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

Prabash K.V.A.S.

(IT19117492)

B.Sc. (Hons) Degree in Information Technology
Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2022

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

Prabash K.V.A.S.

(IT19117492)

The dissertation was submitted in partial fulfillment of the requirements for the B.Sc. Special Honors degree in Information Technology.

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2022

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning, and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student ID	Signature
Prabash K.V.A. S	IT19117492	

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

Signature of the supervise Date

…… …….

(Dr. Anuradha Jayakody)

……. ……..

(Ms. Shashika Lokuliyana)

ABSTRACT

In today's world, mathematics is a core essential for all human beings. It is critical for everyone to enhance their math skills in college or through self-study. Students with vision impairments, on the other hand, have limited access to printed materials. They will need the assistance of a human reader if they want to learn through printed materials. However, it is unbiased due to human expense and availability. but the Braille materials are in the possession of the VI student at the same time, the braille materials are few, and the braille language is complex.

VI students benefit greatly from video and audio technology resources as the other sources provide a limited contribution to their studies. For those students, the audiobook is difficult to comprehend. But they must continue to listen to the audiobooks. However, these limited resources had a negative impact on the math skills of VI students. Text-to-speech Technic is extremely popular across the world. VI students can use the camera to read printed texts. However, just a few applications for mathematical reading have been developed. Math Talk and Math-Player [1-2] are both available in English. Math Genie is available in English, French, and German,[3] as well as AudioMath, which is available in European Portuguese [4]. Regrettably, their mobile apps have a lot of math reading and listening limitations. Many applications cannot read or recognize advanced mathematical equations (ME). So, we are developing a mobile application capable of reading ME. As a result, VI students can use their mobile phones to read printed content without interruption or external assistance.

Keywords: VI Students, Mathematics, Braille System, Mobile Application

ACKNOWLEDGMENT

First and foremost, I'd want to express my heartfelt gratitude to Dr. Anuradha Jayakody, our project supervisor, for her continual guidance and support, which aided me in completing my undergraduate research successfully. Apart from my supervisor, I'd like to express my gratitude to Ms. Shashika Lokuliyana, the co-supervisor of this research project, for always being eager to lend a hand. Many folks' perseverance was necessary to complete this dissertation. Throughout the duration of our study endeavor, their invaluable input and time aided me in achieving perfect success. Then, I'd like to offer my heartfelt gratitude to my teammates and friends for their unwavering support during this effort. Finally, I would want to express my gratitude to my parents for their unwavering support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
IAPB	The International Agency for the Prevention of Blindness
TTS	Text-to-Speech
CNN	Convolutional Neural Networks
ME	Mathematical Equations
VI	Visual Impairments
LNE	Line Number Encoding
ML	Machine Learning

1. INTRODUCTION

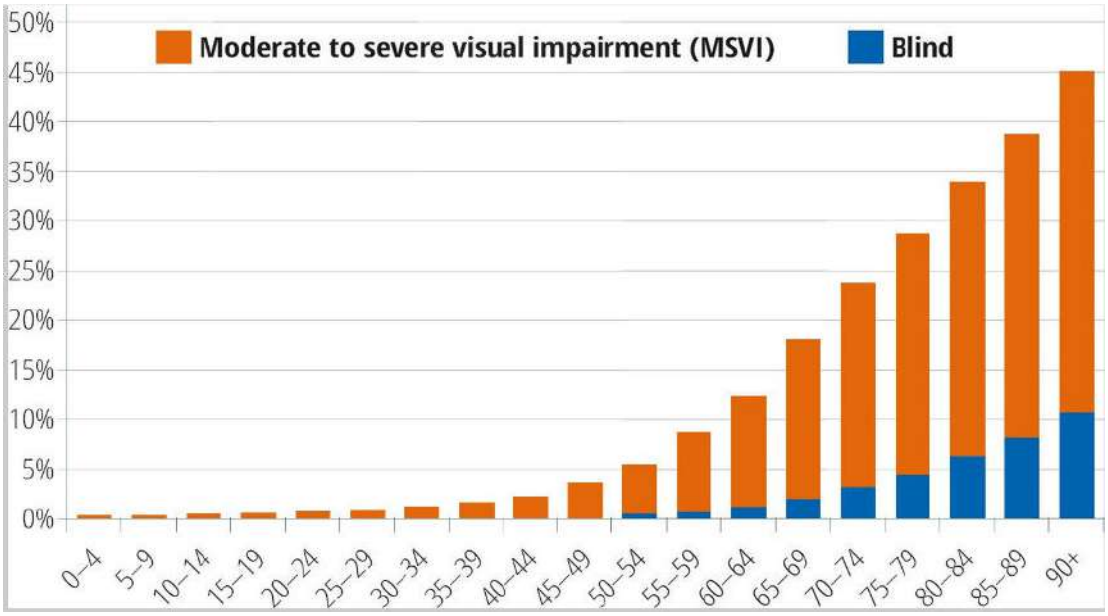
This section introduces the background of the research problem. And it considers different ways to address the problems of visually impaired students. It also discusses the ways and patterns in which visually impaired students learn. The use of Braille was discussed here. This chapter stated the research objectives as well as the critical research questions. This section further discussed literature research used to understand gaps in research.

1.1. BACKGROUND LITERATURE

Over the past three decades, the growing number of visually impaired and blind people has been declining. But the problem is that the decreasing point cannot be taken as a good sign compared to the rapidly growing population of the world.

The International Agency for the Prevention of Blindness (IAPB) Vision Atlas Group [5] published surveys of distance vision loss, near vision loss and blindness from 1980 to 2014 using 98 countries. In 2015, 223 million people worldwide are living with visual impairment. Of these, 36 million people are blind, and 217 million people are visually impaired. And the survey shows that 55% are women and 89% of people live in middle-income countries. 80% of people aged 50 years or older experience visual impairment worldwide (Figure 2).

Figure 1.1 The global blindness and visual impairment count in age



According to United Nations data, when predicting the future population, the world expects 9.7 billion people in 2050. So, expect 4.2 billion people over the age of 50 and therefore the problem of vision impairment may increase.

Many blind people in the world face many difficulties in their daily activities [6]. Most non-blind people latch onto blind people differently. It makes them very uncomfortable.

One of the biggest challenges for blind people is navigating around places. Also, visually impaired people face difficulties in moving around. And some commercial spaces are accessible to blind people because of tactile tiles. But it doesn't happen everywhere, and they face problems while traveling.

And blind people try to read the material they need to read, but many materials are inaccessible to them. When they use the internet, and many website reading options and image descriptions are not supported.

And because they did not identify and arrange their clothes correctly, they were very embarrassed when they went out in inappropriate clothes. And some non-blind people try to help blind people too much. Therefore, blind people face trouble in those situations.

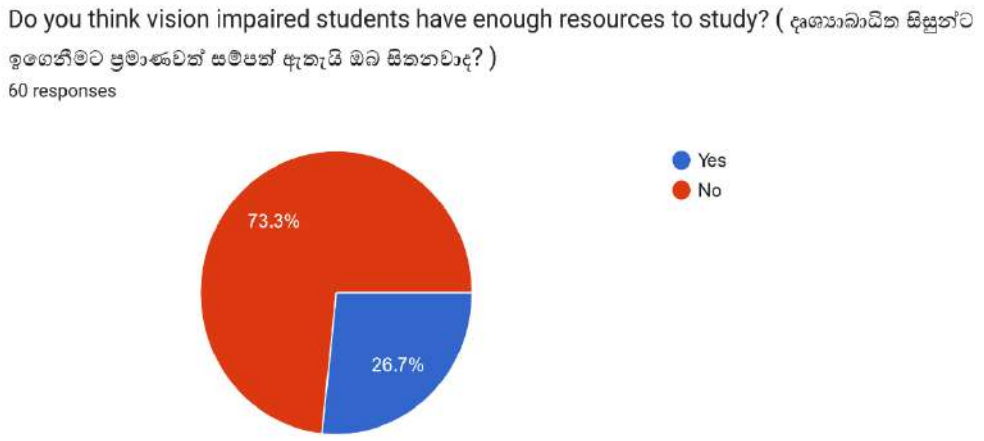
Blind people mainly used Braille for reading and writing. Braille is not a language. It represents alphabetic, numeric, and punctuation using dotted cells [7]. Braille functions similarly to written text as a code that reproduces the sounds, phonology, and meanings of a language.

