR Optimization Passes ar
e enabled (registered 2)
2022-09-06 13:36:38.079357: I tensorflow/stream_executor/cuda/cuda_dnn.cc:369] Loaded cuDNN version 8005
1271/1271 [==============================] - 3750s 3s/step - loss: 3.9865 - male_loss: 0.2264 - smiling_loss: 0.5278 - young_loss:
0.3735 - eyeglasses_loss: 0.0974 - mso_loss: 0.5989 - nobeard_loss: 0.2769 - mustache_loss: 0.1326 - hair_color_loss: 0.8977 - hair
_type_loss: 0.8526 - male_accuracy: 0.9054 - smiling_accuracy: 0.7353 - young_accuracy: 0.8409 - eyeglasses_accuracy: 0.9667 - mso_
accuracy: 0.6751 - nobeard_accuracy: 0.8745 - mustache_accuracy: 0.9580 - hair_color_accuracy: 0.6007 - hair_type_accuracy: 0.6071
- val_loss: 3.3753 - val_male_loss: 0.1353 - val_smiling_loss: 0.4200 - val_young_loss: 0.3454 - val_eyeglasses_loss: 0.0649 - val_
mso_loss: 0.5105 - val_nobeard_loss: 0.2356 - val_mustache_loss: 0.1297 - val_hair_color_loss: 0.7761 - val_hair_type_loss: 0.7554
- val_male_accuracy: 0.9460 - val_smiling_accuracy: 0.8112 - val_young_accuracy: 0.8509 - val_eyeglasses_accuracy: 0.9789 - val_mso
_accuracy: 0.7532 - val_nobeard_accuracy: 0.8879 - val_mustache_accuracy: 0.9496 - val_hair_color_accuracy: 0.6574 - val_hair_type_
accuracy: 0.6491
```

Figure 4.1.15 - Accuracy of the generated attributes

As shown above, the accuracy for the attributes smiling, young, hair color, hair type, eyeglasses, mustache, male female, beard are respectively 87%, 85%, 75%, 78%, 97%, 94%, 97%, and 92%.

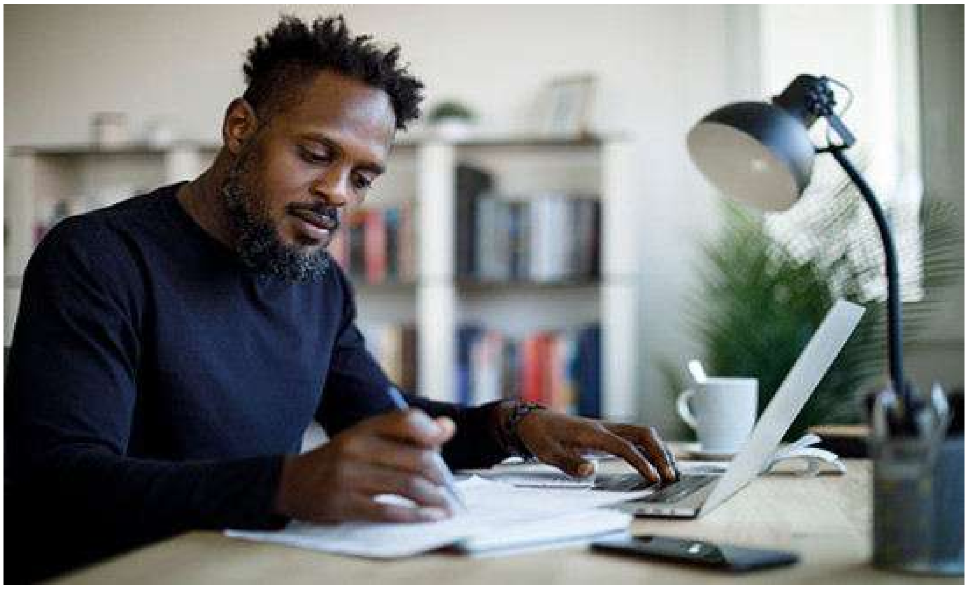
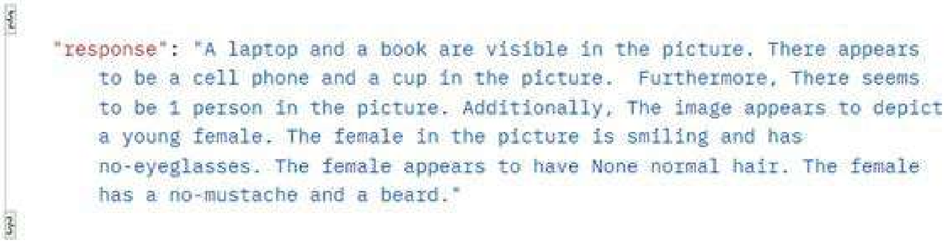


Figure 4.1.16 Sample image tested with man

Below Figure dispalys the caption generated for the figure through the algorithm.



