KIDS BUDDY: APPLICATION FOR SLOW LEARNING CHILDREN, WITH FLEXIBLE LEARNING METHODS

Project Id: 2022-267

Raveena Sachini Amarasiriwardena

(IT19035918)

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

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2022

.

KIDS BUDDY: APPLICATION FOR SLOW LEARNING CHILDREN, WITH FLEXIBLE LEARNING METHODS

Project Id: 2022-267

Raveena Sachini Amarasiriwardena

(IT19035918)

Dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science in Information Technology Specializing in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2022

·

DECLARATION

I declare that this is my own work and this dissertation 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 my 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. Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Student Name – Raveena Sachini Amarasiriwardena

Registration Number – IT19035918

Signature: …………………………… Date: 2022/09/09… …

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

………………………………… Date: …………………….

Signature of the Supervisor

(Ms. Rrubaa Panchendrarajan)

………………………………… Date: …………………….

Signature of the Co-Supervisor

(Ms. Thilini Buddhika Jayasingha)

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere thanks to my parents for believing in me and supporting me throughout in my degree, as well as for supporting me emotionally, mentally and financially during the challenges I faced throughout my degree.

I extended my deepest gratitude to our supervisor, Ms. Rrubaa Panchendrarajan, for the excellent direction and encouragement she has provided to me and my team. Additionally, it gives us great pleasure to express our gratitude to our co-supervisor, Ms. Thilini Buddhika Jayasingha, for his tremendous support and direction during the entire research process. I would like to give my gratitude to my follower colleagues Galikha Mannage, Piumi Ranawansha and Piyumi Fernando for the cooperation and continuous encouragement which helped me immensely in this project. I would like to thank Sri Lanka Institute of Information Technology for providing research materials to guide us throughout the research journey. In addition, I would like to thank Mr. Jagath Wikramarathne, who is the current lecturer-in-charge of the Research Module who guided us during each phase of the project.

Additionally, I like to express our sincere gratitude to Dr. Nayana Henwelagedara, Consultant Paediatrician, Kandy National Hospital and Mrs. Sandya Rajapakshe, teacher at WP/KL/Medhananda Vidyalaya Kalapugama, Moronthuduwa, for providing their necessary guidance and support during this research.

Last but not least, everyone who supported this research is gratefully acknowledged.

Raveena Sachini Amarasiriwardena
Department of Computer Science and Software Engineering,
Faculty of computing,
Sri Lanka Institute of Information Technology.

.

ABSTRACT

There are various stages that children go through while they learn new things. Every child is unique, and they do not all have the same learning skills. Early childhood education and good guidance are crucial for a child's development. When compared to a normal child, slow learners are lacking in fundamental skills, according to interviews with parents, teachers, and doctors. Lack of object recognition is the main deficiency/weakness that we found among those slow learning children while conducting these interviews. Sri Lanka is a country with 75% of Sinhala native speakers. Most children in Sri Lanka find learning disabilities within themselves. The main purpose of this research is to develop a mobile application in Sinhala for slow-learning children in Sri Lanka whose age is between 4-6 years. Under the object recognition objective, this system will identify the objects and generate a description about the identified objects in Sinhala. Technologies like deep learning, YOLOv4 and image processing are used to create this feature and as for the tools, TensorFlow lite is used. The purpose of this overall system is to help the slow-learning child to overcome their learning disabilities and improve their basic education.

Key words: Deep Learning, Image Processing, YOLOv4, Slow learning children, Darknet

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

LIST OF ABBREVIATION

Abbreviation	Description
YOLO	You Look Only Once
UI	User Interface
DL	Deep Learning
COCO	Common Objects In Context
CNN	Convolutional Neural Network
API	Application Programming Interface
Fast R-CNN	Region-based convolutional neural network

.

1. INTRODUCTION

Basic education is the initial step of every human being's learning process. This is because, if a child does not get proper education in their early stage of the child development it causes an issue in the education process of their future. Among those children, slow- learning children need special attention to improve their learning skills. This paper discusses the system which is helpful to slow-learning children are age d between 4-6 years in Sri Lanka. This chapter details the background problem, research objectives, research gap and literature review of the research work.

1.1 Background

At birth, a child's brain has already expanded to 25% of its mature size, according to the process of human brain growth. The brain has grown to 80% of adult size by the time a child is 3 years old [1]. When the child reaches the age of 5 years, the brain had reached to 90% of adult size [1]. After passing those two stages, a child's curiosity, and drive to learn more about the world all grow. A child's desire to learn new things and their natural curiosity aside, every child struggles with the process of learning. Some children aptitude for learning is not average, but rather, below average. So, in their conversations, mathematical skills, pronunciation, object recognition may be wrong/faulty/lacking in these children (Age 4 - 6).

Before they turn three, children begin to recognize items, and when they turn three, their object identification abilities are more advanced than they had previously been. The ability of a child to observe their environment begins to develop, and they become fascinated by items in their surroundings. When a child recognizes an object, they will attempt to name it on their own, whether the name is accurate or not. They sometimes mispronounce the correct name because children do not memorize the object names correctly at once. Therefore, it is crucial to use the right names when referring to the objects.

Sri Lanka consists of 75% of Sinhala and 24% of Tamil native speakers [2]. Those languages are most Sri Lankan children's native languages since their birth. Children attempt to speak and read in those languages before beginning their primary education.

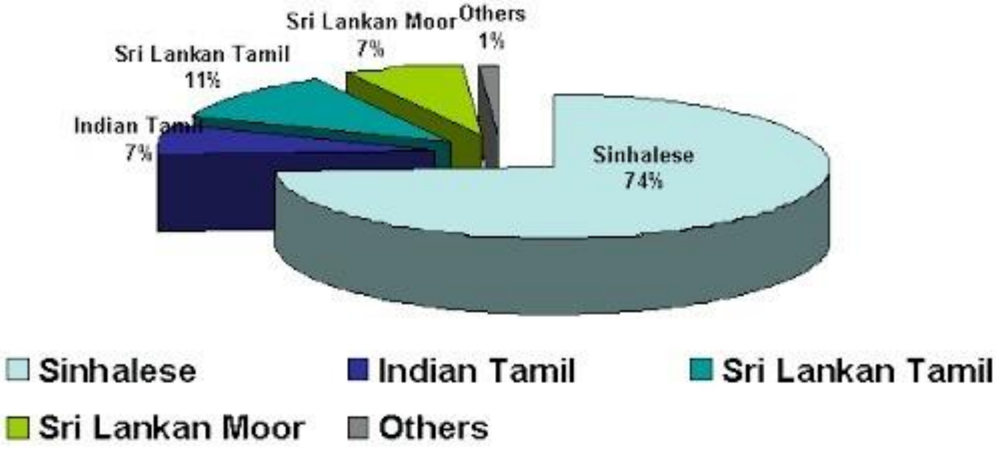


Figure 1 - Ethnic composition of Sri Lanka

There are some existing systems, which are developed for normal children to improve their learning skills, but those applications do not include the 4 main learning skills which we are intending to improve among slow-learning children. For example, in Sri Lanka there is an application call ‘Hapana’ which only includes mathematical quizzes for normal children. Next, if we look at the ‘Buddy.Ai’ app, which was developed in England, it only includes the pronunciation trainer and the chatbot. In addition to that, we specially developed this application in Sinhala, because in the Sri Lanka mobile application market there is no such application implemented for slow learning children are aged between 4-6 years. Before starting their proper primary education, children begin their learning process since birth. Normally when teaching a child there always needs to be a person who will guide the child to correct their mistakes or to help them when they face difficulties. However, with this application, parents or responsible people who look after the children do not have to be with the children while they learn from this application.

Object recognition is an essential skill that every child should enhance. Children’s object recognition skills will be improved with their increase of memory. The main difficulty that those slow learning children face is their lack of memory or their lack of skill when trying to recall the things that they have already learned. This research section discusses object recognition and description generation of the recognized object and how these functions assist slow learning children to improve their object recognition learning skills.

This application will save the child’s and parents’ time when they are learning new things. This is because the main stakeholders who use this application do not have to assign specific times for improving their learning skills and they can allocate the time according to the child’s availability. So, we found that the developed system is very helpful to the slow-learning children and their parents. The following sections of this report explain the literature review of the related research, research gap and research problems, methodology of the research and the testing and discussion part, as well as the conclusion which will conclude the overall research and future work.

1.2 Literature Review

By analyzing and conducting the literature survey for existing research works with similar functionalities and technologies, knowledge can be gained about the research impact, research problems and research gaps. Some of the prominent research works are reviewed here.

Deep learning has achieved considerable improvements in the field of object identification over the last few years [3] and it is slowly developing rapidly across many disciplines, providing concepts for various object recognition. Deep learning object detection methods can be classified into two categories for a given image. The first kind is the two-stage region-based region generation target detection method. In this class, the algorithm first creates a number of candidate frames before categorizing the goals in the candidate frames. These networks are R-CNN [4], Fast R-CNN [5], and Faster R-CNN [6].

Despite having good accuracy, the approaches are nonetheless slow and challenging to utilize in real-time detection environments. One-stage target detection is the second category of regression-based approaches. The object localization is finished in this one-stage object detection while the object classification is looked at. Both the SSD and YOLO series [7, 8] are important networks. This method class can respond to requests instantly and has a high recognition rate. Using quicker R-CNN, YOLO v3 and v4 deep learning algorithms, object identification patterns are produced in typical scenarios.

The three new results on the same dataset are merged, and the variations in item identification performance are studied [10] to determine the optimum approach to the object recognition problem in difficult and intricate situations. We can implement the YOLOv4 model with the number of images we decide to use. Convolutional layers will make up the whole network that Yolo constructs. Using 75 convolutional layers, YOLO is built to take hop connections and sample layers into consideration [11].

The open-source framework for Darknet was C and CUDA. In addition to supporting both CPU and GPU processing, it is speedy and convenient. In each bounding box, YOLOv3 uses logistic regression to predict the object score state [12]. Furthermore, a bounding box surrounds every real item to contain it.

A caption for an image is created using an image caption generator. Case structure, action hierarchy, and verb patterns were employed by A. Kojima [13] to create captions for human activities in a fixed setting. A method for creating image captions by P. Hede [14] uses a list of object names kept in dictionaries as a database to produce captions for preset image content, but it is unable to provide captions for real-world events. Deep neural networks are used to produce captions in a system that was proposed in [15]. A. Farhadi [16] suggested an information retrieval-based method for creating image captions, in which each object in an image receives a score, which is then matched to scores from other photos to produce captions. A ranking-based picture captioning system was presented by M. Hodosh [17], in which the captions are produced using a sentence-based image captioning ranking system. To construct a semantic sentence, Y. Yang [18] presented a sentence-making technique that uses verbs, nouns, and prepositions. The image is identified using trained detectors, and English corpus was used for the estimates of the image. To express the visual significance of the image, R. Socher [19] proposed recursive neural networks based on decision trees. To produce captions for an image, O. Vinyals [20] presented a generative model that integrates computer vision and machine learning. A model of semantic attention that deals with the semantics contained in a hidden layer of neural networks and fuses them to obtain more semantically rich sentences was presented by Q. You et al. [21].

Deep Learning is utilized for object detection and identification in images. For feature extraction, a convolution neural network (CNN) technique is applied [22]. In this scenario, deep learning models can produce an image caption [23]. However, there is a problem with LSTM; rather than considering the content of the image, the next anticipated word in the caption mostly depends on the last predicted word. To get around this, a modified LSTM cell with an extra gate (read-only-unit) can provide captions that are more accurate. The loss value of the LSTM was 2.85 while the loss values for the read-only unit, RNN, and LSTM cell were all 1.85. [24] Zero-Shot Learning is applied. It employs Long-Short Term Memory (LSTM), which makes use of captions from training data and a word embedding that has been learned on external corpora. Three modules make up this system. The first module is an algorithm for creating image captions. The second module is an algorithm for detecting objects. The third module is a semantic word embedding. On sentences, we use transfer learning CNN, and using a deep Fisher kernel, we extract the picture representation [25]. Every extracted activation is aggregated to Fisher vectors in the Fisher kernel. Finally, the vectors are combined with MPP to create a final vector, and they concentrated on improving the representation of a picture to exceed current caption generating algorithms. The sentence embedding in this paper should be able to gather information from diverse descriptions by scanning from left to right and top to bottom [26]. This determines how robustly one may comprehend picture detail.

1.3 Research Gap

Getting the advanced and updated impacts of the new technologies is a significant concern today. Technology advancement plays a big function in making daily tasks easier. One of the most well-known fields, smart technology, is a significant contributor to a variety of applications that are used in the modern world. In this scenario, the functionalities that are perform by the relevant applications are used to rank the applicants. This research focuses on image processing and Deep Learning (DL), two recently developed technologies that have been used for object detection.

To rank the applicants in modern world, certain applications are used. Applications like Buddy.Ai and Sinhala Kids, can be rated based on a variety of attributes and are available on the information technology market.