There are two main types of Braille in common use [8]. Grade 1 Braille has 26 letters of the alphabet and is often used by children when they are first learning to read and write. Grade 2 Braille is the most widely used type of Braille in the world. It also has 26 letters of the alphabet and is more complex compared to Grade 1. In this Braille system, one dot represents a word. Apart from that, there are many other types of brails in the world.

Visually impaired students learn in different ways [9]. They use braille, audio tape, enlarged print, or a screen reader. Braille readers take more time to read compared to others. Therefore, most students use screen readers [10]. But the problem is that not every part can be read using readers. When students need printed materials, many are not available in Braille or audio format. So, they have to wait for materials to be produced as needed. Hence, blind students have to lag behind other students.

According to the findings, a public literature survey acknowledges that blind people lack the resources to learn (Figure 1.2).

Figure 1.2 Summary response of VI students to enough resources to study



Many blind students get help from other blind people in their education. But they have to do their own thing when it comes to higher education and living. Because there are not enough people to follow their needs.

Mathematics is considered a universal language. Learning mathematics is very difficult for visually impaired students as learning mathematical formulas and their nature is very complicated. Reading and writing math is completely different from reading and writing regular letters [11].

There are two types of Braille. Dots six and eight represent letters on a flat surface. The problem when talking about braille is that there are no international standards for printing braille math for math characters, but it is done by local experts.

The Braille system has limited character in mathematical formulas such as algebra, square roots, logarithms, series, additions, products, theorems, integration, limits, and trigonometry. But Braille displays a limited number of characters. If 10 upper and lower case letters and numbers and common punctuation are included, 6 dots represent a total of 64 (26) characters. Therefore, Braille characters span up to 8 dots. Then get 256 (28) characters. But increasing the number of Braille dots also increases the difficulty of Braille [12].

Figure 1.3 Difference of Brail 6 dots and 8 dots [13]

6-dot	Braille 6-dots	8-dot	Braille 8-dots
	5		8
	24		24
	5,24		248

Mathematical equations have highly strict syntax constraints, and even minor markup errors can have severe effects (forgetting to close a parenthesis). Furthermore, the causal links between symbols extend beyond left-to-right interactions, including a wide range of horizontal to-vertical relationships (e.g. fractions and matrices). Such semantic linkages are especially challenging since some of them result in the reduplication of included letters into tiny ones.

Optical Character Recognition (OCR) [14] is main part of identifying the text part in visual data. But the OCR is also used to identify the math equations and formulas. Conditional random fields, support vector machines, and k-nearest neighbor classifiers are just a few examples of OCR models that rely on pre-defined feature representations of the underlying data.

Using Image to Sequence architecture [15], it can extract text from an image and accurately sequence it without any constraints on the text and non-text orientation, layout, and size. Furthermore, it can also be trained to generate auxiliary markups related to formatting, layout, and content. It can be used to predict the next output and input sequences that may or may not be of the same length or at the same time.

It is very useful to convert math images into LATEX and create vocabulary by letter for the math reading section. A state-of-the-art paragraph-level reorganization can be achieved through vocabulary modeling. Hence after getting the LATEX formula using the image it has to convert a readable math paragraph.

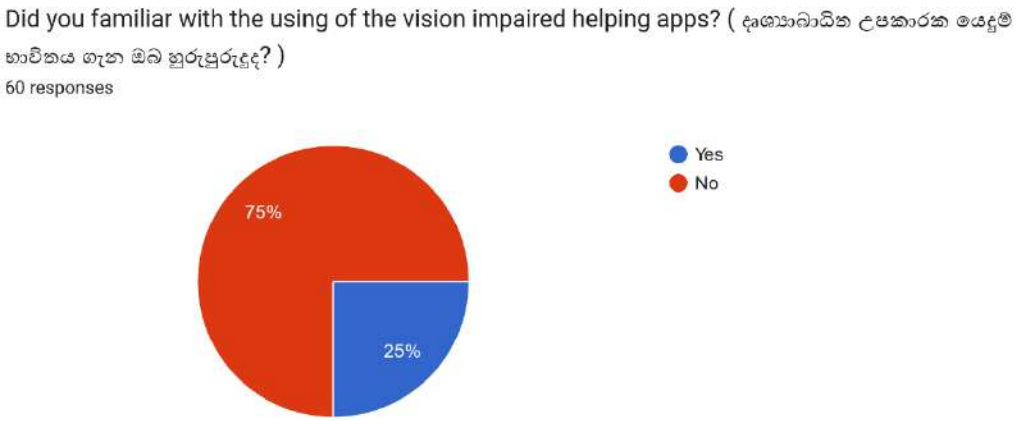
Math reading is none of the same as normal reading. Math formulas have different ways.

$\int dx\, f(x) = 2x + 2$ = Integrate dx f open bracket x close bracket equal two multiply x plus two

Because of that ME have to be read in separate mathematical ways.

Now a days most of the blind students and the vision impaired peoples used mobile phones to makes their daily work easier. Hencs they have many third party applications in now.

Figure 1.4 Summary response to Vision impaired app usage



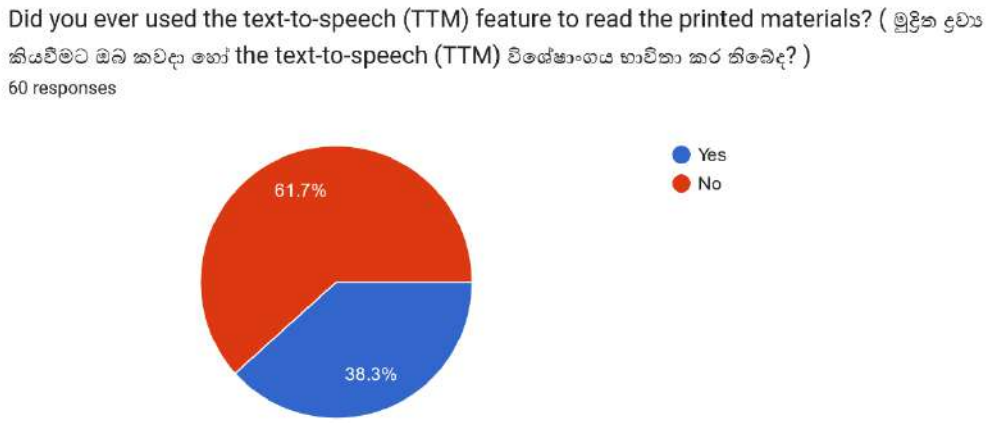
Reading the passage, checking the color of the surface, identifying the objects and much more can be done using mobile phone applications. So, users can do very useful things using the mobile phone. But the problem is that many applications do not output correctly to blind users. So users have to face severe difficulties.

Students can use a mobile phone to accelerate their education. Using the mobile phone they can read every printed material without the help of others or Braille translation.

It's very useful feature is Text-to-Speech (TTS) [16]. Therefore, most blind or non-blind people can listen to printed material. Because of that feature, students can listen and understand lessons without others.

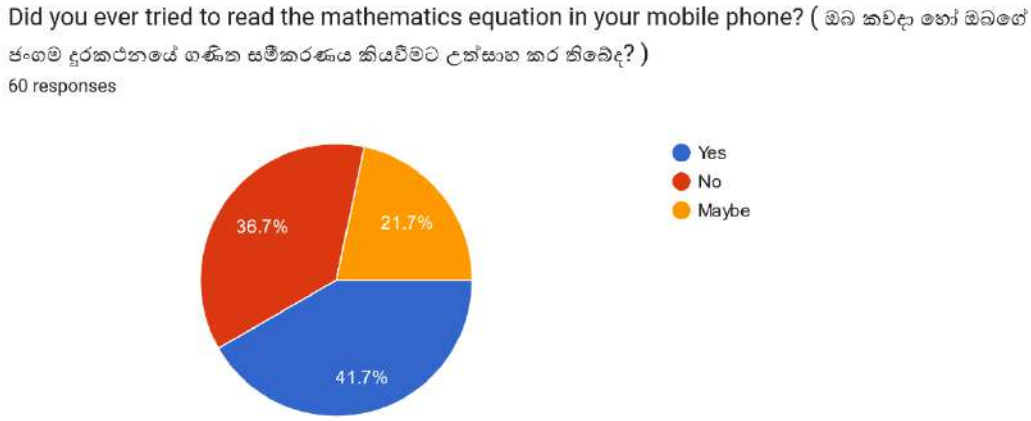
In the survey, we discovered students using the TTS process. We found that students tried to read math and try to translate math equations into English using TTS.