```
{
    "response": "A laptop and a book are visible in the picture. There appears
        to be a cell phone and a cup in the picture.  Furthermore, There seems
        to be 1 person in the picture. Additionally, The image appears to depict
        a young female. The female in the picture is smiling and has
        no-eyeglasses. The female appears to have None normal hair. The female
        has a no-mustache and a beard."
}
```

Figure 4.1.17 Results got for the sample image a man working

Figure 4.1.2 and Figure 4.1.3 shows two results that are collected from the implemented image interpretation component. As you can see the predicted descriptions are approximately equal to the real-time human generated captions. So, it is clearly shown that the implemented mobile application gives a high accurate description and is capable to identify males and female separately and describe the images by indicating the other sub objects in the image.

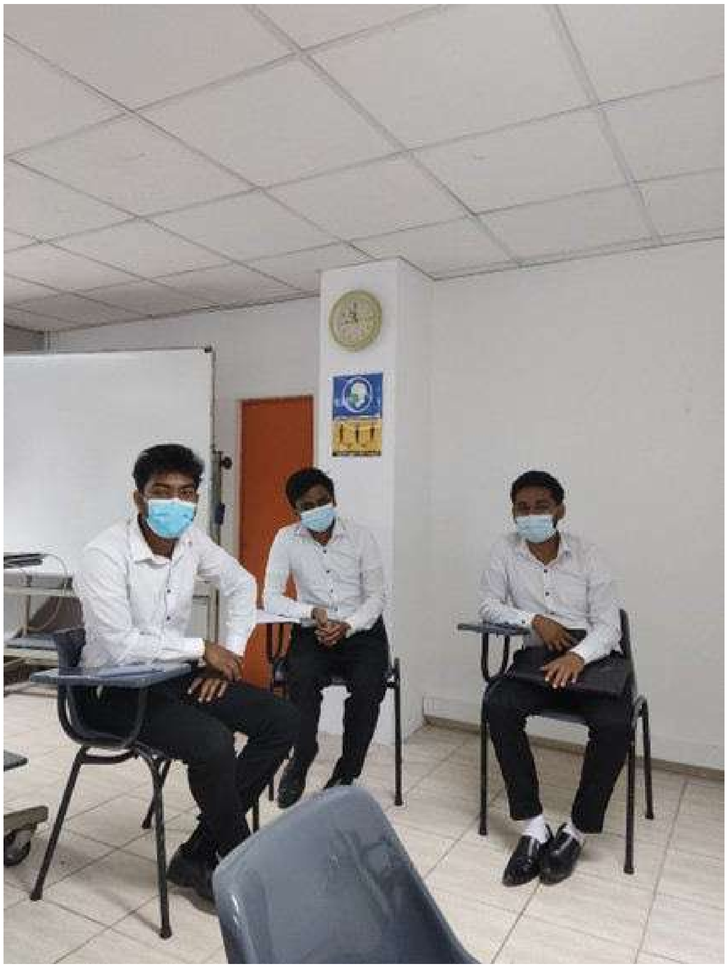
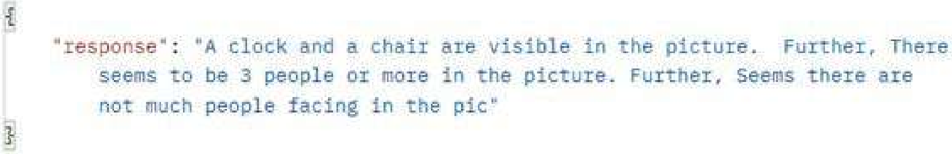


Figure 4.1.18 - Sample image tested with three people sitting

For the above Figure the generated caption through the models that are trained for thr image interpretation will generate the caption as shown in the figure



```
{
    "response": "A clock and a chair are visible in the picture.  Further, There
        seems to be 3 people or more in the picture. Further, Seems there are
        not much people facing in the pic"
}
```

Figure 4.1.19 - Result got for the image of three people siting

- Results of mathematics interpretation

Using and training the model within the designed architecture of a modern neural network To train this model, we use graphics cards with a 1060Ti GTX processor. PyTorch is used to create an implementation of this concept. A fixed learning rate of 0.000001 was utilized throughout the training process, which utilized a total of 200 batches.