Features	Buddy.Ai	Our System [Kids Buddy]
Identify different objects in different categories (Animals, v egetables, fruits, etc.)	No	Yes
Describe the identified object	No	Yes

Table 1 - Comparison between related products

There were various kinds of object recognition models after examining and evaluating the preceding research articles. The studies that were reviewed focused on several object categories (Example: Animals, Birds, Vegetables, Fruits, Vehicles etc.). Those studies claim that there are no specific features to present information about the recognized object. Children who struggle to learn are the target age group for our study, which is 4 to 6 years. Their inability to remember or memorize an object after seeing it a number of times is their biggest obstacle to object recognition. The object must be seen and recognized again until it is ultimately ingrained in their memory. There has not been any research done in this field to complete this objective. Therefore, our suggested solution includes an object recognizer component to close this gap. This component will offer an object recognizer model, which will be able to recognize objects across all categories and provide a description of the recognized object.

1.4 Research Problem

The goal of this research is to develop a system that addresses problems that slow-learning children face when they are improving their learning skills. This chapter will describe the research problem that slow-learning children may face in their object recognition skills. In the present world, countless objects exist. To learn those objects they must learn and memorize the objects by name. So, slow leaning children must learn those object names slowly according to their speed of learning. To do that, children must see those objects several times and memorize the object names for identifying those objects by themselves. According to the process, it will take time for those slow-learning children to identify the object by themselves. But later this process will help them to identify the object separately.

There are some mobiles apps available on the software market to improve learning skills among children. The existing applications do not focus specially on slow learning children, but they are generally focused on all children.

The existing applications have some of the features, that were mentioned in the above sections. This study aims to resolve several problems that were seen in existing systems.

Below are some research problems considered when carrying out this research.

- Throughout the object detection procedure, there are some opportunities for error. There is a possibility that the description will not be generated correctly for the object.
- Actual description is essential when creating the description.
- When object does not exist in the dataset, it is important to generate an error message to the child

We create an application to support slow learners in enhancing their learning abilities to provide solutions to these issues. The child can periodically assess their learning capacity by using this system.

1.5 Research Question

- How to identify different objects separately?
- How to categorize the objects?
- How to collect different object data and how to store it?
- How to use image processing for object identification?
- What is the most suitable deep learning approach for object detection?

1.6 Research Objectives

1.6.1 Main Objective

The main objective is to build a mobile application for the learning process of slow-learning children aged between 4 to 6. Through the application, those children get an opportunity to enhance their object recognition skills in the Sinhala language. Children can select an area that they need to develop and practice with the application. This research is based on how we had collected the necessary data and train the object recognition model using YOLOv4 model and apply that model to identify the objects that child the has included.

1.6.2 Specific Objectives

1. Train the Model

The YOLOv4 object detection model will be trained using the COCO dataset. 80 different object classes are included in COCO datasets.

2. Provide information about the identified object.

The following stage after finding the object is to show the Sinhala description of the already found identified object. Google and Wikipedia API are used to produce object descriptions.

3. Develop a simple UI to use the system user-friendly and respond quickly.

Child can interact with the system by inserting necessary details to get the results within a short period of time.

·

2. METHODOLOGY

The relevant functions of the system are handled and implemented according to the methodology that is described in this chapter. The methodological approach to research, requirement gathering, designing, and execution is covered in this chapter. It is crucial to choose the right tools and technologies for implementation based on literature reviews. Object detection feature is based on the YOLOv4 model, and the below list mentions the main parts in this component, which will be elaborated upon in the next few chapters.

a. Object identification
b. Sinhala description generation for the identified object
c. Display error message for the non-identified object

Every component depends on each other component, and those components are implemented using various techniques and algorithms.

2.1 System Overview Diagram

A system diagram is a representation of the system that shows the parts and how they work together, along with any supporting materials. The total system planning is shown in the system diagram below. The diagram shows the four components in the e-learning system which was developed for children who struggle with learning (ages 4-6). These unique functions serve as the answers to some of the issues that slow learners in e-learning systems encounter. The various functionalities for each module in Figure 2 below are designed to enhance slow learners' learning capacity.

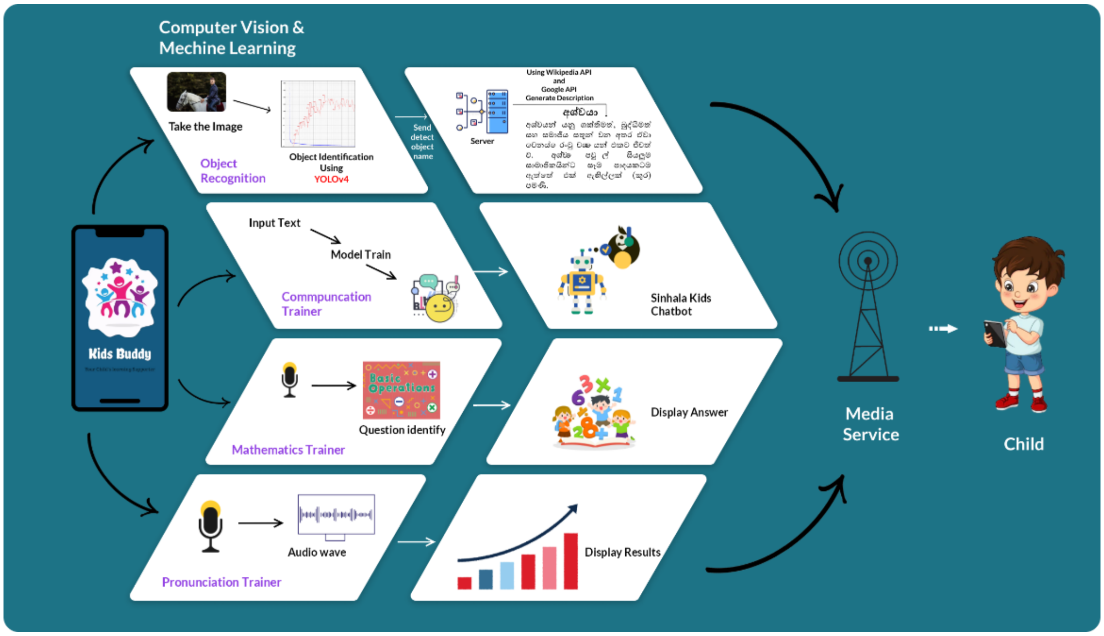


Figure 2 - System Overview Diagram

In this research paper, we have to discuss only the part up to the object detection and description generator. The below system diagram is for the specific part that contains the components and interactions between them. The following functional diagram clearly shows what are the steps that need to be done in order to identify the object and generate a description about the identified object.

2.2 Functional Overview Diagram

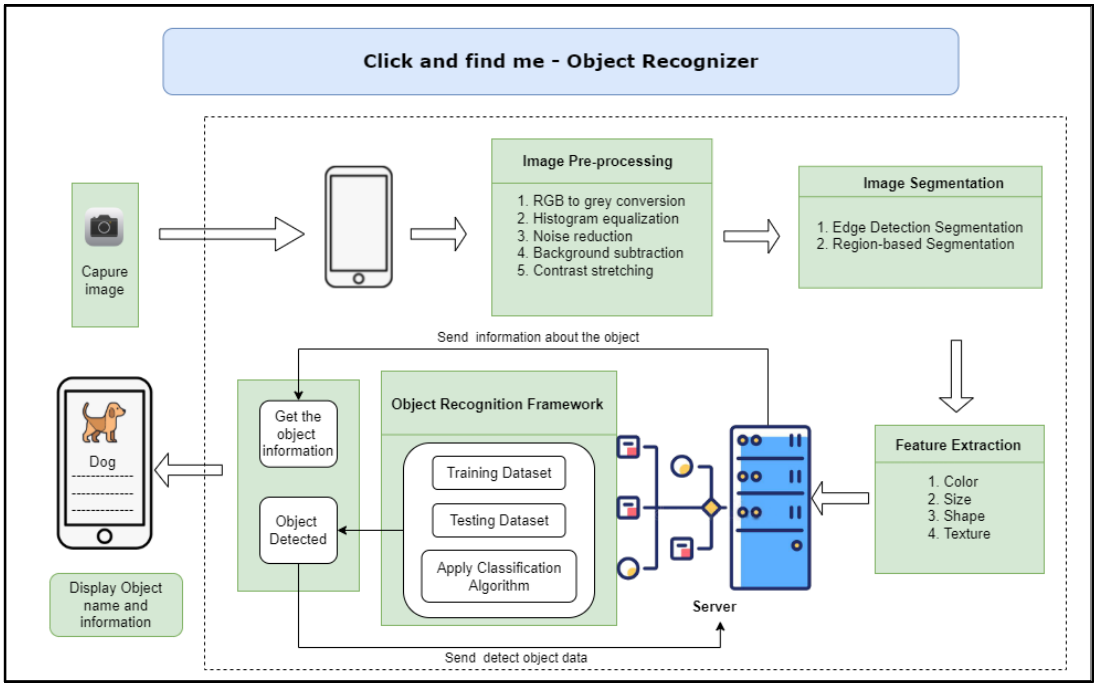

Click and find me - Object Recognizer
Capure image
Image Pre-processing
1. RGB to grey conversion
2. Histogram equalization
3. Noise reduction
4. Background subtraction
5. Contrast stretching
Image Segmentation
1. Edge Detection Segmentation
2. Region-based Segmentation
Send information about the object
Feature Extraction
1. Color
2. Size
3. Shape
4. Texture
Server
Object Recognition Framework
Training Dataset
Testing Dataset
Apply Classification Algorithm
Get the object information
Object Detected
Send detect object data
Dog
Display Object name and information

Figure 3 - Functional System Diagram

The given functional diagram (figure 3) is a visual diagram of the object detection component, which shows the basic functionality to teach the slow learning children to identify the different objects and learn about the identified object with the Sinhala description. In addition, this can be a great opportunity for both teachers and the parents to observe and recognize the learning abilities of slow-learning children.

2.3 Use case diagram

2.3.1 Use case diagram for object detection and description generate

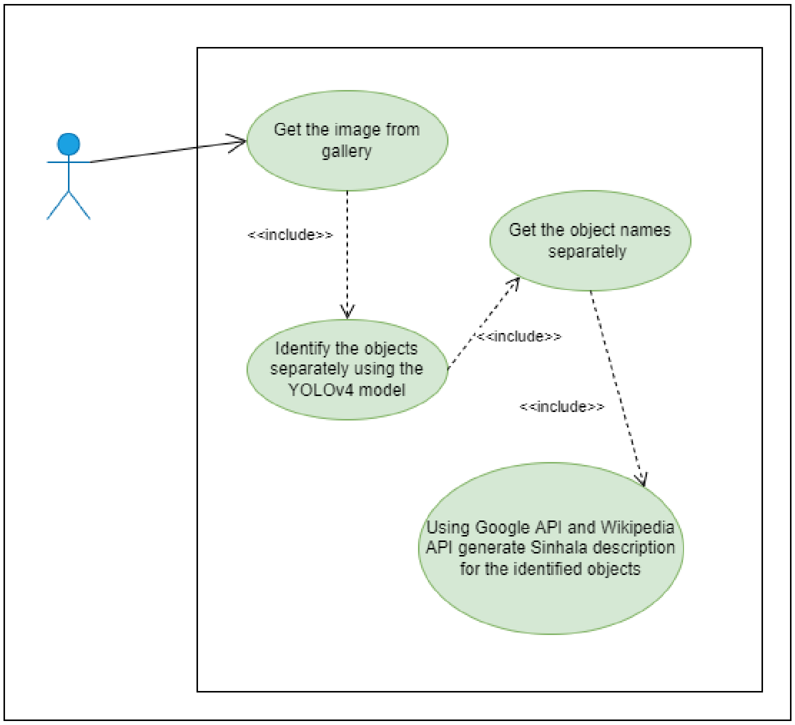

Get the image from gallery
<<include>>
Get the object names separately
<<include>>
Identify the objects separately using the YOLOv4 model
<<include>>
Using Google API and Wikipedia API generate Sinhala description for the identified objects