Figure 1.5 Summary response to attempts to read mathematical equations using TTS



Using the TTS feature can read the paragraph easily. But the problem is when it comes to the mathematics equation reading it can't properly read the math equation. Some applications can't identify the correct math equation or badly read the identified equation.

Figure 1.6 Summary response to attempts to read mathematical equations using TTS



1.2. RESEARCH GAP

According to the survey, most respondents say there is a lack of resources for blind people. 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 ME 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 [17] (Research A) 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 [18] (Research B) 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 [19] (Research C) 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.

Table 1.1 Comparison of former research

	Research A	Research B	Research C	
Explain the mathematic equations and plain text in simple English.	✗ (Text only)	✓ (Math only)	✓ (Math only)	✓ (Math and Text)
Mathematics equation recognition	✗	✓	✓	✓
Optimized for mobile, cloud use	✗	✗	✗	✓
Add grammar rule to explain mathematical equation	✗	✓	✗	✓

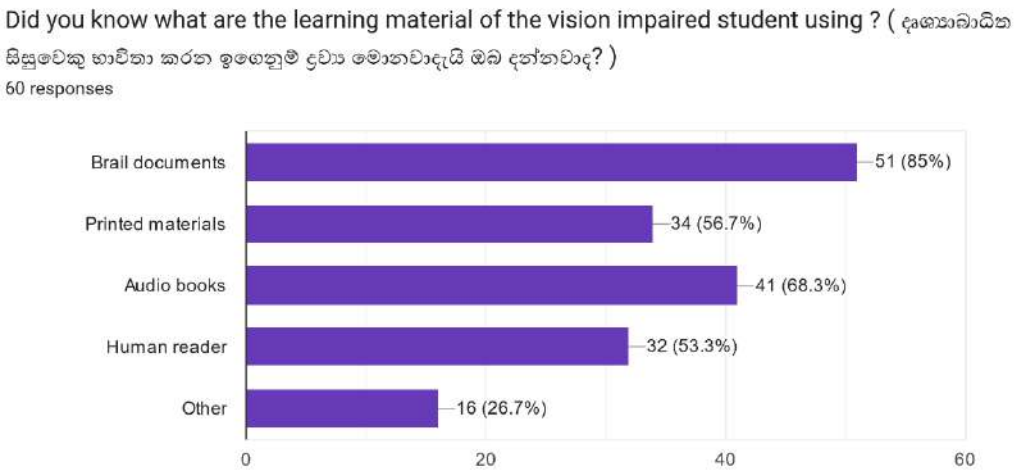
In the all researches compares and then give the modem solution to blind and vision impaired students to use mobile application and read the math equation in mathematical ways.

2. RESEARCH PROBLEM

In my opinion, everyone in the world ought to keep learning until the day they die. People who are blind or visually challenged are also able to learn. However, persons who are blind have greater challenges than the average person. In addition to this, kids are required to spend additional time reading and studying. due to the fact that they do not have sufficient resources.

According to the survey, many peoples say that students with visual impairments (VI) have certain materials they need to learn. Audiobooks, Braille language books, and human readers support are just a few of them.

Figure 2.1 Summary response of materials used by visually impaired students



Visually impaired students take classes from preschool to high school to use the above material to study. But as the subjects of learning expand, they face many difficulties [20].

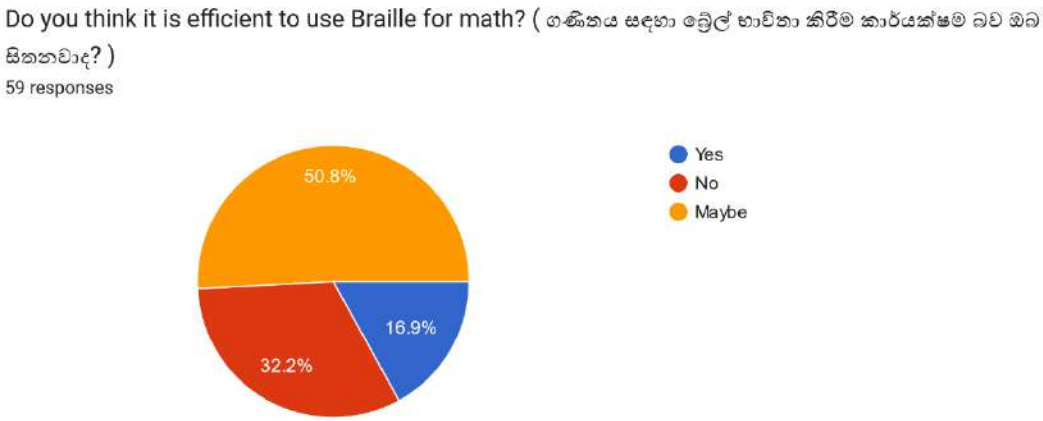
- Using Braille, the advanced mathematical formula equation becomes complex.
- Lack of the audiobooks
- Human assistance is a very expensive and time-consuming problem.

- Lack of Braille books
- Many books are printed material and cannot be used for vision-impaired students.

In the pre-school to ordinary levels, Vision Impaired students can get support from human readers like the teachers, and others. But after the ordinary level's Vision, impaired students have to learn by themselves. After also they can get the support of the human readers. But it is too expensive. And maybe they do have not much time to spend with them. Therefore, it is not a very practical solution. But some may be joining the blind and vision impairment community and they try to get help from others and help each other. It is much more effective comparing to learning themselves.

They can then be studied using Braille material. But the problem is that Braille is also difficult. For example, visually impaired students may want to learn advanced math, but they may have some problems with a lack of characters [21]. This is because the Braille language has two different modes, which are 6 dots, and the expanded version of the previous version is 8 dots. Then the language also expanded and very complicated. If they want to learn higher math so they have to learn first brail language complicated structure. So therefor it is time wasted to comparing to other nonblind peoples. If using the expanded brail, it helps to identify the math formula and specially sign of equation. If they can not understand the math symbol, then they read the formula completely inaccurately.

Figure 2.2 Summary response to learning mathematics in Braille



They can therefore use audio books. These days, several authors publish their novels alongside audiobooks. Because visually challenged students may easily get audiobooks and listen to them wherever they live, audiobooks are incredibly beneficial. However, the issue is that not all works are offered as audiobooks. Students who are blind or visually challenged cannot read all the needed books. And the biggest issue is that a lot of people dislike listening to books. Continuously listening to audiobooks is monotonous. The results of our survey are shown below. As a result, it is reasonable to assume that many students dislike audio books