The IAM2Latex 100k dataset was utilized for the training of the model. And it has more than a hundred thousand mathematical image formulas. Develop the model's vocabulary before beginning training, and then train it in order of complexity using the vocabulary.

After going through the training process, the model was found to have an accuracy rate of 6.5% given the model and a loss rate of 4.5%. If the current findings are used as a comparison, it is possible to get increased accuracy by doubling the number of graphics cards that are being used.

As the amount of training data increases, there is a minor improvement (less than 1%) in performance that is associated with increasing the model size and/or image resolution.

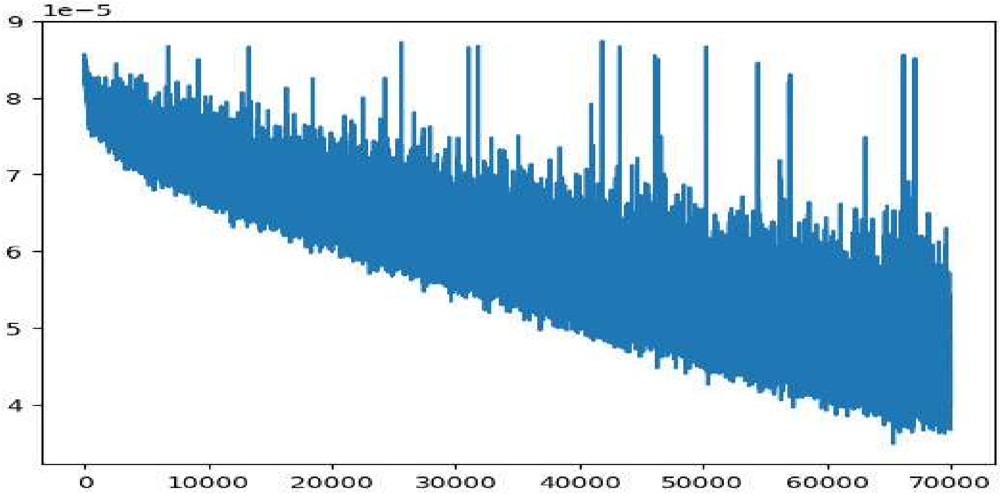


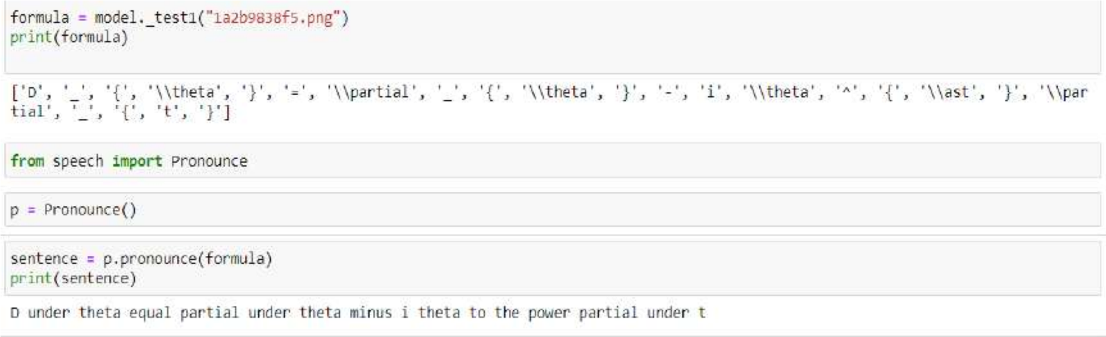
Figure 4.1.20 Mathematics model training loss value

And after the model training we get the Latex output and it have to convert the mathematical way. Therefore, it used vocab list to the convert data to readable math part. E.g. $D_{\theta}=\partial_{\theta}-i\theta^{\partial_{\theta}}$

```
It converts as the ‟ D under theta equal partial under theta minus i
theta to the power partial under t
”
```

$F=ma$

It converts as the “ F equal m a ”



```
formula = model._test1("1a2b9838f5.png")
print(formula)
```

```
['D', '_', '{', '\\theta', '}', '=', '\\partial', '_', '{', '\\theta', '}', '-', 'i', '\\theta', '^', '{', '\\ast', '}', '\\par
tial', '_', '{', 't', '}']
```

```
from speech import Pronounce
```

```
p = Pronounce()
```

```
sentence = p.pronounce(formula)
print(sentence)
```

```
D under theta equal partial under theta minus i theta to the power partial under t
```

Figure 4.1.21 Mathematical way sentence

4.2. Research Findings

This research is mainly focused on building a mobile application to read document content for vision impaired people to fill he lack of available resources in accessible mediums. So much research has been done to implement a device for print disabled people to interpret printed documents. But still there are 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. So, the main outcome of this study is to develop 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, our solution 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.