Figure 4 - Use case diagram for object detection and description generates

2.4 Flow chart

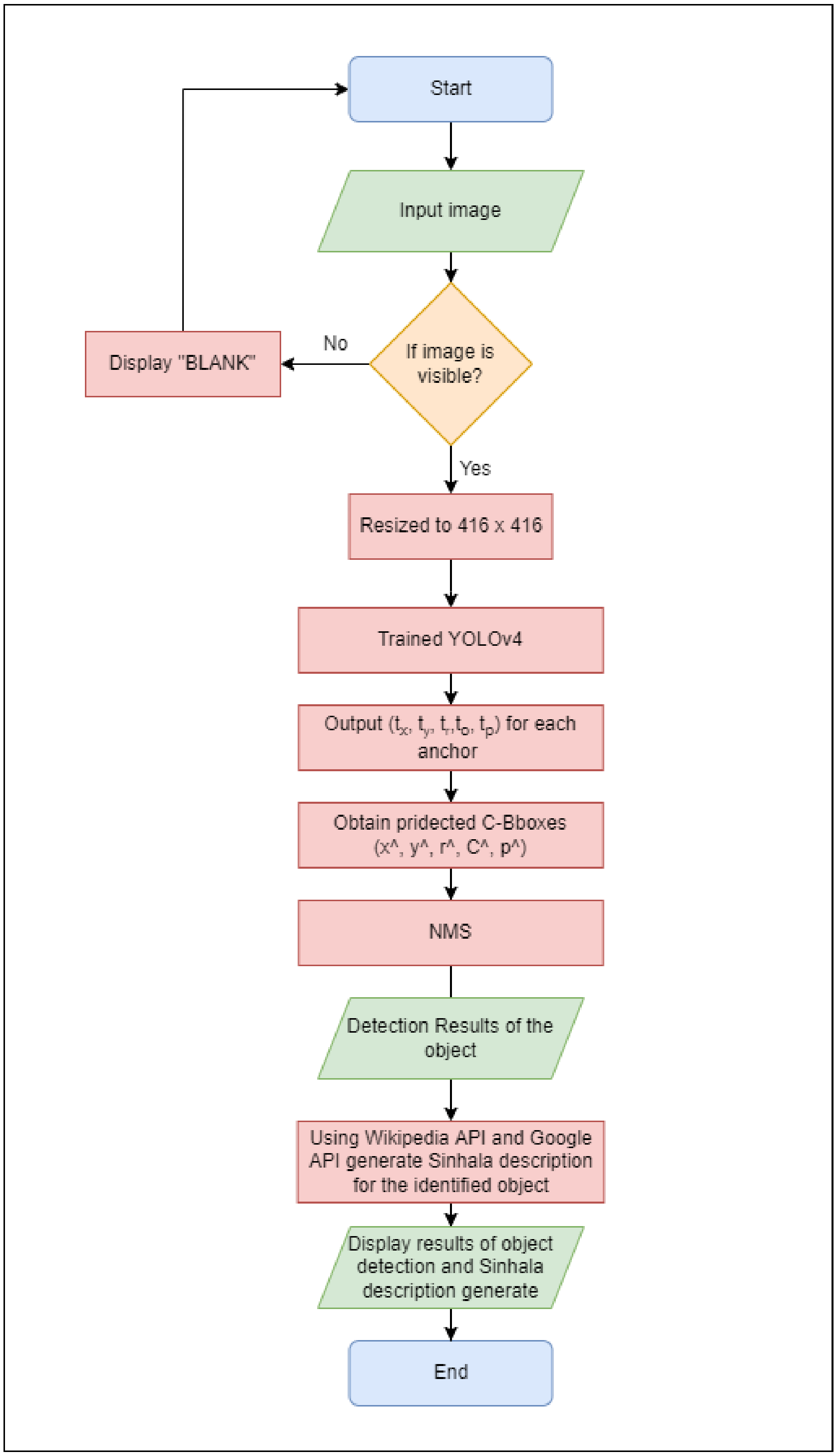

Start
Input image
If image is visible?
No
Display "BLANK"
Yes
Resized to 416 x 416
Trained YOLOv4
Output (t_x, t_y, t_r,t_o, t_p) for each anchor
Obtain pridected C-Bboxes (x^, y^, r^, C^, p^)
NMS
Detection Results of the object
Using Wikipedia API and Google API generate Sinhala description for the identified object
Display results of object detection and Sinhala description generate
End

Figure 5 - Flow chart of the system

2.5 Development Process

The agile SCRUM and the Iteration development model are among the process models that are considered to be the most appropriate for this research because these procedures heavily rely on the outcomes of the phase before them.

The overall project requirements are broken down into various phases by the iteration model. Each iteration of the development process includes the steps of requirements gathering, design, implementation, and testing. This model was chosen because:

This model was chosen because,

- Over time, some of the functionality can change or develop. Before completing the implementations, adjustments must be considered.
- Depending on the needs of the process, new technologies may be implemented during the project's development phase. Therefore, if there is a change like this, it is critical to iterate the steps.

The agile SCRUM and iteration model is considered as the most appropriate approach based on these criteria.

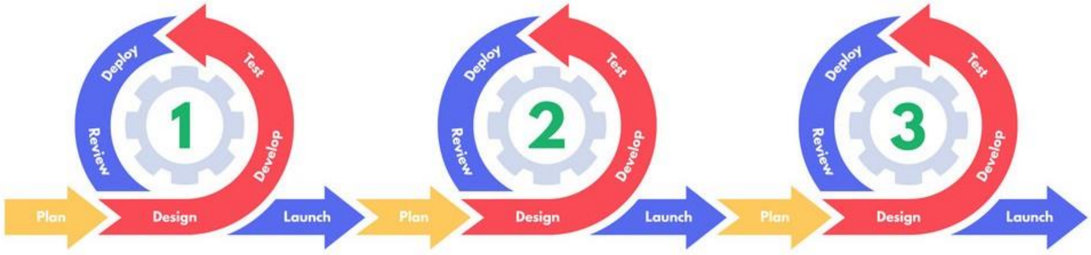


Figure 6 - Agile Scrum process

2.6 Work Breakdown Structure

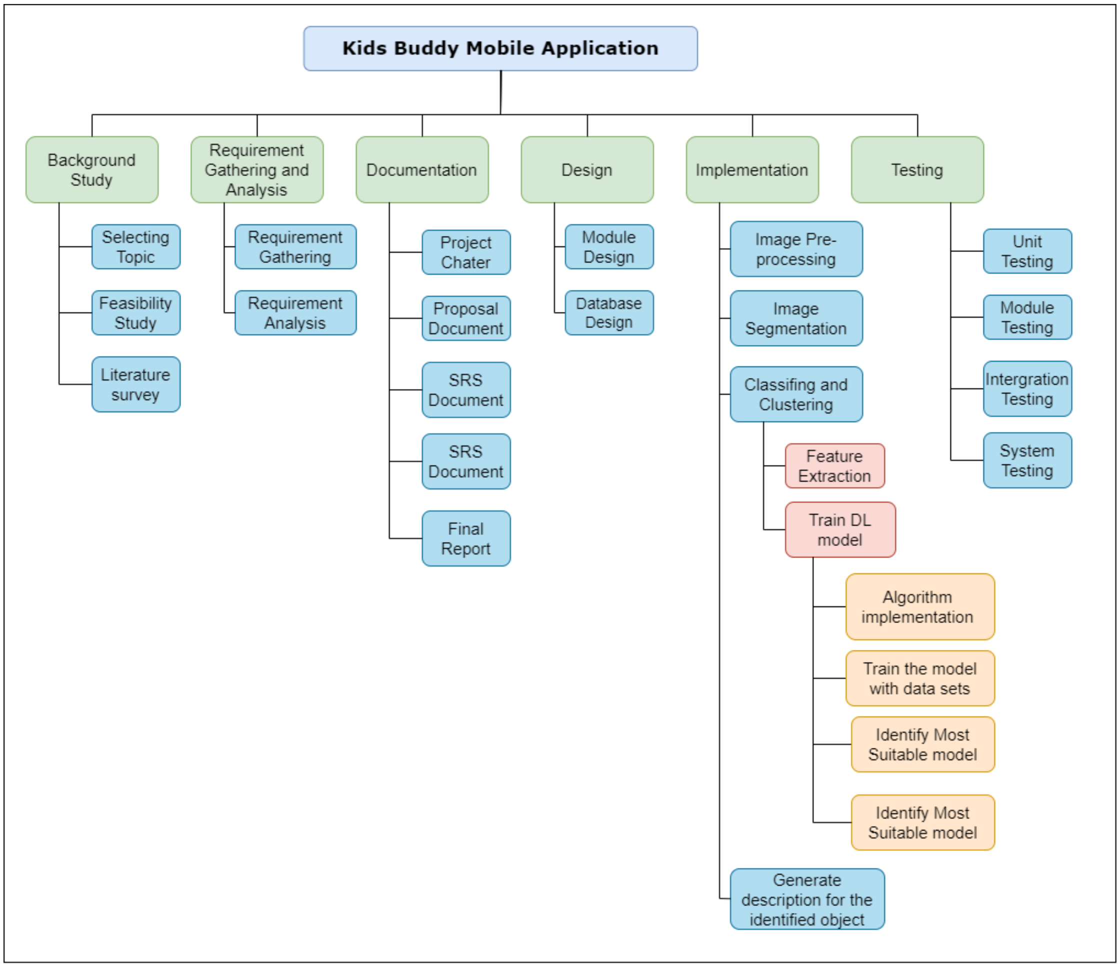


Figure 7 - Object Recognition and description generate Work breakdown Structure

2.7 Feasibility Study

The technological resources required for the project should be identified and considered during feasibility assessment. Technologies such as YOLOv4 (You Look Only Once) configuration used for implementation, and TensorFlow Lite are used for converting the YOLOv4 model to tflite model. In addition, Wikipedia API and Google API are used for Sinhala description generation. Therefore, this research project is almost technically feasible.

The following tools and technologies are used to develop this system.

1. Design Tools - UI design – Figma
   Diagram – Draw.io
2. Implementation - Deep Learning – YOLOv4, TensorFlow Lite
   User Interfaces – Android Studio
   Mobile Application Backend – Android Studio
   Version Control – GitLab
   Task Planning – Microsoft Planner
   Database - SQLite
3. Documentation - Research Paper/ Final Report – MS Word 365
   Presentation – MS PowerPoint 365

2.8 Requirement Gathering

This phase is one of the key steps that must be completed prior to the installation of any system. It is important to investigate and read a lot of the prior research related to this project in order to understand the types of implementations to be done, the technologies used in the previous research, and the research gaps in those works. The best location to find what you need is in research papers. Additionally, Google and relevant users of other platforms are reviewed for the specifications.

- **Literature Review**

  We discussed the research project concept with our previous supervisor, and she assisted us in gathering data from online conferences, papers, and articles that informed us of the project's supporting technology and the topic we would be working on.

- **Dataset**

  A large image dataset called MS COCO (Microsoft Common Objects in Context) contains 328,000 images of everyday objects and people. You can use the dataset's annotations to train machine learning models to identify, categorize, and describe items.

The types of features that MC COCO offers are as follows:

- Object detection - 80 types of objects classes with full segmentation masks and bounding box coordinates
- Adding captions—descriptions of each image in everyday language.
- Key points: The dataset consists of over 200,000 photos of more than 250,000 people that have been annotated with key points such the right eye, nose, and left hip.
- "Stuff image" segmentation consists of pixel mappings of 91 different types of amorphous backdrop areas, such as grass, walls, and skies.
- Panoptic—full scene segmentation, identifying objects in the image in accordance with 91 categories of "stuff" and 80 categories of "things" (cat, pen, fridge, etc.). (Road, sky, water, etc.).
- Dense pose: The dataset contains over 56,000 persons in over 39,000 photos, and each labeled object is annotated with an instance id and a mapping between body pixels for that object and a template 3D model.

The below table 2 shows the COCO (Common Objects in Context) dataset object classes [28, 29].

ID	Object Name	Supper Category
1	person	person
2	bicycle	vehicle
3	car	vehicle
4	motorcycle	vehicle
5	airplane	vehicle
6	bus	vehicle
7	train	vehicle
8	truck	vehicle
9	boat	vehicle
10	traffic light	outdoor
11	fire hydrant	outdoor
12	stop sign	outdoor
13	parking meter	outdoor

14	bench	outdoor
15	bird	animal
16	cat	animal
17	dog	animal
18	horse	animal
19	sheep	animal
20	cow	animal
21	elephant	animal
22	bear	animal
23	zebra	animal
24	giraffe	animal
25	backpack	accessory
26	umbrella	accessory
27	handbag	accessory
28	tie	accessory
29	suitcase	accessory
30	frisbee	sports
31	skis	sports
32	snowboard	sports
33	sports ball	sports
34	kite	sports
35	baseball bat	sports
36	baseball glove	sports
37	skateboard	sports
38	surfboard	sports
39	tennis racket	sports
40	bottle	kitchen
41	wine glass	kitchen
42	cup	kitchen
43	fork	kitchen
44	knife	kitchen
45	spoon	kitchen
46	bowl	kitchen
47	banana	food
48	apple	food
49	sandwich	food
50	orange	food
51	broccoli	food
52	carrot	food
53	hot dog	food
54	pizza	food
55	donut	food
56	cake	food
57	chair	furniture
58	couch	furniture
59	potted plant	furniture

60	bed	furniture
61	dining table	furniture
62	toilet	furniture
63	tv	electronic
64	laptop	electronic
65	mouse	electronic
66	remote	electronic
67	keyboard	electronic
68	cell phone	electronic
69	microwave	appliance
70	oven	appliance
71	toaster	appliance
72	sink	appliance
73	refrigerator	appliance
74	book	indoor
75	clock	indoor
76	vase	indoor
77	scissors	indoor
78	teddy bear	indoor
79	hair drier	indoor
80	toothbrush	indoor

Table 2 - COCO dataset details

There are primarily functional, non-functional, and user requirements in addition to the aforementioned requirement. These requirements are,

Functional Requirement

- Identify the objects
- Describe an identified object
- Provide a save option to save the details about the identified object

Non-Functional Requirement

i) Correctness

- The system should display correct results according to the data
- The system should be run without error/bug-free

ii) Performance

- The app should support slow learning children according to their ability and give a better user experience.
- System response time should be between 1-4 seconds.