Figure 2.3 Summary response to attempts to use audiobooks

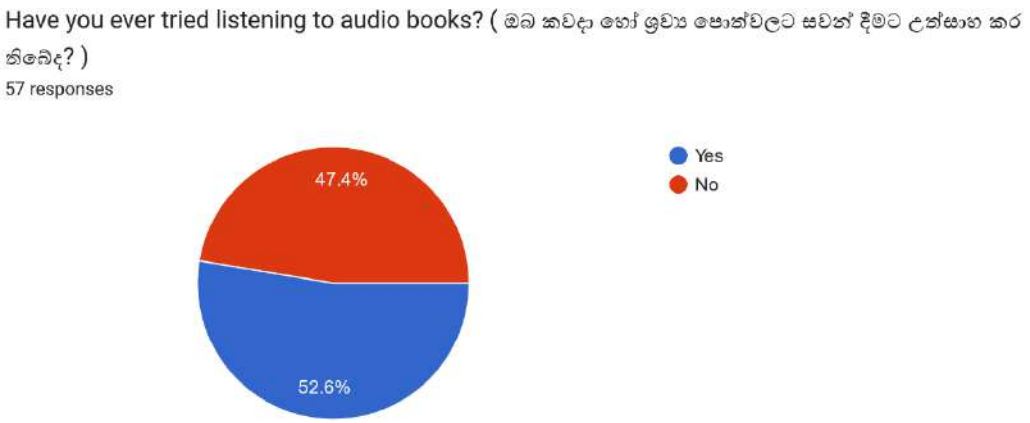
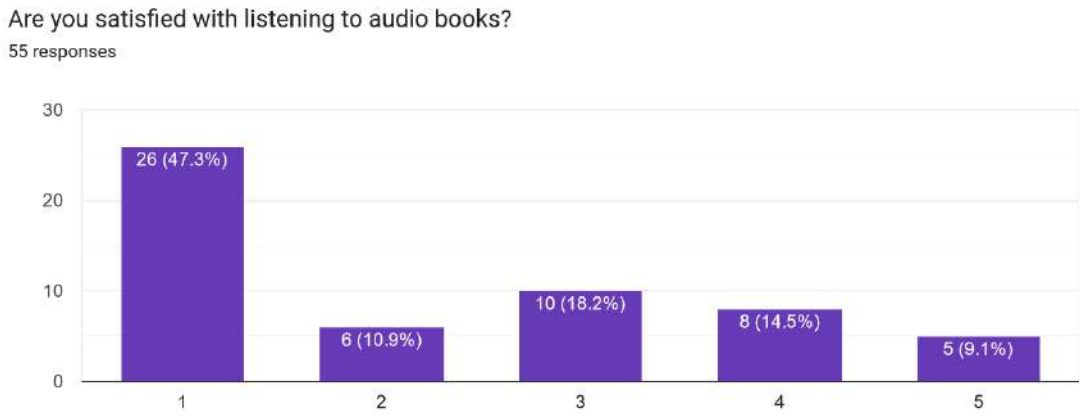


Figure 2.4 Summary Response to Listening to Satisfied Students' Audio Books



Visually impaired students also use the screen reader option when using computers. The GUI-based interface allows them to type whatever they want and read what they type. Now a days it very trendy feature is it. Most of the websites and the software's have this feature and most of the vision impairment peoples use it. But if a visually impaired student writes a ME, it reads like a word. As a result, they may not have a good understanding of what they want. Because math reading is completely difference reading normal paragraph.

Therefore, the TTS and OCR processes are quite significant. The text-to-speech (TTS) and optical character recognition (OCR) systems for English and a number of other languages are based on computers and mobile phones. The optical character recognition process makes printed text readable by computers. After that, TTS will do a reverse translation of the text and will play back the audio that was produced by the computer's processing.

However, due to their inability to comprehend the formula problem, pupils who are blind or visually handicapped have some challenges when attempting to acquire mathematical knowledge. Reading mathematical expressions with TTS and OCR does not result in very accurate readings. This is due to the fact that the mathematical equation is read as a word. If that's the case, the mathematical argument falls apart. Since reading math requires a distinct set of skills than reading regular text. If students who are visually challenged can read arithmetic equations using a smartphone app, it will be incredibly helpful for them to achieve the goals they have set for themselves in life.

3. RESEARCH OBJECTIVE

In research, we have to try to conclude one main objective as well as several secondary objectives.

3.1 MAIN OBJECTIVE

The fundamental objective of this study is to develop an application for mobile phones that is capable of recognizing mathematical expressions or uploading mathematical visuals via a camera, and then converting that information into a math method through the use of a straightforward English voice. A camera will serve as the foundation for this application.

3 SUB OBJECTIVE

- Identified the mathematics formula.

  Through the use of the camera, the mobile app should be able to correctly recognize the formula. Because the camera takes pictures, some of which may have shadows and others may contain noisy data. In addition to this, the formula is brought up to speed with the rest of the text. Therefore, the program ought to completely ignore the text data and instead just record the formula data.

- Identified mathematics symbols.

  Formula recognition is included into the majority of mobile apps. As a consequence of this, the model that has been proposed has to detect complex mathematical formulas. It is possible to recognize integral, alpha, theta, beta, differentiation, and symbols, among other mathematical concepts. In addition,

the recognized formula ought to have a comparable connection to the other formula.

- Convert math to LATEX formula

  As a result of the findings of the literature review, upload or capture images of formulas so that they can be converted. Following the successful acquisition of the image, it is processed by the model and then transformed into a latex formula. Converting the latex formula uses little time and effort because it is simple to isolate the bits that can be read. In addition, latex is capable of separating mathematical components and can read mathematical notation.

- Convert to the Latex formula to the readable parts.

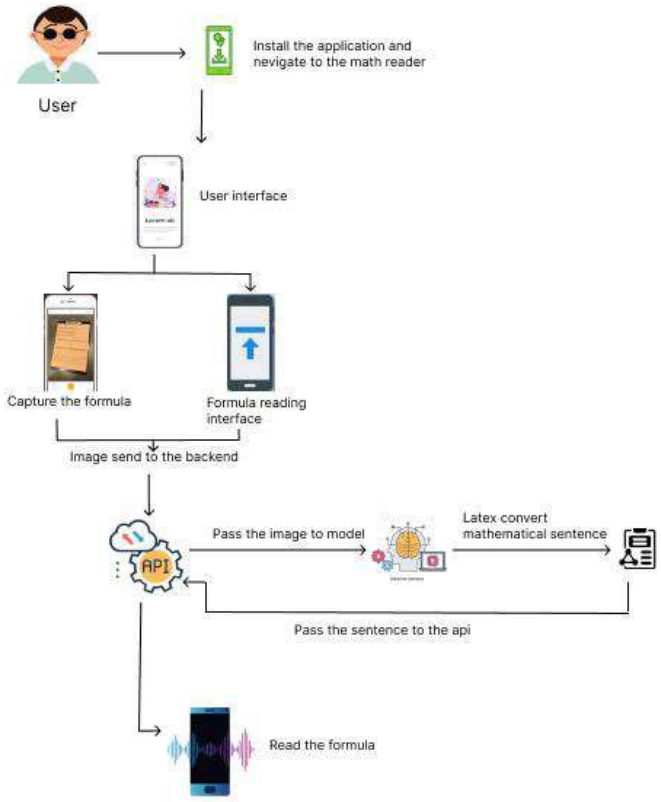
  It could be simpler to divide the components into readable parts once the formula has been converted to a latex format. because the formula for the latex gave it structure. In addition to that, it can be utilized to the creation of formulas in the form of sentences. To begin, it is necessary for us to develop a vocabulary for the latex formula. By using it, we can change it into a formula that can be read in its full sense.

4 METHODOLOGY

In this section, the methodology of the entire system, as well as the model diagram, the methodologies used, and the commercialization components of the product, are described.