When considering the accuracy of the implemented overall system, it has achieved an overall accuracy of more than 90%. For the implementation of the system, as this study is mainly under image processing, so many image processing and deep learning techniques have been used. As the study is heavily based on image processing, there are so many datasets that can be used to train all the models with the needs. So, by implementing this mobile application, the significant gap in reading rights and equality between vision impaired people and the general community can be solved for a considerable extent.

4.3. Discussion

In this research the major objective was to develop an assistive mobile application which enables print-disabled individuals to comprehend printed materials. Several factors have been discovered by the development and testing of the application. A considerable proportion of the world's population is disabled, and "printing disability" is one of them. Braille is one of their solutions that they have used to interpret textsand content. But, some content, including charts, graphs, tables, figures, and math or algebraic equations, cannot be represented by the braille system; therefore, these people must seek the assistance of a third party. With such advanced technology in today's modernized world, it's past time we have better options for assisting handicapped persons in several sectors. As a result, we decided to create a tool to address this issue.

Earlier, some research was conducted on the subject in order to gain a better understanding of the population of visually impaired persons. The literature survey included data from the worldwide vision data base gleaned from population-based studies of blindness and visual impairment, as well as age-specific prevalence of blindness and the number of blind persons by age. We wanted some firsthand data on the subject after conducting some background research. That's when we decided to conduct an online survey to gather some important research data. The survey was carried out via distributing a Google form.

The following questions were mostly posed to the audience in order to gain insight into the use of currently popular tools: Do individuals believe that a mobile application can solve their issues related when interpreting a document? Do they know of any alternate means for visually impaired folks to interpret images, charts,tables, graphs, mathematical equations and so on? Which of these instruments are they acquainted with? What are the requirements for the aforementioned tools? Also, we learned from this poll that many people had seen a visually impaired person and faced such challenges while assisting them.

Furthermore, many comments noted that this type of invention would be extremely beneficial given the difficulties faced by the blind community. The majority of respondents indicated applications like E reader, Screen reader, and Voice assistant. In addition, our goal was to discover the limitations of the already available tools so that we could propose answers to these issues and improve the features of our program. This study looked at not just the technological aspects of the problem, but also the biology origins of vision impairment. We discovered that lenses, which develop with age, are the leading cause of full or partial blindness. It is critical to have a solid understanding of the target population when conducting such research. It is vital to distinguish age groups, genders, and so on. The document will be captured from the user's mobile device initially. The backend will then receive the input and collect it for processing. In the backend, the document will be divided into numerous areas, and the regions will then be categorized. The fragments will then be routed to the appropriate algorithms for interpretation. After the analysis is completed, a voice output will prompt an explanation. Furthermore, the system will guide the user through the entire process.

4.4. Summary of student contribution

4.4.1. Student: Priyashan Sandunhetti S. H. S - IT19187242

Contribution: Here, this member is responsible for developing a proper algorithm to classify the content in the scanned material, direct relevant portions of the image to suitable interpretation algorithms. The input of this component is the scanned document, and the outputs are the classified portions of the scanned document, and these outputs will be directed to relevant algorithms to interpret. Furthermore, this member implemented the chart classification and interpretation algorithms to explain the detected charts (Vertical/Horizontal bar charts, Line charts and Pie charts) in plain English. In addition, the scanning process of the printed material will be assisted with the developed in-device real-time edge detection method.

Completed tasks:

- Developed the mobile application camera with real-time edge detection methods
- Implemented the Django backend to expose services like document segmentation, chart classification and interpretation to be used by the mobile application
- Implemented document segmentation algorithm and added it to the Django backend
- Implemented the CNN to classify chart images
- Collected images of charts and created a dataset to train the classification algorithm
- Trained the chart classification neural network and added it to the backend
- Implemented four different algorithms to extract data from chart images (Vertical bar charts, Horizontal bar charts, Pie charts and line charts) and to generate sentences using extracted data.