.

iii) Reliability

- The mobile application should be available at all times.
- Least down error time.
- The system can be used numerous times.

iv) Security

- Antiviruses Guard
- Network Security
- Google’s built-in malware protection

User Requirement

- Mobile phone with a camara
- Internet Connection

2.9 Resources Used

2.9.1 Soft Boundaries

- Android Studio:

  It is a source-code editor that is freeware and is compatible with Linux, Windows, and macOS. It serves as the IDE for this project's mobile application development.

- GitLab

  A Git repository is made available for the continuous integration built by GitLab Inc. through the web-based DevOps lifecycle tool. Version control and project integration are both used for this project.

- Darknet Framework

  An open-source neural network framework called Darknet was created in C and CUDA. It supports both CPU and GPU computing, is quick, and is simple to install.

2.9.2 Hardware Boundaries

- Memory: 4 GB
- Graphics Card: NVIDIA GeForce GTX 970
- CPU: Intel Core i5-4590
- File Size: 2 GB
- OS: Windows 7, Windows 8, Window 10, Window 11

2.10 Commercialization aspect of the product

When taking into account its potential for the slow-learner children, the mobile application that we suggested offers several benefits. The main goal of implementing a learning management system for the slow learners was to boost their self-esteem and help them succeed in the academic environment. They definitely benefit greatly from it because they have never had access to such a learning environment. As a result, children will not feel as though they are losing out on anything in the community.

The suggested system's application is straightforward and user-friendly, extremely trustworthy, highly accurate, and quick enough to follow the fundamental learning skills. When applying to a high demand area for children with sluggish learning, these traits will have a significant impact. Along with that, we can guarantee that the children who benefit from it will stand out in society and excel because they will not have any obstacles to learning new things or developing new abilities.

This system's commercial worth lies in giving slow learners a platform to advance their fundamental learning abilities. Users of this system can use it for free, and it is a very user-friendly application.

The key characteristics and benefits of this system are summarized in the chart below.

Feature	Advantages
Time	Less time when comparing with manual method. Easy to give the result within a short time period

Error	Less error In the manual method, user can make some mistakes when analyzing result.
Easy to use	Child can insert the answers to the system. The system will output the results
User -friendly	Simple UI that everyone can easily understand.
Accurate	Can get accurate result from the applicants.

Table 3 - Features and advantages of the system

2.11 Implementation

In this chapter, system's implementation of the features, which include the ones listed below, will be explored.

a) Image Classification

b) YOLOv4 Backbone Network - Feature Formation

c) YOLOv4 Neck Network - Feature Aggregation

d) YOLOv4 Head – The detection Step

e) Description generation and generate error message for non-existing object

f) Mobile Application

g) Database

2.11.1 Image Classification

YOLOv4 model was used to implement the DL model. We can process various images using the YOLOv4 classifier and then categorize them as necessary. YOLOv4 was built using convolutional neural networks (CNN) and fast region-based convolutional neural networks (Fast R-CNN). This topic will be covered in the chapter that follows. COCO dataset was used for the object detection and each object will be classified under a class. There will be 80 classes, 11 super classes and 328,000 pictures. In below figure 8 shows the 80 classes and 11 super classes of the COCO dataset. To reduce the cost of sensory device here we used a vision-based approach to focus on images/mobile camara input to gain data from the user end.

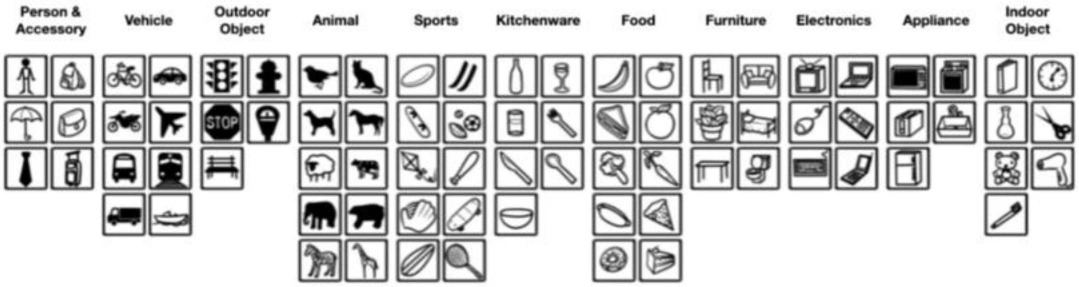


Figure 8 - COCO dataset 80 classes and 11 super classes

The user's data is extracted using a **Gaussian blur filter**, which speeds up preparation. Although there are various techniques, arithmetic mean and Gaussian mean are the most appropriate/suitable for data extraction. The general formula to calculate threshold value, T, is as follows. According to the Gaussian mean, pixel values further from the region's (x, y) coordinate center contribute less to the threshold value computation as a whole. The equation below displays the Gaussian mean.

$$\mathbf{T = mean\ (I_L) - C}$$

Here, T is a constant that we can use to change the threshold value, **(I_L)** is the local sub-region of the image, I, and C.

After applying the filter, the image goes through several steps in the YOLOv4 model to detect an object.

Each object detector takes an image as input and compresses its information using a convolutional neural network as its foundation. Predictions can be based on these backbones, which are the network's central nodes in image segmentation. Since object recognition requires several bounding boxes to be built around images in addition to classification, the feature layers of the convolutional backbone must be combined and held up in comparison to one another. The backbone's layers come together at the neck.

One-stage detectors and two-stage detectors are the two categories into which object detectors can be subdivided. Detection occurs in the head. Two-stage detectors are used to divide the tasks of object localization and classification for each bounding box.

One-stage detectors make predictions for both object localization and classification at the same time. You Only Look Once refers to the one-stage detector known as YOLO.

Below figure 9 shows the one stage and two stage detectors, in object recognition.

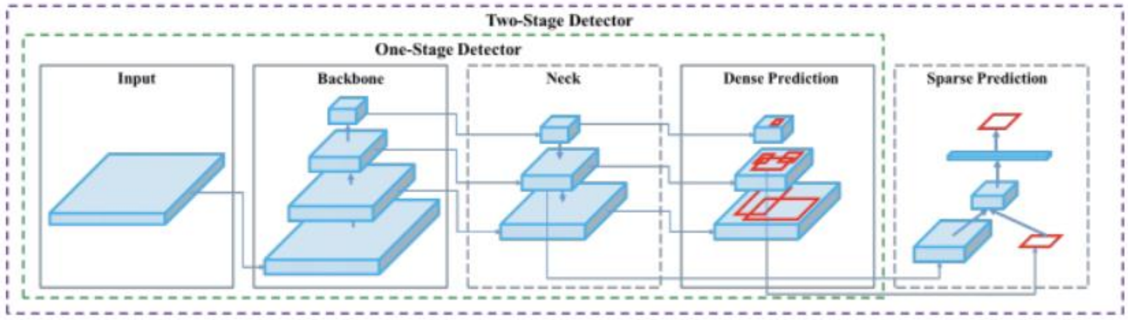


Figure 9 - Object detector (0ne Stage and Two Stage)

2.11.2 YOLOv4 Backbone Network - Feature Formation

Typically, pre-training on ImageNet categorization is given to the backbone network of an object detector. Even when the network has already been trained to recognize key elements in an image, pre-training refers to the process of changing the network's weights for the new goal of object recognition. Backbones for the YOLOv4 object detector include,

- CSPResNext50
- CSPDarknet53
- EfficientNet-B3

The CSPResNext50 and the CSPDarknet53 are both built on top of DenseNet. In order to increase feature propagation, encourage feature reuse, reduce the number of network parameters, and solve the vanishing gradient problem, DenseNet was developed to connect layers in convolutional neural networks.

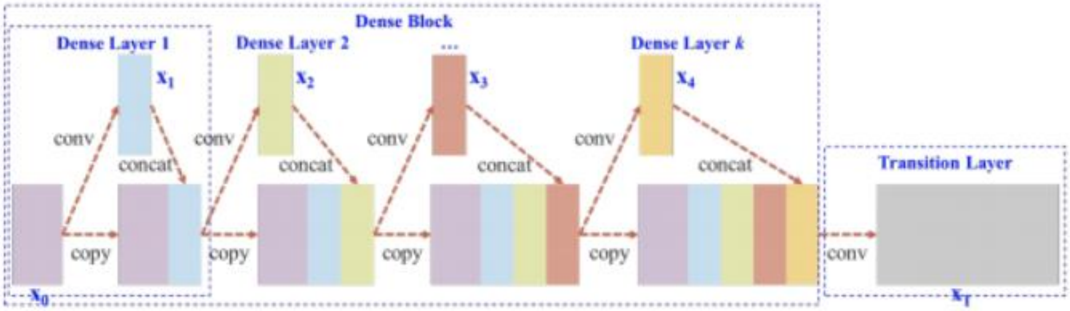


Figure 10 - DenseNet

In terms of image classification, EfficientNet performs better than other networks of equivalent size.

2.11.3 YOLOv4 Neck Network - Feature Aggregation

To be prepared for the detection step of object detection, the characteristics developed in the ConvNet backbone must now be merged and integrated. YOLOv4 considers several different neck scenarios, including:

- FPN
- PAN
- NAS-FPN
- BiFPN
- ASFF
- SFAM

In figure 11 shows the Feature network design in neck network.

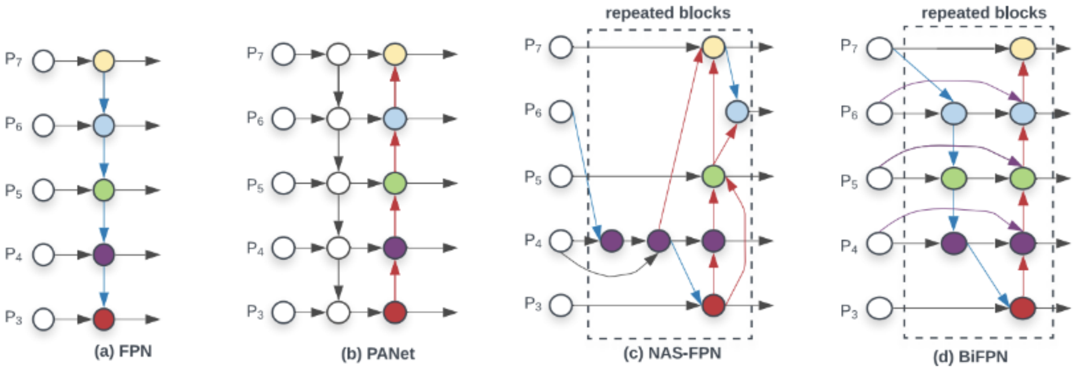


Figure 11 - Feature network design

Each of the aforementioned points to a different feature tier, as determined by the CSPDarknet53 backbone. The image shown above is from YOLOv4's predecessor, EfficientDet. YOLOv4 chooses PANet for the network's feature aggregation.

YOLOv4 adds an SPP block after CSPDarknet53 to increase the receptive area and separate the most important characteristics from the backbone.

2.11.4 YOLOv4 Head – The detection Step

YOLOv4 uses the same YOLO head as YOLOv3 for detection with the anchor-based detection steps and three degrees of granularity. Below figure 12 shows the Dense prediction in YOLOV4 head.

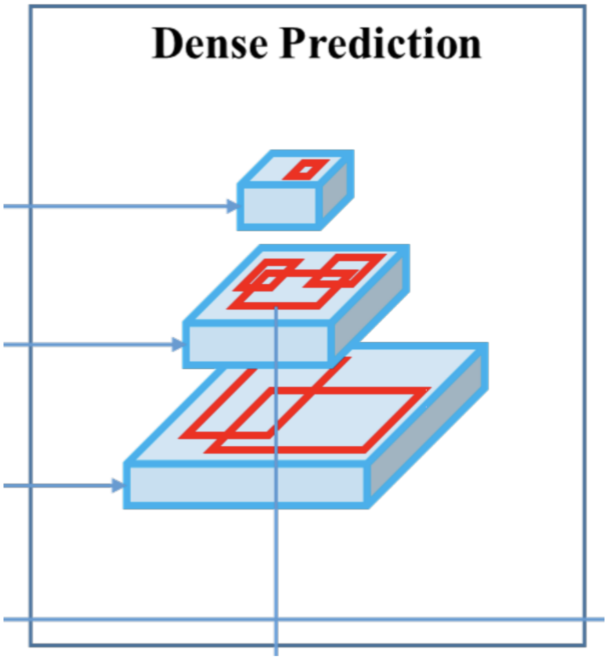


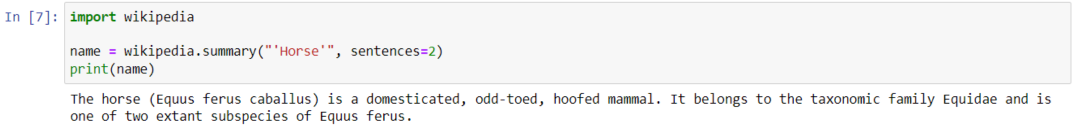
Figure 12 - Dense prediction

The network determines the confidence score and bounding box coordinates for a class. The goal of YOLO is to divide the image into a grid of multiple cells, with each cell employing anchor boxes to estimate the probability that an object will be present. The outcome is a vector containing bounding box coordinates and probability classes. Additionally, post-processing methods like non-maxima suppression are utilized at the conclusion.