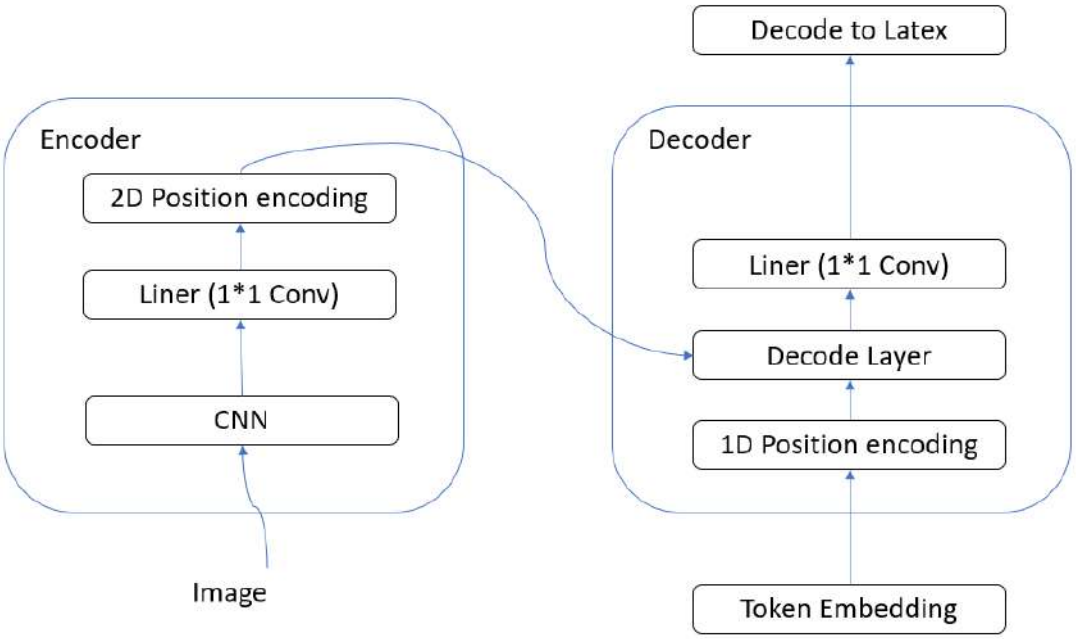
4.1 Overall system

Figure 4.1 System overview diagram



4.2 Model diagram

Figure 4.2 Model architecture



4.3 Used techniques and dataset

Image preprocess

All of the photographs in the collection have been scaled to the exact same dimensions, both in terms of width and height. After the image size has been translated to the size of the formula, the image size of each image must be maintained consistently; hence, the photos are inserted into the padding. Due to the fact that the image is transformed into a 2D matrix before the model and the vocabulary can be used.

Resnet architecture

There is a growing trend in machine learning (ML) for using more distributed architectures, where several processing nodes are used to speed up the training of models. We use the PyTorch ResNet model. The PyTorch ResNet module is a building block of a deep learning network that is used to model the geometry of images by stacking several blocks together to form a deep neural network. It consists of two basic parts: the input layer and the output layer. For each block in the network, an input image is passed through a series of nonlinear transformations such as pooling and convolution and then a new image is produced as the output of the block. And also in this architecture can increase the layers as the developers needed. Running the ResNet architecture can be use the GPU performance and can be increase the GPU power.

Image converts to sequence architecture

In order to accomplish sequence learning, it devised a code-decoder architecture that was based on two RNNs. To be more specific, the RNN encoder takes a sequence that has a variable length and transforms it into a context variable that has a fixed shape. Next, the RNN decoder takes the created tokens and the context variable and generates the output sequence token by token. However, despite the fact that not all input tokens can be utilized for the decoding of a specific token, the context variable that encodes the entirety of the input sequence is still employed at each step of the decoding process.

Math token vocabulary

Develop a vocabulary using both letters and symbols that is integrated. Additionally, the data may be in lowercase or uppercase form. Training data and validation data were utilized during its development. It is helpful in recognizing the LATEX formula [22] and giving output that is flawless. After going through the data in the dictionary, a math token vocabulary is made, which may then be allocated to a distinct math formula node in the LaTeX code. This is done after the data has been examined.

Math reading vocabulary

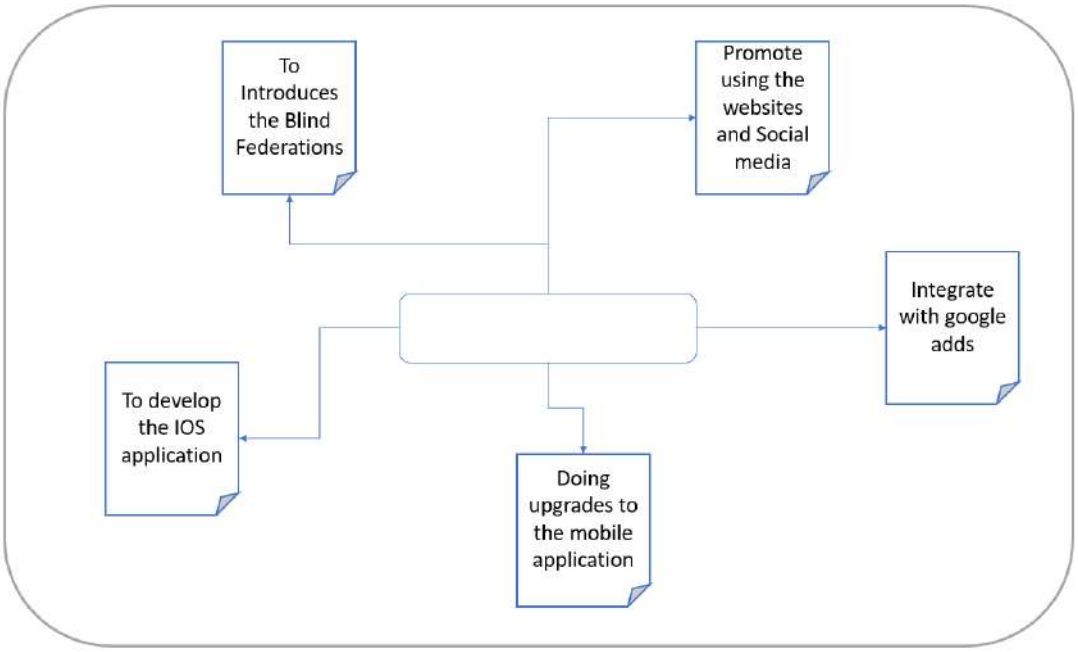
With the use of data obtained from LATEX, a different vocabulary was built. In the read mode, it is used to generate mathematical formulas. In the write mode, it is used to write mathematical formulas. Following the successful conversion of the LATEX formula, the user will produce the words and a new phrase that are required in order to read the mathematics. By following these basic guidelines, you will be able to create your own LATEX symbols.

Dataset

A dataset that has already been pre-built for use in OpenAI's work on the Image-to-latex system. 100,000 equations and images, each of which was split into a train, validation, and test set. Formulas extracted from the LaTeX source code Every image is a PNG file with a predetermined size. The formula is written in black, while the rest of the image is translucent except for where the formula is.

4.4 Commercialization of the product

Figure 4.3 Commercialization of the project



As the above-mentioned characteristics can be used the commercialize the application product. Most of the research doing the math equation converting to LATEX [23-25]. But here it interpreted with the mobile application. It can be very useful for blind and vision impairment users in the world. This application can use every age of peoples. And there is no limit to the usage.

This application is user-friendly enough that even young children may use it without any problems. This capability can be utilized not just in Sri Lanka but also everywhere else in the world because it was developed in the English language, which is a language that is understood everywhere. Because of this, everyone who is able to converse in English will be able to take use of the benefits that this program provides. Utilizing today's available technology, users' awareness of this application can be increased as a result.

By introducing the community of blind federations to the application, it will be possible to gather feedback, increase the number of users, and assist those who need assistance. Additionally, it is possible to obtain a large amount of feedback regarding the application and what needs to be improved, as well as the challenges that users have while using the application, and to maintain an updated version of the application based on the input users provide.

5 Testing & Implementation

This section content two main sections implementation and the testing. Used to implementations tools and technologies mention in sub section of implementation. The way of developing the application describe in the implementation section. And the testing the model and the application mention in the testing section.

5.1 Implementation

Based on the findings and the literature review it identified the there is a lack of resource to the vision impairment peoples. And there are huge need of math reader to that people. And it should be the identified math formula and read the math formula in mathematical way using the mobile phone.

For the implementation it must do a requirement analysis. Therefore, at the first stage, gathered the requirements into three main categories such as, user requirements, functional requirements, and non-functional requirements. Some of the examples relates to those requirement types are listed in below,

User requirements

- User should have basic knowledge of the app manage
- Capture the text correctly through the camara.

Functional requirements

- This app should be equally impressive the vision impaired peoples and normal peoples.
- After the text capture give the alert to the user
- Capture result delivers high volume audio in Sinhala

Non - functional requirements

- Usability – VI students can easily use this app without any hassle
- Availability – This app should be able to use anywhere they want.
- Performance – Capturing the text of the printed material is with the high accuracy
- Reliability – Always provide high accuracy of capture output

Using the found requirement and the new ideas implement the model to the formula image to convert to the Latex formula.