4.4.2. Student: Dilitha Ranjuna G.P - IT19156484

Contribution: This member's responsibilities align with developing proper table comprehension model within the system. The inputs of this component are the tables and diagrams within the scanned document and the output will be a plain English description of the inputs. There will be a speech engine to assist users when using the application. This speech engine will be useful when guiding the user to navigate through the application and to explain the printed material to the user.

Completed tasks:

- Developed the model to interpret the tables and convert them to simple plain English.
- Developed the speech engine for give commands and information to the user.

4.4.3. Student: Prabash K. V. A. S - IT19117492

Contribution: The tasks of this member align with to develop an explanation of the text and complex mathematical expressions to the user. the input of the components to take the mathematical equations and describe the equation in a mathematical way and also to improve the quality of the input images by removing shadows and darkness. This member will develop the proper user interfaces which will be easily accessible for any user.

Completed tasks:

- Enhanced the main image quality so that every algorithm to output more accurate results.
- Developed the model to interpret text and mathematics to simple plain English.
- Developed proper user interfaces suited for every possible user in the system.

4.4.4. Student: Sanduni Madara P.G - IT19392172

Contribution: The tasks of this member align with developing a system to interpret images included in the scanned material. The inputs of this component are the image portions of the scanned material, and the output will describe the images in plain English. Furthermore, this member will be responsible for the tactile feedback system of the application which will help the disabled user when navigating through the application.

Completed tasks:

- Developed the mobile application
- Implemented the backend with Django to get the image interpretation services by the application
- Developed image captioning and describing models.
- Implemented image interpreting algorithms to extract the image content
- Developed haptic feedback system in the proposed mobile application.

5. 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, our solution 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 printed documents or books into digitalized text and reading the content to the user. The developed solution supports text, mathematical equations, table data, chart data and image interpretation.

There are limitations to this solution as the accuracy of the whole solution heavily depends on the quality of the smartphone camera. To avoid this, one can introduce an external camera with higher quality into the solution. 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] “Blindness and vision impairment,” Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 03-Sep-2022].

[2] 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.

[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: 03-Sep-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, pp. 279–284.

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

[9] D. A. Borges Oliveira and M. P. Viana, “Fast CNN-Based Document Layout Analysis,” in 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017.

[10] 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.

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

[12] 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.

[13] 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.

[14] S. Paliwal, D, Vishwanath, R. Rahul, M. Sharma, and L. Vig, “TableNet: Deep Learning model for end-to-end Table detection and Tabular data extraction from Scanned Document Images,” 2020.

[15] M. Namysl, A. M. Esser, S. Behnke, and J. Köhler, “Flexible table recognition and semantic interpretation system,” 2021.

[16] K. A. Hashmi, M. Liwicki, D. Stricker, M. A. Afzal, M. A. Afzal, and M. Z. Afzal, “Current status and performance analysis of table recognition in document images with deep neural networks,” 2021.

[17] 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.

[18] 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.

[19] 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.

[20] R. Noor, K. Mejbaul Islam, and M. J. Rahimi, “Handwritten Bangla numeral recognition using ensembling of convolutional neural network,” in 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018.

[21] X. Bian, B. Qin, X. Xin, J. Li, X. Su, and Y. Wang, “Handwritten mathematical expression recognition via attention aggregation based Bi-directional mutual learning,” Proc. Conf. AAAI Artif. Intell., vol. 36, no. 1, pp. 113–121, 2022.

[22] S. Bender, M. Haurilet, A. Roitberg, and R. Stiefelhagen, “Learning fine-grained image representations for mathematical expression recognition,” in 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), 2019.

[23] Z. Shen, R. Zhang, M. Dell, B. C. G. Lee, J. Carlson, and W. Li, “LayoutParser: A unified toolkit for deep learning based document image analysis,” 2021.

[24] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, ‘Detectron2’, 2019. [Digital edition]. Available at: https://github.com/facebookresearch/detectron2.