2.11.5 Description generation and generate error message for non-existing object

After the object is successfully identified, the model will generate a description about the identified object in English using Wikipedia API. Then using Google API, the model will translate the generated description into Sinhala language.

If model will not identify an object in an image or if the detected image confidence level is less than 20%, then the system will generate an error message to the user. Below figure 13 and 14 shows Wikipedia and Google API calling for the description generation in Sinhala for an identified object.

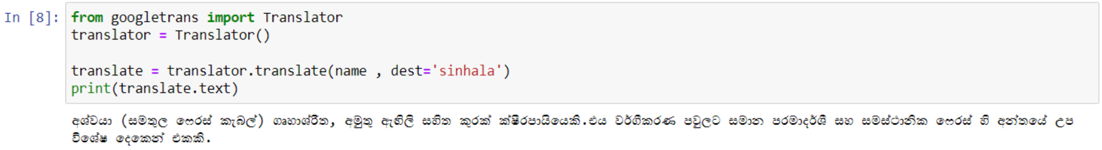


```
In [7]: import wikipedia

        name = wikipedia.summary("'Horse'", sentences=2)
        print(name)

        The horse (Equus ferus caballus) is a domesticated, odd-toed, hoofed mammal. It belongs to the taxonomic family Equidae and is
        one of two extant subspecies of Equus ferus.
```

Figure 13 – Wikipedia API calling for the description generation



```
In [8]: from googletrans import Translator
        translator = Translator()

        translate = translator.translate(name , dest='sinhala')
        print(translate.text)

        අශ්වයා (සමතුල පෙරස් කැබල්) ගෘහාශ්රිත, අමුතු ඇඟිලි සහිත කුරක් ක්ෂීරපායියෙකි.එය වර්ගීකරණ පවුලට සමාන පරමාදර්ශී සහ සමස්ථානික පෙරස් හි අන්තයේ උප
        විශේෂ දෙකෙන් එකකි.
```

Figure 14 - Google API calling for the description generation translation

2.11.6 Mobile Application

The functionalities need to be implemented in mobile application, for the child to use themselves and test it. However, since Sri Lankan children between the ages of 4-6 will be our main audience, we must concentrate on building an engaging user interface (UI) to improve their experience. As a result, we decide to create the frontend using Java and the Android Studio IDE. The UI should be easier to use and more child-friendly, as indicated above. The graphics below show the basic UI designed for object identification and description generation that are created using Figma.

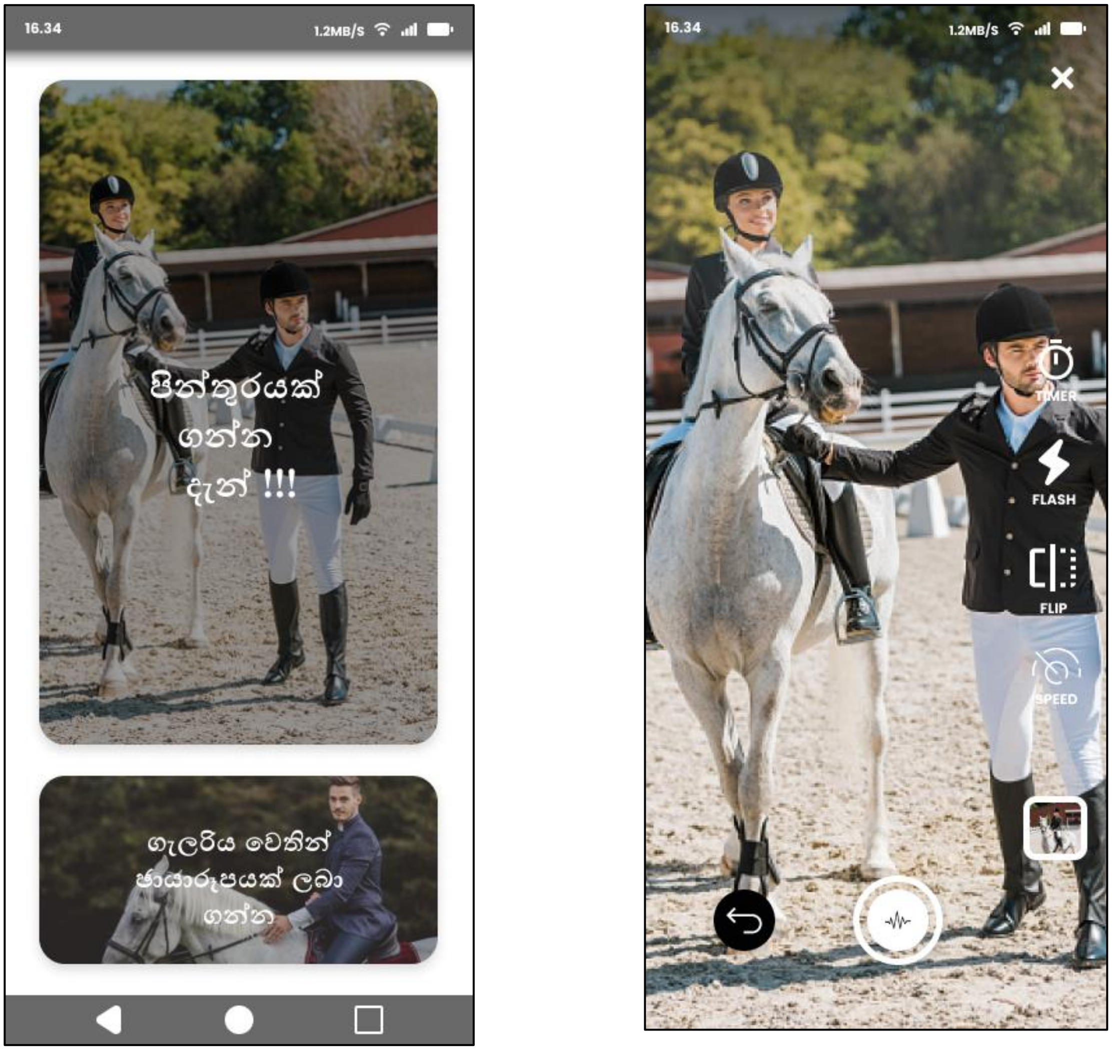
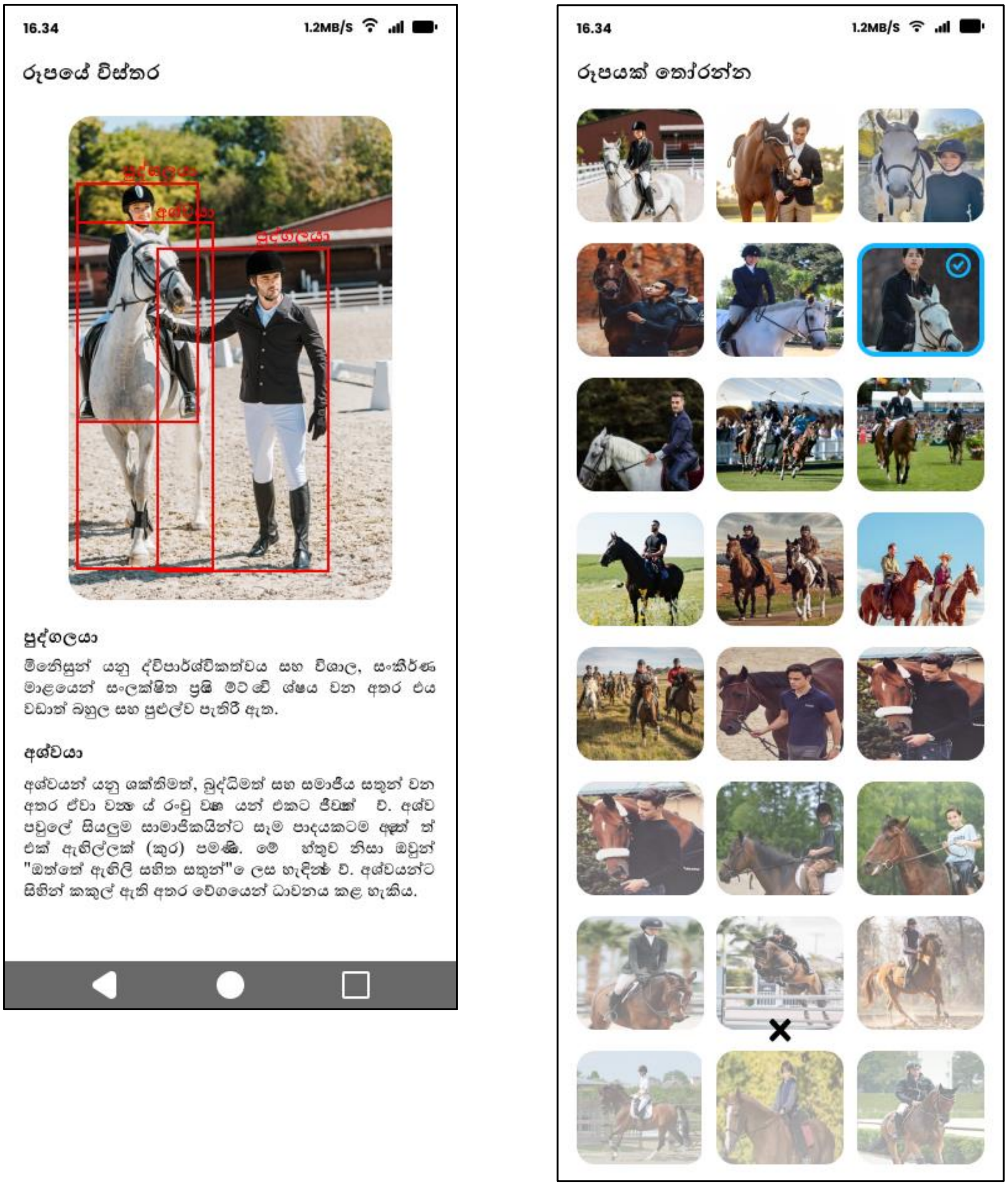


Figure 15 **-** *User interface of the MainClickAndFindMe main page and camara view pages*



රූපයේ විස්තර

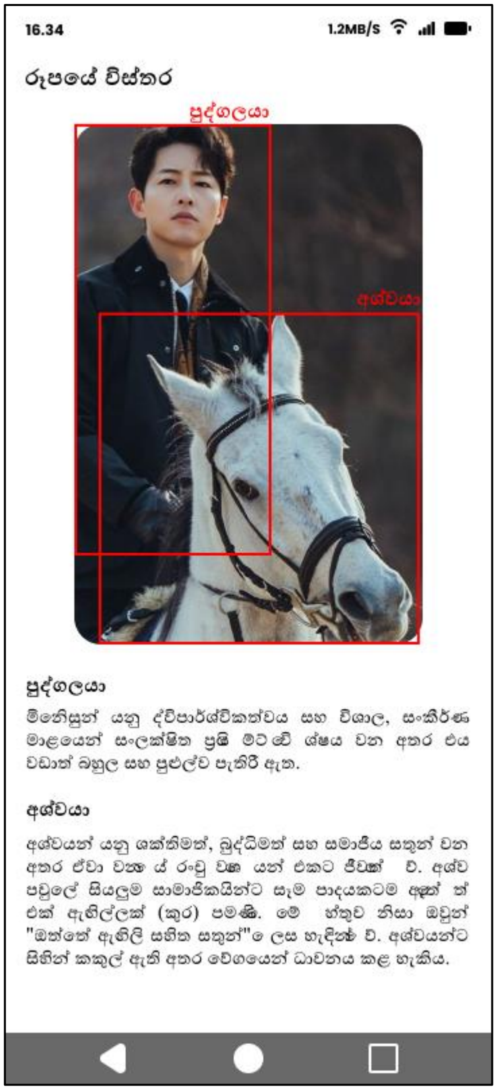
පුද්ගලයා

මිනිසුන් යනු ද්විපාර්ශ්විකත්වය සහ විශාල, සංකීර්ණ මාළයෙන් සංලක්ෂිත ප්‍රෂි මට වේ ශ්ෂය වන අතර එය වඩාත් බහුල සහ පුළුල්ව පැතිරී ඇත.

අශ්වයා

අශ්වයන් යනු ශක්තිමත්, බුද්ධිමත් සහ සමාජීය සතුන් වන අතර ඒවා වන ය් රංචු වෂ යන් එකට ජීවක් වේ. අශ්ව පවුලේ සියලුම සාමාජිකයින්ට සෑම පාදයකටම අඟුත් ත් එක් ඇඟිල්ලක් (කුර) පමණි. මේ හේතුව නිසා ඔවුන් "ඔත්තේ ඇඟිලි සහිත සතුන්"ෙ ලස හැඳින් වේ. අශ්වයන්ට සිහින් කකුල් ඇති අතර වේගයෙන් ධාවනය කළ හැකිය.