Dataset images

Using the IM2Latex 100k dataset in the dataset has the different sizes images and width and height. Therefore, before the using the dataset create a new dataset adding to the padding to images and it crop the size of image for to the text area. And the correction of the image angle to correct way. This creating new dataset helps to use the text recognize part easily. Because if the images have various sizes and width/height it gives the low accuracy of the model and tokenizers can't get the correct token to the map the character.

Image preprocessing

Utilizing the generated dataset, transform photos using the albumenation tool in Python [26]. 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.

Tools and technologies

When focusing on the tools and technologies that are used to construct the functionality of this mobile application, the table that follows provides a full description,

Table 5.0.1 Tools and technologies

Description	**Tools and Technologies**
Programming IDE	Visual Studio Code
Programming language for mobile	Flutter and Dart

application development	
Backend	Django
Machine learning model training	Python language with, Anaconda and Juptyer
Team connectivity	Teams and WhatsApp

5.2 Testing

It tested uploading and capturing mobile equation images by using the final product as the mobile app and conducting the tests. The data that was tested as well as the data th at was output may be found here.

$S_\theta = i\,\vartheta\,n\,.$

Output as the mathematical way = S under theta equal i theta n dot

$Y = mx + c$

Output as the mathematical way = Y equal m x plus c

Following the development, the mobile application ought to operate without any problems. due to the fact that some of the users are blind or have other VI. As a consequence of this, the program in question must be accessible and functional for users who are blind. Utilizing a PC to do something is very different than using a mobile application. Therefore, the mobile application ought to incorporate voice commands as a means of operation.

Table 5.2 Test cases for the mobile application

Test case	Test steps	Results and comments
T001	Using the voice command tell user to how to open camera	pass
T002	Using the voice command tell user to how to upload image	pass
T003	Read equation page navigate	Pass
T004	Equation reader button place inform voice command	Pass
T005	Equation read	Pass
T006	Back to the main frame	pass

6. Budget

Table 6.1 Budget

Item	Cost(Rs)
App publishing cost on google play	5000.00
Backend hosting cost	10000.00
Paper publishing cost	50000.00
Total	65000.00

7 RESULTS & DISCUSSION

In this section, it will provide an analytical description of the most important outcomes of the product, research findings, and a concluding discussion of the product, together with the appropriate diagrams and screen images that are relevant to them.

7.1 Results

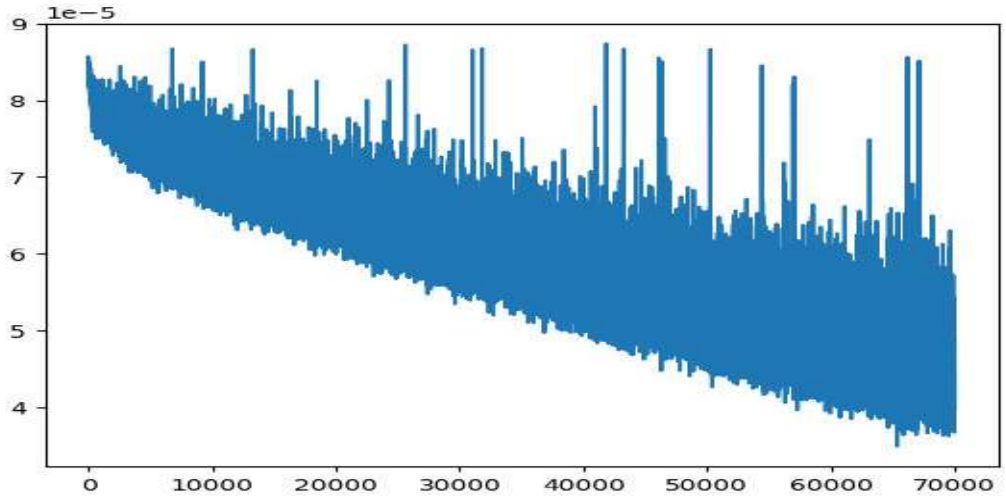
Using and training the model within the designed architecture of a modern neural network To train this model, we use graphics cards with a 1060Ti GTX processor. PyTorch [27] is used to create an implementation of this concept. A fixed learning rate of 0.000001 was utilized throughout the training process, 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 7.1.1 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 "

Figure 7.1.2 Mathematical way sentence

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

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

```
from speech import Pronounce
```

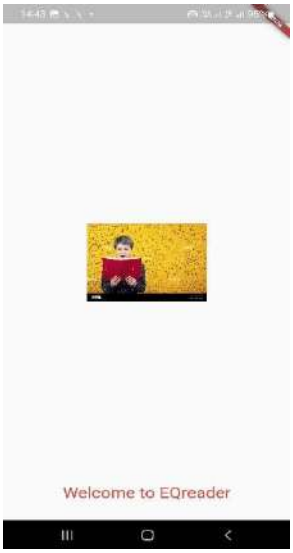
```
p = Pronounce()
```

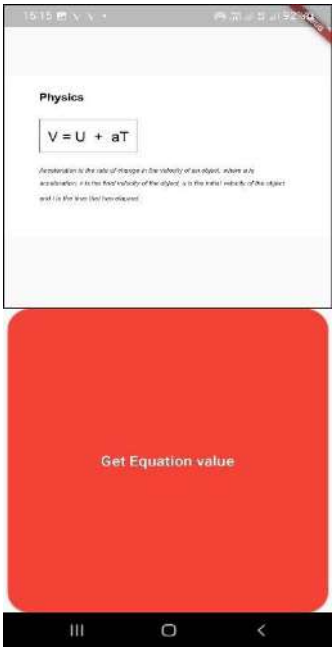
```
sentence = p.pronounce(formula)
print(sentence)
```

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

The way of system work flow for the whole functionality is mentioned in the below with using related user interfaces in the below.

Figure 7.1.3 Mobile application frontend





7.2 Discussions

The mobile application was designed expressly for people who have visual impairments (VI) with the goal of making it as easy as possible for them to use in their day-to-day lives. It is possible that the output of the program will be in the form of words that individuals who have a background in mathematics will be able to understand. This may be accomplished through the use of the device's built-in camera or through the uploading of images to a formula. This application was put into operation after it was determined that the pertinent research articles and the findings of the surveys that were carried out needed to be studied. This function is incorporated into the formula that is used to determine the expected result as a consequence of the findings and the study of the relevant literature. In addition, in order to create products that are more useful, it is necessary to win the support of both the Blind Federation and the larger community as a whole.

8. Conclusions

In the research, we developed the reading of mathematical formulas using the mobile phone. Capturing or uploading the image can read the math formula using the app. After the input image to the application, the application connects to the backend and after converting the formula to latex formula and generating readable text in mathematical form.

There are a lot of apps available for people who are blind or visually impaired. However, the majority of them are things that can be identified and aid with reading, recognizing colors, and listening studies. In addition, there are a number of apps that can read mathematical formulas. The difficulty, however, is that the software has a low level of accuracy and it takes a very long time for it to recognize the formula. A few apps treat the file as if it were ordinary text. Because of this, users who are blind are unable to recognize the equation. There are instances when you need the help of others in order to read through that formula once again.

We proposed the innovative approach by using an attention-based sequence-by-sequence architecture and a tensor-tensor architecture. Our approach incorporates Resnet CNN layers as a component. After being received, the image goes via the encoder and is then turned into metrics by means of 2D excited encoding.

The decoder will be used to map the image to the Latex formula, and the text will be used to decode the image. In addition, the decoder will be sent to the vocabulary that was developed using Latex. After mapping the encoding of the image and the data for the words, it then generates a latex formula that corresponds with the image. For the purpose of training this model, we make use of the IM2Latex 100k dataset. Creates the mathematical path sentence by using the language that is provided as well as the latex formula that was constructed.

In addition, for the purpose of commercializing the app, a YouTube video, flyers, a poster, and a Facebook page were made. Through the application of those methods, commercialization can take place. Create a product demonstration video, then upload it on YouTube and distribute the URL to it to various individuals. Make a Facebook page and let other people know about it. In addition, flyers and posters should be designed and printed in order to provide additional information about the product to members of the general public. In addition for us to generate the application, please.