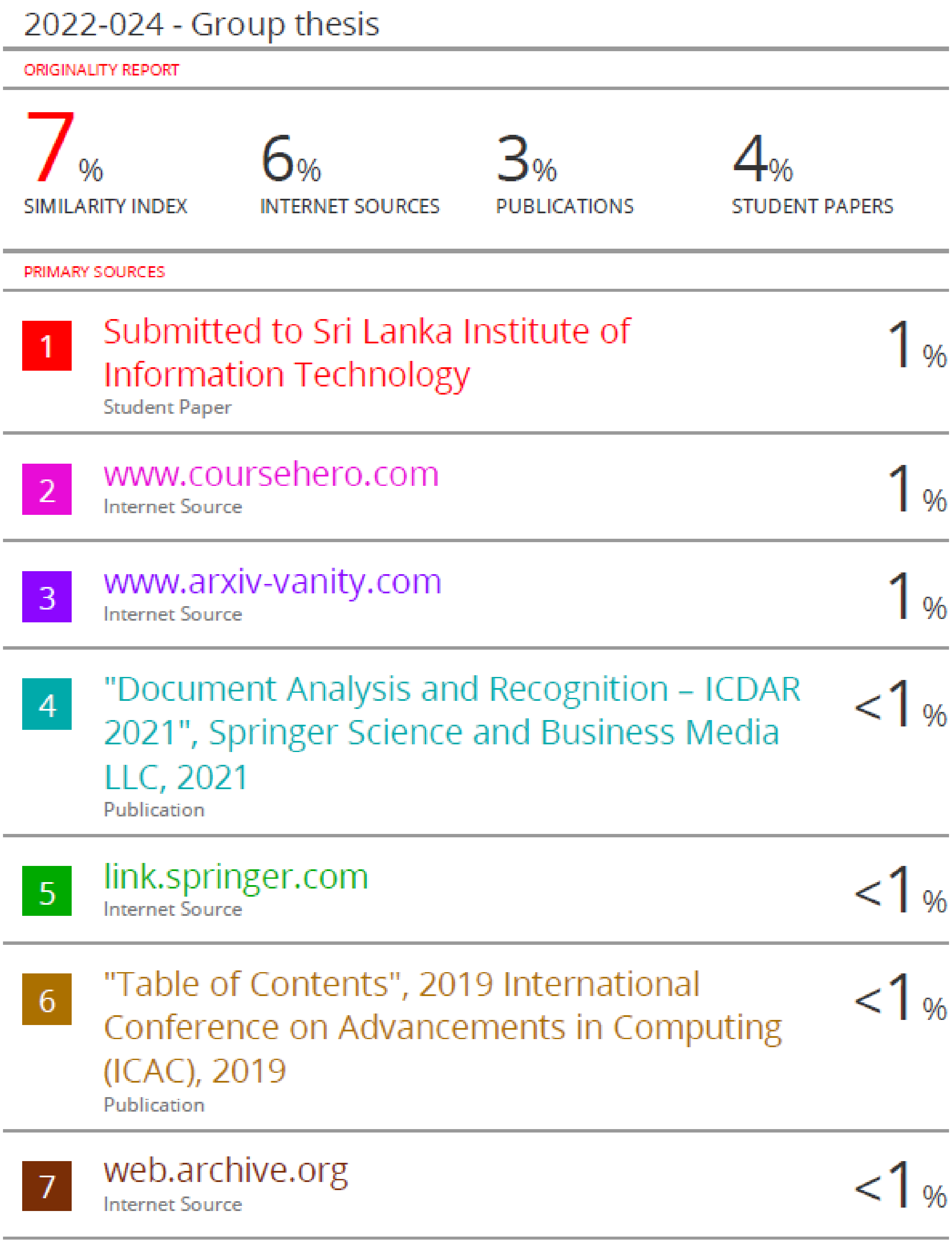
[25] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character region awareness for

text detection," arXiv [cs.CV], 2019.

[26] S. Saoji, R. Singh, A. Eqbal, και B. Vidyapeeth, 'TEXT RECOGINATION AND DETECTION FROM IMAGES USING PYTESSERACT', Journal of Interdisciplinary Cycle Research, τ. XIII, σσ. 1674–1679, 08 2021.

[27] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," 2014.

APPENDICES



2022-024 - Group thesis

ORIGINALITY REPORT

7%	6%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	1 %
2	www.coursehero.com Internet Source	1 %
3	www.arxiv-vanity.com Internet Source	1 %
4	"Document Analysis and Recognition – ICDAR 2021", Springer Science and Business Media LLC, 2021 Publication	<1 %
5	link.springer.com Internet Source	<1 %
6	"Table of Contents", 2019 International Conference on Advancements in Computing (ICAC), 2019 Publication	<1 %
7	web.archive.org Internet Source	<1 %

Appendix A: Plagiarism report