Figure 16 - User interface of the Storage image Sinhala description generate page main page and image storage page



රූපයේ විස්තර

පුද්ගලයා

මිනිසුන් යනු ද්විපාර්ශ්විකත්වය සහ විශාල, සංකීර්ණ මාළයෙන් සංලක්ෂිත ප්‍රෂි මට ශේ ශ්ෂය වන අතර එය වඩාත් බහුල සහ ප්‍රළුල්ව පැතිරී ඇත.

අශ්වයා

අශ්වයන් යනු ශක්තිමත්, බුද්ධිමත් සහ සමාජීය සතුන් වන අතර ඒවා වසා ය් රංචු වශ යන් එකට ජීවත් වේ. අශ්ව පවුලේ සියලුම සාමාජිකයින්ට සෑම පාදයකටම අඟුත් ත් එක් ඇඟිල්ලක් (කුර) පමණි. මේ හේතුව නිසා ඔවුන් "ඔත්තේ ඇඟිලි සහිත සතුන්"ෙ ලස හැඳින් වේ. අශ්වයන්ට සිහින් කකුල් ඇති අතර වේගයෙන් ධාවනය කළ හැකිය.

Figure 17 - User interface of the camara image Sinhala description generate page

User interfaces are created based on children perspective because they are the prime users of this application. Dashboard, navigation between pages, buttons, images/content that display on the screen focus on children's perspective for them to learn and work on this application by themselves.

Below figures shows the real mobile application features that are developed in object detection and Sinhala description generation component which are created for slow-learning children.

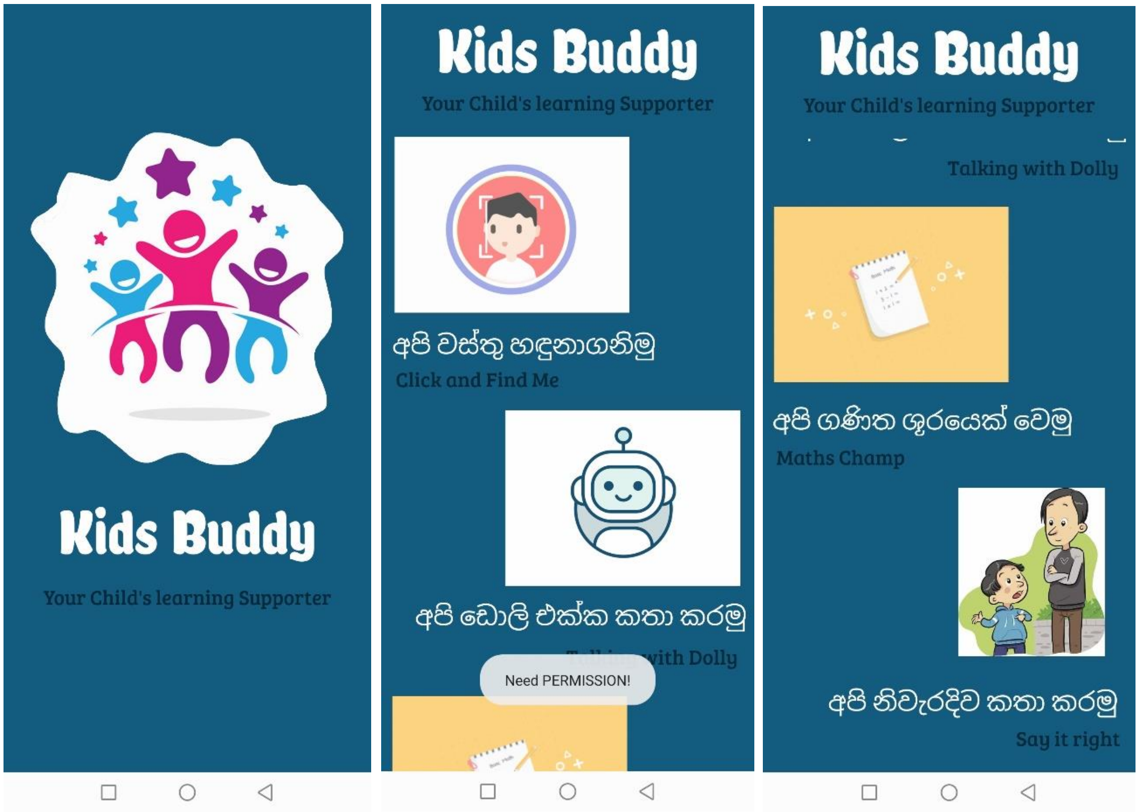
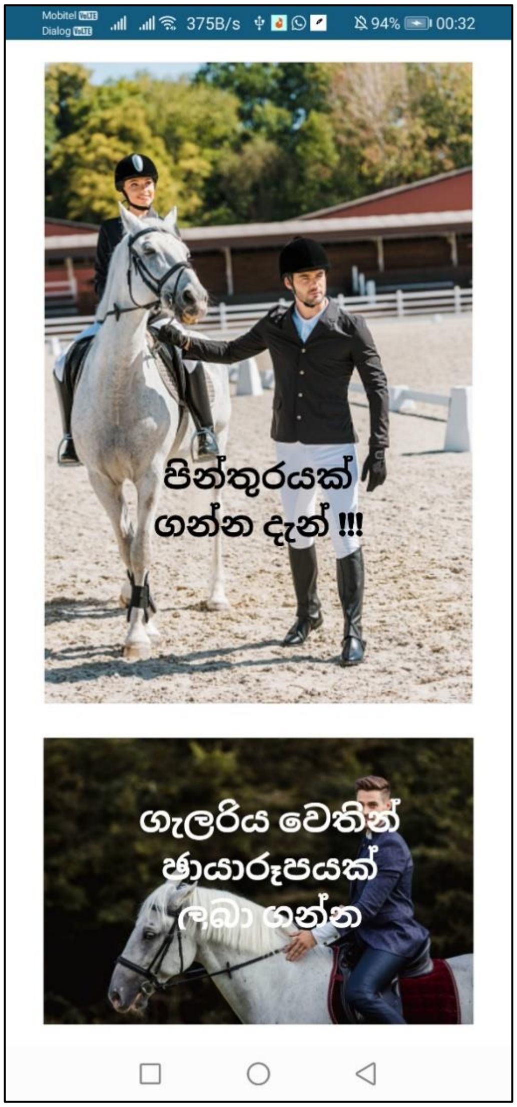
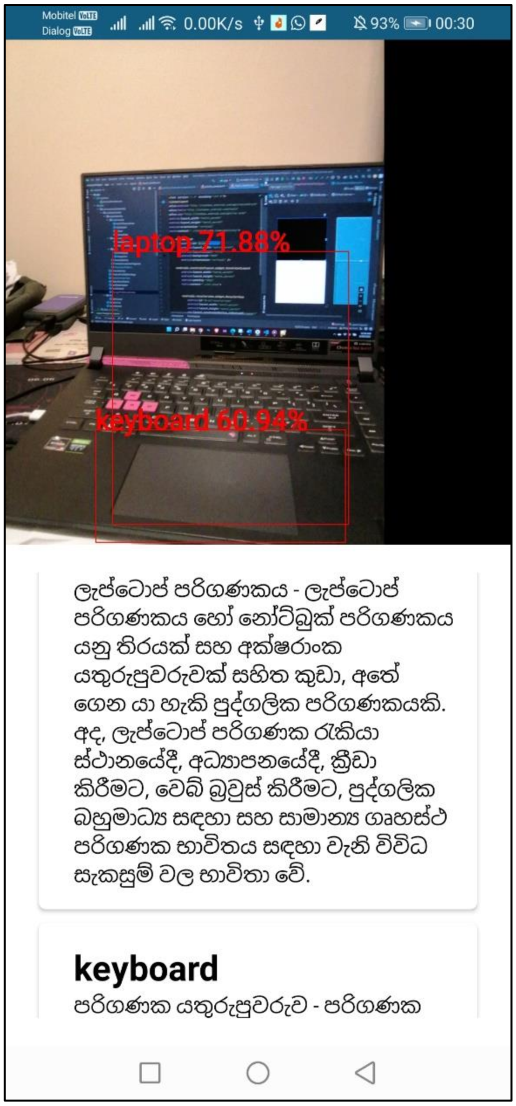


Figure 18 - User interface Kids Buddy mobile application - Splash screen and Dashboard page



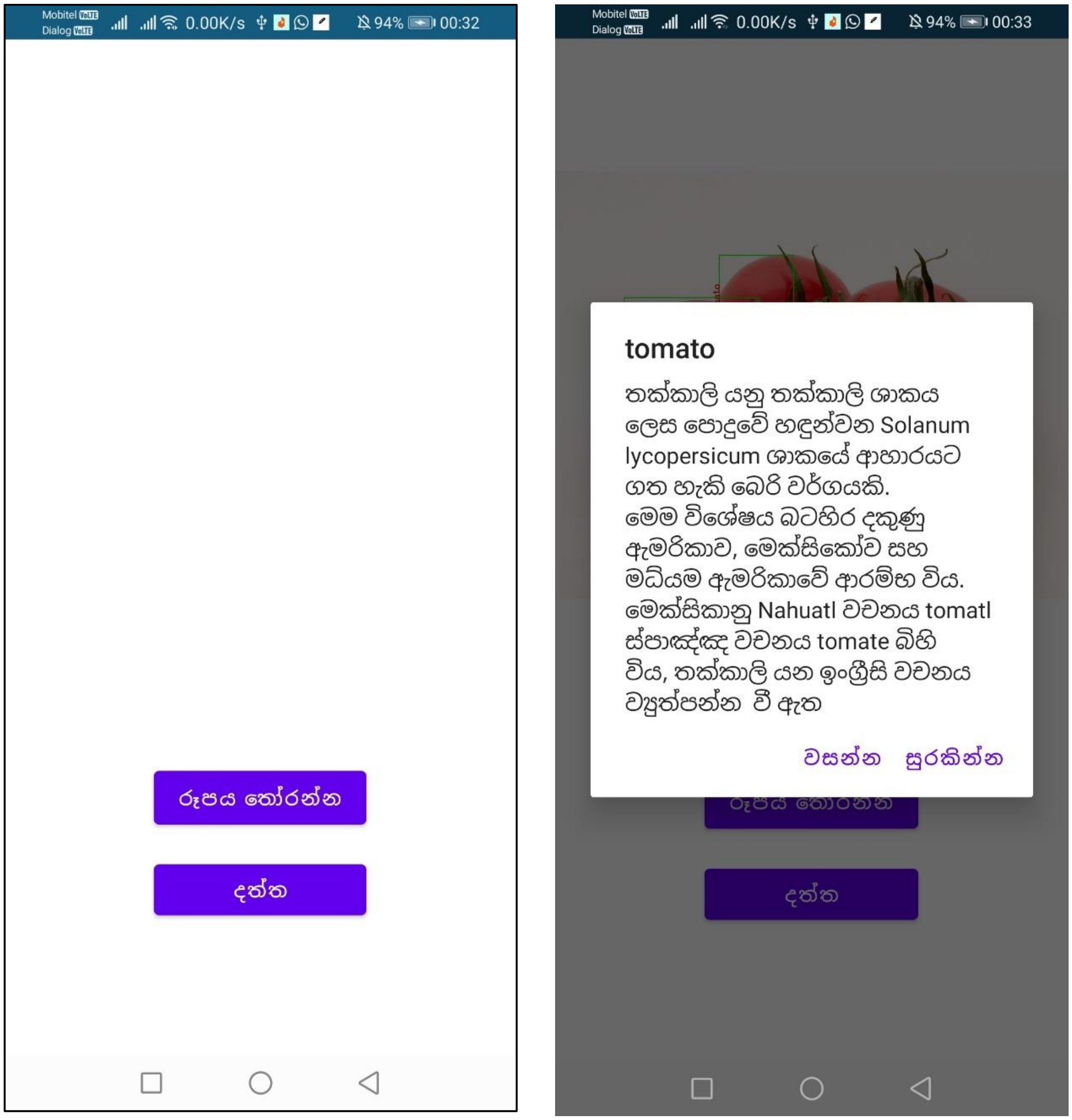


ලැප්ටොප් පරිගණකය - ලැප්ටොප් පරිගණකය හෝ නෝට්බුක් පරිගණකය යනු තිරයක් සහ අක්ෂරාංක යතුරුපුවරුවක් සහිත කුඩා, අතේ ගෙන යා හැකි පුද්ගලික පරිගණකයකි. අද, ලැප්ටොප් පරිගණක රැකියා ස්ථානයේදී, අධ්‍යාපනයේදී, ක්‍රීඩා කිරීමට, වෙබ් බ්‍රවුස් කිරීමට, පුද්ගලික බහුමාධ්‍ය සඳහා සහ සාමාන්‍ය ගෘහස්ථ පරිගණක භාවිතය සඳහා වැනි විවිධ සැකසුම් වල භාවිතා වේ.