After that, we will be able to inform users who are blind or have VI about the app. In addition, by incorporating blind federation into the application, we are able to obtain additional ideas on mobile apps and expedite the spread of information regarding the application.

The majority of the formulas in our model may each be read as if they were written on a single line. It reads the multiline formula, but with a reduced level of accuracy. As we continue this line of study, one of our goals is to modify the model so that it can read the multiline formula.

This dataset is being trained on a GPU with a low power consumption, and our objective is to eventually retrain it on a more powerful GPU set. In addition, with the help of the upgraded computer, we hope to achieve a higher level of precision with the model.

In addition to making the app easier to use. Voice instructions are now available for users of the app. Enhance the functioning of the buttons by giving them vibrating patterns that make them immediately recognizable. After that, people who are blind will have a much simpler time locating the buttons.

REFERENCE

[1]M. I. Susperreguy and P. E. Davis-Kean, “Maternal Math Talk in the Home and Math Skills in Preschool Children,” *Early Education and Development*, vol. 27, no. 6, pp. 841–857, Mar. 2016, doi: 10.1080/10409289.2016.1148480.

[2]“MathPlayer: fully accessible web-based Math. - Free Online Library,” *MathPlayer: fully accessible web-based Math. - Free Online Library*. https://php.radford.edu/~ejmt/deliveryBoy.php?paper=eJMT_v1n3n2 (accessed Oct. 18, 2022).

[3]S. Lecce, F. Bianco, and C. Hughes, “Reading minds and reading texts: Evidence for independent and specific associations,” *Cognitive Development*, vol. 57, p. 101010, Jan. 2021, doi: 10.1016/j.cogdev.2021.101010.

[4]“AudioMath: Blind children learning mathematics through audio | 10.1515/IJDHD.2005.4.4.311,” *Sci-Hub | AudioMath: Blind children learning mathematics through audio | 10.1515/IJDHD.2005.4.4.311*. https://sci-hub.hkvisa.net/10.1515/IJDHD.2005.4.4.311 (accessed Oct. 18, 2022).

[5]“Vision Atlas - The International Agency for the Prevention of Blindness,” *The International Agency for the Prevention of Blindness*, Oct. 17, 2022. https://www.iapb.org/learn/vision-atlas (accessed Oct. 18, 2022).

[6]L. Kumar, “Daily Life Problems, Struggle and Challenges Faced by Blind People,” *WeCapable*, Aug. 14, 2018. https://wecapable.com/problems-faced-by-blind-people/ (accessed Oct. 18, 2022).

[7]“What Is Braille? | American Foundation for the Blind,” *The American Foundation for the Blind*. https://www.afb.org/blindness-and-low-vision/braille/what-braille (accessed Oct. 18, 2022).

[8]J. Jullion, “What Are the Different Types of Braille? | T-Base Communications,” *T-Base Communications*, Feb. 15, 2021. https://tbase.com/what-are-the-different-types-of-braille/ (accessed Oct. 18, 2022).

[9]“Strategies for Learning and Teaching | National Council for Special Education - CPD and In-School Support,” *Strategies for Learning and Teaching | National Council for Special Education - CPD and In-School Support*. https://www.sess.ie/categories/sensory-impairments/visual-impairment/tips-learning-and-teaching (accessed Oct. 18, 2022).

[10]“Screen Reader for Low Vision Students? - Perkins School for the Blind,” *Perkins School for the Blind*, Aug. 03, 2022. https://www.perkins.org/resource/screen-reader-low-vision-students/ (accessed Oct. 18, 2022).

[11]“What is the importance of reading and writing in the math curriculum? - IDRA,” *IDRA*, Mar. 23, 2018. https://www.idra.org/resource-center/what-is-the-importance-of-reading-and-writing-in-the-math-curriculum/ (accessed Oct. 18, 2022).

[12]“The Impact of Braille Systems on Advanced Mathematical Geometry,” *The Impact of Braille Systems on Advanced Mathematical Geometry*. https://ieeexplore.ieee.org/document/9378682 (accessed Oct. 18, 2022).

[13]H. Kacorri and G. Kouroupetroglou, “C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part III, LNCS 8011, pp. 331—340, 2013. ’ Springer-Verlag Berlin Heidelberg 2013 Design and Developing Methodology for 8-dot Braille Code Systems,” *Researchgate.net*. https://www.researchgate.net/profile/Georgios-Kouroupetroglou/publication/256292575_Design_and_Developing_Methodology_for_8-dot_Braille_Code_Systems/links/02e7e5221ad17d9667000000/Design-and-Developing-Methodology-for-8-dot-Braille-Code-Systems.pdf (accessed Oct. 18, 2022).

[14]“Optical character recognition - Wikipedia,” *Optical character recognition - Wikipedia*, Mar. 01, 2013. https://en.wikipedia.org/wiki/Optical_character_recognition (accessed Oct. 18, 2022).

[15]“Making Predictions with Sequences,” *Machine Learning Mastery*, [Online]. Available: https://machinelearningmastery.com/sequence-prediction/

[16]S. G. Wood, J. H. Moxley, E. L. Tighe, and R. K. Wagner, “Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis,” *Journal of Learning Disabilities*, vol. 51, no. 1, pp. 73–84, Jan. 2017, doi: 10.1177/0022219416688170.

[17]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. (accessed Oct. 18, 2022).

[18]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. (accessed Oct. 18, 2022).

[19]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 | Proceedings of the AAAI Conference on Artificial Intelligence," *Handwritten Mathematical Expression Recognition via Attention Aggregation Based Bi-directional Mutual Learning | Proceedings of the AAAI Conference on Artificial Intelligence*, Jun. 28, 2022. https://ojs.aaai.org/index.php/AAAI/article/view/19885 (accessed Oct. 18, 2022).

[20]M. G. Rickey and F. L. Bein, "Students' Learning Difficulties in Geography and Teachers' Interventions: Teaching Cases from K-12 Classrooms," *Journal of Geography*, vol. 95, no. 3, pp. 118–125, May 1996, doi: 10.1080/00221349608978703.

[21] van leendert, A. Doorman, P. Pel, and M. Drijvers, "Teachers' Skills and Knowledge in Mathematics Education for Braille Readers," *Technology, Knowledge and Learning*, 2021. 1-22. 10.1007/s10758-021-09525-2. (accessed Oct. 18, 2022).

[22]"Accessible LaTeX Based Mathematical Document Authoring and Presentation," *ACM Digital Library*. https://dl.acm.org/doi/abs/10.1145/3290605.3300734?casa_token=BbQGbRs-LMIAAAAA:H_kHF3wUb6iQMdAA1YekgjCmAVsrk_2Up0DZc8rohGucqMR3GWxMXNjR6Vs-ISf6WTKcqaL3frCkpKs (accessed Oct. 18, 2022).

[23]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 | Proceedings of the AAAI Conference on Artificial Intelligence," *Handwritten Mathematical Expression Recognition via Attention Aggregation Based Bi-directional Mutual Learning | Proceedings of the AAAI Conference on Artificial Intelligence*, Jun. 28, 2022. https://ojs.aaai.org/index.php/AAAI/article/view/19885 (accessed Oct. 18, 2022).

[24]"Handwritten Mathematical Expression Recognition: An approach on data augmentation," *Handwritten Mathematical Expression Recognition: An approach on data augmentation*. https://ieeexplore.ieee.org/abstract/document/9668219?casa_token=ohQa3by7lgEAAAAA:NQ2llPnJOAj2A4JPzyf5wFkcV4oWUJ-PnqVRWjKpsGWvmBVPSprIl6Nl8jzfbrH2mDh4IJu0CHbmtvs (accessed Oct. 18, 2022).