keyboard

පරිගණක යතුරුපුවරුව - පරිගණක

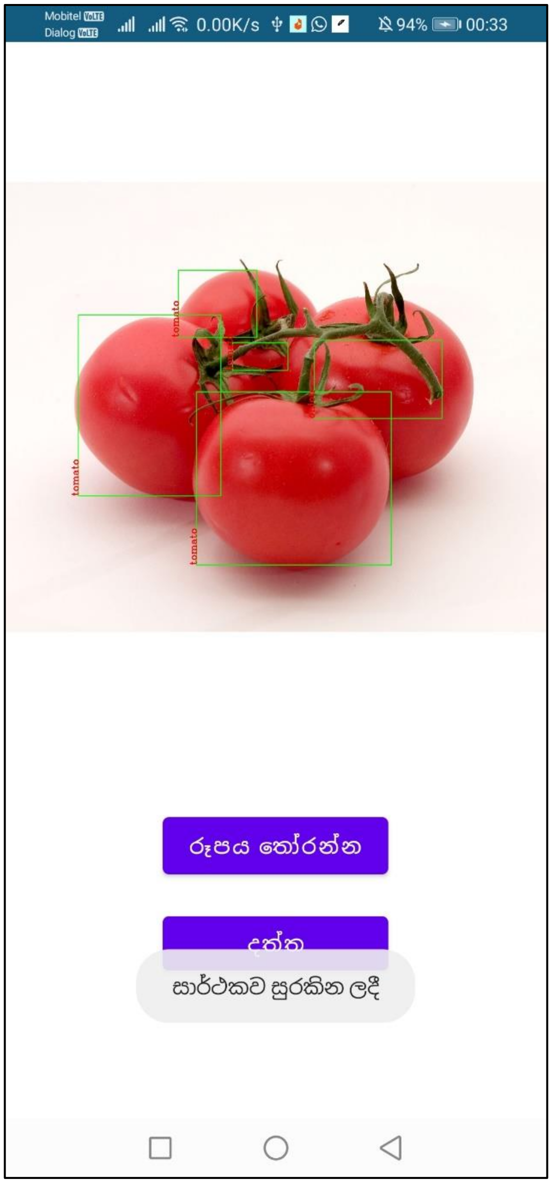
Figure 19 - User interface Kids Buddy mobile application – ClickAndFindMe main page and live object detection and Sinhala description generate page



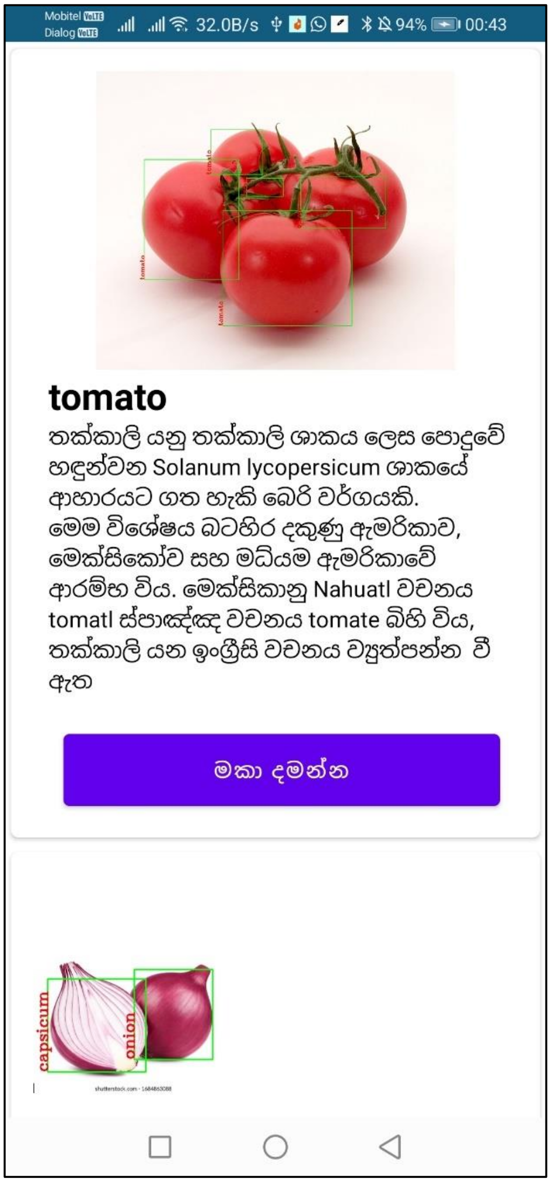
tomato

තක්කාලි යනු තක්කාලි ශාකය
ලෙස පොදුවේ හඳුන්වන Solanum
lycopersicum ශාකයේ ආහාරයට
ගත හැකි බෙරි වර්ගයකි.
මෙම විශේෂය බටහිර දකුණු
ඇමරිකාව, මෙක්සිකෝව සහ
මධියම ඇමරිකාවේ ආරම්භ විය.
මෙක්සිකානු Nahuatl වචනය tomatl
ස්පාඤ්ඤ වචනය tomate බිහි
විය, තක්කාලි යන ඉංග්‍රීසි වචනය
ව්‍යුත්පන්න වී ඇත

Figure 20 - User interface Kids Buddy mobile application – Storage Image main page and storage image object detection and Sinhala description generate page



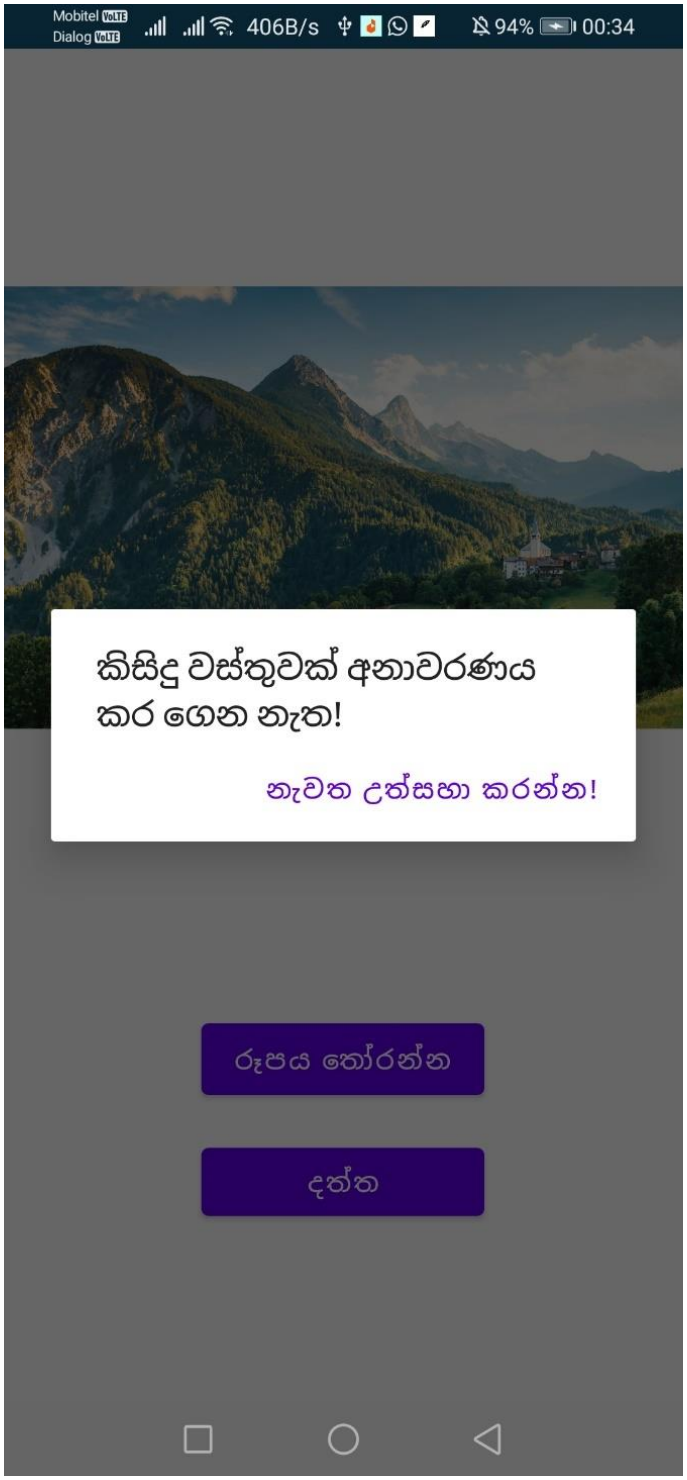
සාර්ථකව සුරකින ලදී



tomato

තක්කාලි යනු තක්කාලි ශාකය ලෙස පොදුවේ හඳුන්වන Solanum lycopersicum ශාකයේ ආහාරයට ගත හැකි බෙරි වර්ගයකි. මෙම විශේෂය බටහිර දකුණු ඇමරිකාව, මෙක්සිකෝව සහ මධ්‍යම ඇමරිකාවේ ආරම්භ විය. මෙක්සිකානු Nahuatl වචනය tomatl ස්පාඤ්ඤ වචනය tomate බිහි විය, තක්කාලි යන ඉංග්‍රීසි වචනය ව්‍යුත්පන්න වී ඇත

Figure 21 - User interface Kids Buddy mobile application – storage image object detection page and detected image store page with the Sinhala description



කිසිදු වස්තුවක් අනාවරණය කර ගෙන නැත!

Figure 22 - User interface Kids Buddy mobile application – Error msg generate for non-identified objects

2.11.7 Database

In this object detection component, the database that is used for storing the necessary object detection images and the relevant Sinhala description in the system is SQLite which is a NoSQL database. In this object detection and Sinhala description generated function, after image is used for object detection, the child can store those object detection and Sinhala description generation details in the mobile phone using the SQLite database. In addition, the child can also delete the stored details. This function is added for child's later studies and help to improve the object identification skills within them.

2.12 Testing

At various stages of the development life cycle, the system needs to be tested using a variety of methodologies. These tests assist in identifying any system flaws. Testing is a difficult and important step in the development of the application. Usability, performance, security, functional, and non-functional elements are all included in application testing. Testing will improve the quality of the final product. Early detection of the system's flaws is essential for preventing such issues. Bugs and problems can be fixed by creating test cases for each operation. For this applicant ranking application, unit, module, integration, and system testing should be performed.

2.12.1 Unit Testing

Every module is tested individually to ensure that it meets all the standards and has all the necessary functionalities. The components can be easily merged with other components if they are error-free. Individual operations, such as those involved in model training and description creation, are examined in this research.

2.12.2 Module Testing

This testing method involves checking each and every component and subsystem. Everyone in the group contributes to it; the user is not required to test their own modules.

2.12.3 Integration Testing

Following the integration of components or subsystems, integration testing is carried out. The system is completed by integrating elements like object identification, communication training, pronunciation training, and mathematics training. Regression testing should be carried out on the system once these components have been integrated to make sure that the extra pieces have not altered the system's functionality.

2.12.4 System Testing

The system does not need to be aware of the internal structure of the component; it is completed once all the components have been integrated. Here, the system is checked with actual outputs and expected outputs of the applicants.

The test cases for each testing task are listed below.

Test Case No	Test case 01
Description	Train the model
Test Procedure	1. Run the YOLOv4 model
Input	COCO dataset
Expected Output	Model should be trained
Actual results	Model is trained
Pass/Fail	Pass

Table 4 - Test case for train the model

Test Case No	Test case 02
Description	Identified the objects
Test Procedure	1. Insert the images to the system 2. Run the system
Input	Images with objects
Expected Output	System should identify the objects
Actual results	System identified the objects
Pass/Fail	Pass

Table 5 - Test case for identified the objects

Test Case No	Test case 03
Description	Checks whether the system will generate Sinhala description
Test Procedure	1. Insert images to the system 2. Click more details button
Input	Resumes and the job requirements
Expected Output	Generate Sinhala Description about the identified object
Actual results	Generate Sinhala Description about the identified object
Pass/Fail	Pass

Table 6 - Test case for Generate Sinhala Description for the identified object

Test Case No	Test case 04
Description	Checks whether the system will generate an error message for the unidentified object
Test Procedure	1. Insert images to the system
Input	Image
Expected Output	Show the error message
Actual results	Show the error message
Pass/Fail	Pass

Table 7 - Test case for Display the error message

3. RESULTS & DISCUSSION

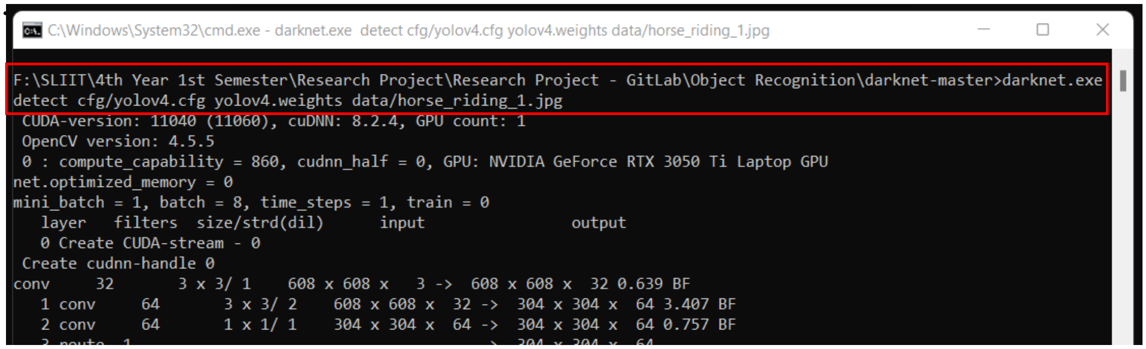
3.1 Results

The findings from the application's development are covered in this chapter. Identifying objects is an essential skill that child must improve, as it is important for development of the child's brain. Normal children can improve this object recognition skill in a short time period, but slow learners need more time and assistance to develop this skill. So, analyzing resumes become heavy work for them. To improve the basic educational skills of slow-learning children, applications like these would be very helpful. This program was created primarily with slow children are aged between 4 to 6. As was already said, this approach has four individual components and each part will help slow learners to develop four different skills.