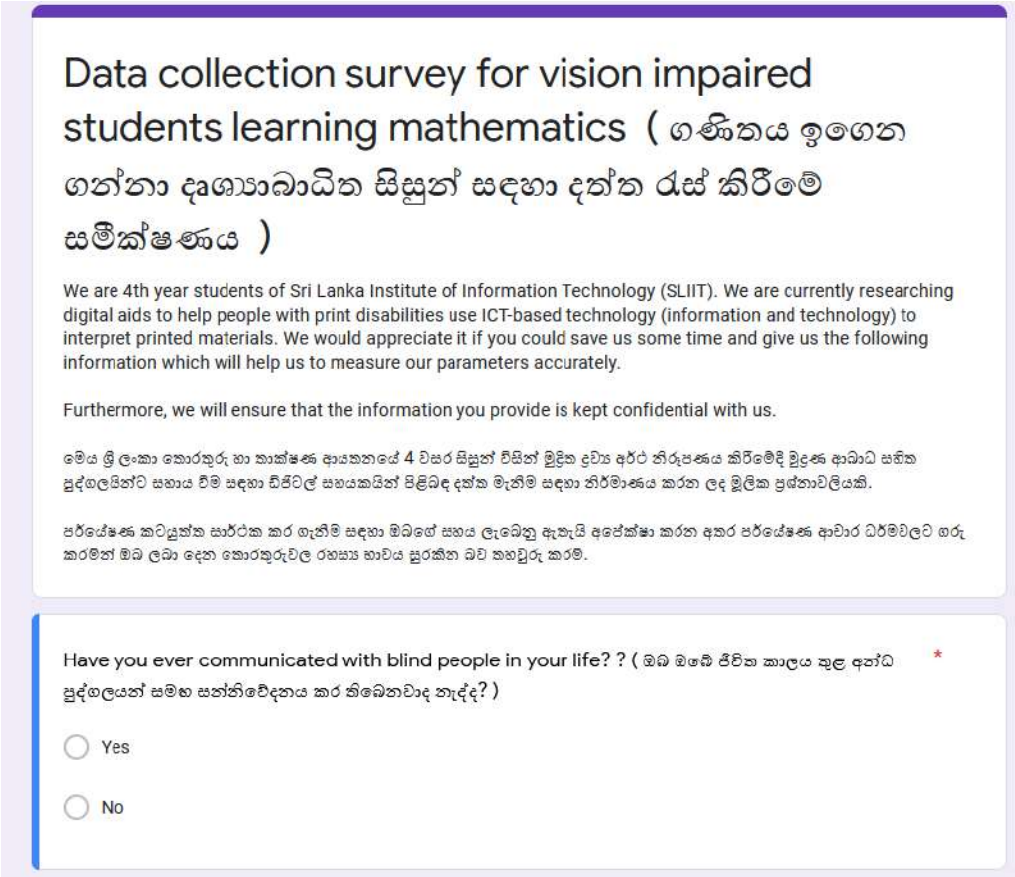
[25]H. Q. Ung, C. T. Nguyen, H. T. Nguyen, T.-N. Truong, and M. Nakagawa, "A Transformer-Based Math Language Model for Handwritten Math Expression Recognition | SpringerLink," *A Transformer-Based Math Language Model for Handwritten Math Expression Recognition | SpringerLink*, Sep. 02, 2021.

https://link.springer.com/chapter/10.1007/978-3-030-86159-9_29 (accessed Oct. 18, 2022).

[26]“Albumentations,” *Albumentations: fast and flexible image augmentations*. https://albumentations.ai/ (accessed Oct. 19, 2022).

APPENDIX

Online Survey



Data collection survey for vision impaired students learning mathematics (ගණිතය ඉගෙන ගන්නා දෘශ්‍යාබාධිත සිසුන් සඳහා දත්ත රැස් කිරීමේ සමීක්ෂණය)

We are 4th year students of Sri Lanka Institute of Information Technology (SLIIT). We are currently researching digital aids to help people with print disabilities use ICT-based technology (information and technology) to interpret printed materials. We would appreciate it if you could save us some time and give us the following information which will help us to measure our parameters accurately.

Furthermore, we will ensure that the information you provide is kept confidential with us.

මෙය ශ්‍රී ලංකා තොරතුරු හා තාක්ෂණ ආයතනයේ 4 වසර සිසුන් විසින් මුද්‍රිත ද්‍රව්‍ය අර්ථ නිරූපණය කිරීමේදී වුදෙණ ආබාධ සහිත පුද්ගලයින්ට සහාය වීම සඳහා ඩිජිටල් සහයකයින් පිළිබඳ දක්ක මැනීම සඳහා නිර්මාණය කරන ලද මූලික ප්‍රශ්නාවලියකි.

පර්යේෂණ කටයුත්ත සාර්ථක කර ගැනීම සඳහා ඔබගේ සහය ලැබෙනු ඇතැයි අපේක්ෂා කරන අතර පර්යේෂණ ආචාර ධර්මවලට ගරු කරමින් ඔබ ලබා දෙන තොරතුරුවල රහස්‍ය භාවය සුරකින බව කහවුරු කරමි.

Have you ever communicated with blind people in your life? ? (ඔබ ඔබේ ජීවිත කාලය තුළ අන්ධ පුද්ගලයන් සමඟ සන්නිවේදනය කර තිබෙනවාද නැද්ද?) *

- Yes
- No

Did you know what are the learning material of the vision impaired student using ? (දෘශ්‍යාබාධික සිසුවෙකු භාවිතා කරන ඉගෙනුම් ද්‍රව්‍ය මොනවාදැයි ඔබ දන්නවාද?) *

- [ ] Brail documents
- [ ] Printed materials
- [ ] Audio books
- [ ] Human reader
- [ ] Other

If you know other materials using the vision impaired student using, please let them know below? (දෘශ්‍යාබාධික සිසුන් භාවිතා කරන වෙනත් ද්‍රව්‍ය ඔබ දන්නේ නම්, කරුණාකර පහතින් ලියන්න.)

Short answer text

Do you think vision impaired students have enough resources to study? (දෘශ්‍යාබාධික සිසුන්ට ඉගෙනීමට ප්‍රමාණවත් සම්පත් ඇතැයි ඔබ සිතනවාද?) *

- ◯ Yes
- ◯ No

Have you ever tried listening to audio books? (ඔබ කවදා හෝ ශ්‍රව්‍ය පොත්වලට සවන් දීමට උත්සාහ කර තිබේද?) *

- ◯ Yes
- ◯ No

Are you satisfied with listening to audio books?

	1	2	3	4	5	
low	◯	◯	◯	◯	◯	High

Do you think it is efficient to use Braille for math? (ගණිතය සඳහා බ්‍රේල් භාවිතා කිරීම කාර්යක්ෂම බව ඔබ සිතනවාද?) *

- ◯ Yes
- ◯ No
- ◯ Maybe

If you say 'yes', what are the reasons for it? ('ඔව්' කිව්වොත් ඒකට හේතු මොනවද?)

Short answer text

Did you familiar with the using of the vision impaired helping apps? (දෘශ්‍යාබාධිත උපකාරක යෙදුම් භාවිතය ගැන ඔබ හුරුපුරුදුද?) *

- Yes
- No

If you say 'Yes', please let know the what are the apps? (ඉහත ඔබ 'ඔව්' කිව්වොත්, කරුණාකර දුරකතන යෙදුම් මොනවාදැයි දන්වන්න?)

Short answer text

Did you ever used the text-to-speech (TTM) feature to read the printed materials? (මුද්‍රිත ද්‍රව්‍ය කියවීමට ඔබ කවදා හෝ the text-to-speech (TTM) විශේෂාංගය භාවිතා කර තිබේද?) *

- Yes
- No

Plagiarism Checker

ORIGINALITY REPORT

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

PRIMARY SOURCES

1	Submitted to Sri Lanka Institute of Information Technology Student Paper	3%
2	www.arxiv-vanity.com Internet Source	1%
3	Sumeet S. Singh, Sergey Karayev. "Chapter 4 Full Page Handwriting Recognition via Image to Sequence Extraction", Springer Science and Business Media LLC, 2021 Publication	1%
4	cvhci.anthropomatik.kit.edu Internet Source	<1%
5	dl.lib.uom.lk Internet Source	<1%
6	"South-East Asia Eye Health", Springer Science and Business Media LLC, 2021 Publication	<1%
7	www.grin.com Internet Source	<1%