The basic idea of this application is to develop an application that helps the slow learners to improve their skills. This system is checked with some sample images. This chapter describes the results with the code segments for each function or the components.

3.1.1 Output of YOLOv4 Trained model

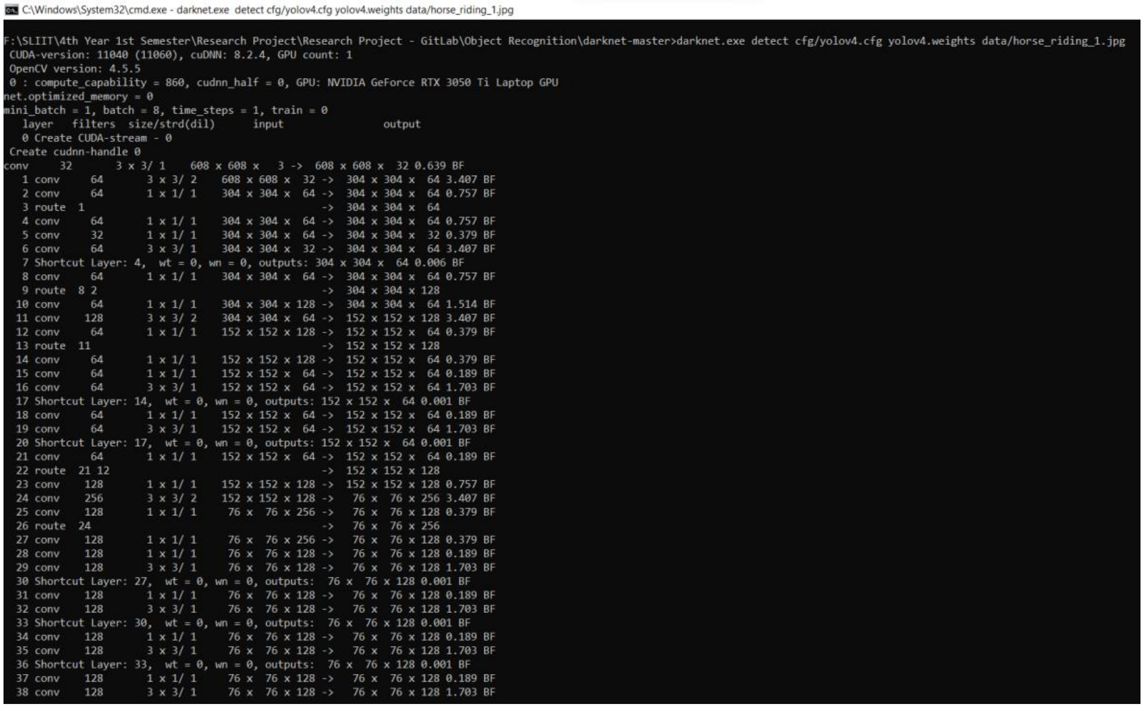
First, when the image is included, path location runs the model. The below figures 21 and 22 show the image path which needed to include for run the YOLOv4 model to identify the objects and output of the lunched object detection.



```
C:\Windows\System32\cmd.exe - darknet.exe  detect cfg/yolov4.cfg yolov4.weights data/horse_riding_1.jpg

F:\SLIIT\4th Year 1st Semester\Research Project\Research Project - GitLab\Object Recognition\darknet-master>darknet.exe
detect cfg/yolov4.cfg yolov4.weights data/horse_riding_1.jpg
 CUDA-version: 11040 (11060), cuDNN: 8.2.4, GPU count: 1
 OpenCV version: 4.5.5
 0 : compute_capability = 860, cudnn_half = 0, GPU: NVIDIA GeForce RTX 3050 Ti Laptop GPU
net.optimized_memory = 0
mini_batch = 1, batch = 8, time_steps = 1, train = 0
   layer   filters  size/strd(dil)      input                output
   0 Create CUDA-stream - 0
 Create cudnn-handle 0
conv     32       3 x 3/ 1    608 x 608 x   3 ->  608 x 608 x  32 0.639 BF
   1 conv     64       3 x 3/ 2    608 x 608 x  32 ->  304 x 304 x  64 3.407 BF
   2 conv     64       1 x 1/ 1    304 x 304 x  64 ->  304 x 304 x  64 0.757 BF
   3 route  1                                      ->  304 x 304 x  64
```

Figure 23 - Image path include to run the YOLOv4 model



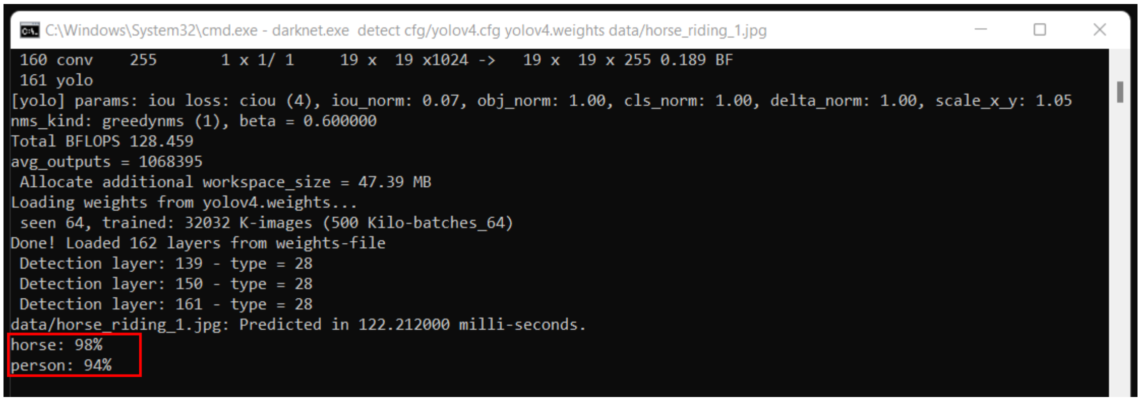
```
C:\Windows\System32\cmd.exe - darknet.exe  detect cfg/yolov4.cfg yolov4.weights data/horse_riding_1.jpg

F:\SLIIT\4th Year 1st Semester\Research Project\Research Project - GitLab\Object Recognition\darknet-master>darknet.exe detect cfg/yolov4.cfg yolov4.weights data/horse_riding_1.jpg
 CUDA-version: 11040 (11060), cuDNN: 8.2.4, GPU count: 1
 OpenCV version: 4.5.5
 0 : compute_capability = 860, cudnn_half = 0, GPU: NVIDIA GeForce RTX 3050 Ti Laptop GPU
net.optimized_memory = 0
mini_batch = 1, batch = 8, time_steps = 1, train = 0
   layer   filters  size/strd(dil)      input                output
   0 Create CUDA-stream - 0
 Create cudnn-handle 0
conv     32       3 x 3/ 1    608 x 608 x   3 ->  608 x 608 x  32 0.639 BF
   1 conv     64       3 x 3/ 2    608 x 608 x  32 ->  304 x 304 x  64 3.407 BF
   2 conv     64       1 x 1/ 1    304 x 304 x  64 ->  304 x 304 x  64 0.757 BF
   3 route  1                                      ->  304 x 304 x  64
   4 conv     64       1 x 1/ 1    304 x 304 x  64 ->  304 x 304 x  64 0.757 BF
   5 conv     32       1 x 1/ 1    304 x 304 x  64 ->  304 x 304 x  32 0.379 BF
   6 conv     64       3 x 3/ 1    304 x 304 x  32 ->  304 x 304 x  64 3.407 BF
   7 Shortcut Layer: 4,  wt = 0, wn = 0, outputs: 304 x 304 x  64 0.006 BF
   8 conv     64       1 x 1/ 1    304 x 304 x  64 ->  304 x 304 x  64 0.757 BF
   9 route  8 2                                    ->  304 x 304 x 128
  10 conv     64       1 x 1/ 1    304 x 304 x 128 ->  304 x 304 x  64 1.514 BF
  11 conv    128       3 x 3/ 2    304 x 304 x  64 ->  152 x 152 x 128 3.407 BF
  12 conv     64       1 x 1/ 1    152 x 152 x 128 ->  152 x 152 x  64 0.379 BF
  13 route  11                                     ->  152 x 152 x 128
  14 conv     64       1 x 1/ 1    152 x 152 x 128 ->  152 x 152 x  64 0.379 BF
  15 conv     64       1 x 1/ 1    152 x 152 x  64 ->  152 x 152 x  64 0.189 BF
  16 conv     64       3 x 3/ 1    152 x 152 x  64 ->  152 x 152 x  64 1.703 BF
  17 Shortcut Layer: 14,  wt = 0, wn = 0, outputs: 152 x 152 x  64 0.001 BF
  18 conv     64       1 x 1/ 1    152 x 152 x  64 ->  152 x 152 x  64 0.189 BF
  19 conv     64       3 x 3/ 1    152 x 152 x  64 ->  152 x 152 x  64 1.703 BF
  20 Shortcut Layer: 17,  wt = 0, wn = 0, outputs: 152 x 152 x  64 0.001 BF
  21 conv     64       1 x 1/ 1    152 x 152 x  64 ->  152 x 152 x  64 0.189 BF
  22 route  21 12                                  ->  152 x 152 x 128
  23 conv    128       1 x 1/ 1    152 x 152 x 128 ->  152 x 152 x 128 0.757 BF
  24 conv    256       3 x 3/ 2    152 x 152 x 128 ->   76 x  76 x 256 3.407 BF
  25 conv    128       1 x 1/ 1     76 x  76 x 256 ->   76 x  76 x 128 0.379 BF
  26 route  24                                     ->   76 x  76 x 256
  27 conv    128       1 x 1/ 1     76 x  76 x 256 ->   76 x  76 x 128 0.379 BF
  28 conv    128       1 x 1/ 1     76 x  76 x 128 ->   76 x  76 x 128 0.189 BF
  29 conv    128       3 x 3/ 1     76 x  76 x 128 ->   76 x  76 x 128 1.703 BF
  30 Shortcut Layer: 27,  wt = 0, wn = 0, outputs:  76 x  76 x 128 0.001 BF
  31 conv    128       1 x 1/ 1     76 x  76 x 128 ->   76 x  76 x 128 0.189 BF
  32 conv    128       3 x 3/ 1     76 x  76 x 128 ->   76 x  76 x 128 1.703 BF
  33 Shortcut Layer: 30,  wt = 0, wn = 0, outputs:  76 x  76 x 128 0.001 BF
  34 conv    128       1 x 1/ 1     76 x  76 x 128 ->   76 x  76 x 128 0.189 BF
  35 conv    128       3 x 3/ 1     76 x  76 x 128 ->   76 x  76 x 128 1.703 BF
  36 Shortcut Layer: 33,  wt = 0, wn = 0, outputs:  76 x  76 x 128 0.001 BF
  37 conv    128       1 x 1/ 1     76 x  76 x 128 ->   76 x  76 x 128 0.189 BF
  38 conv    128       3 x 3/ 1     76 x  76 x 128 ->   76 x  76 x 128 1.703 BF
```

Figure 24 - Output of the lunched object detection

After running the object detection model, it will generate the predicted object identification percentage values with the image.

The below figures 23 and 24 show the terminal output of the tested image using YOLOv4 model and the output of the tested image.



```
C:\Windows\System32\cmd.exe - darknet.exe  detect cfg/yolov4.cfg yolov4.weights data/horse_riding_1.jpg

 160 conv    255       1 x 1/ 1     19 x  19 x1024 ->    19 x  19 x 255 0.189 BF
 161 yolo
[yolo] params: iou loss: ciou (4), iou_norm: 0.07, obj_norm: 1.00, cls_norm: 1.00, delta_norm: 1.00, scale_x_y: 1.05
nms_kind: greedynms (1), beta = 0.600000
Total BFLOPS 128.459
avg_outputs = 1068395
 Allocate additional workspace_size = 47.39 MB
Loading weights from yolov4.weights...
 seen 64, trained: 32032 K-images (500 Kilo-batches_64)
Done! Loaded 162 layers from weights-file
 Detection layer: 139 - type = 28
 Detection layer: 150 - type = 28
 Detection layer: 161 - type = 28
data/horse_riding_1.jpg: Predicted in 122.212000 milli-seconds.
horse: 98%
person: 94%
```

Figure 25 - Terminal output of the object detection

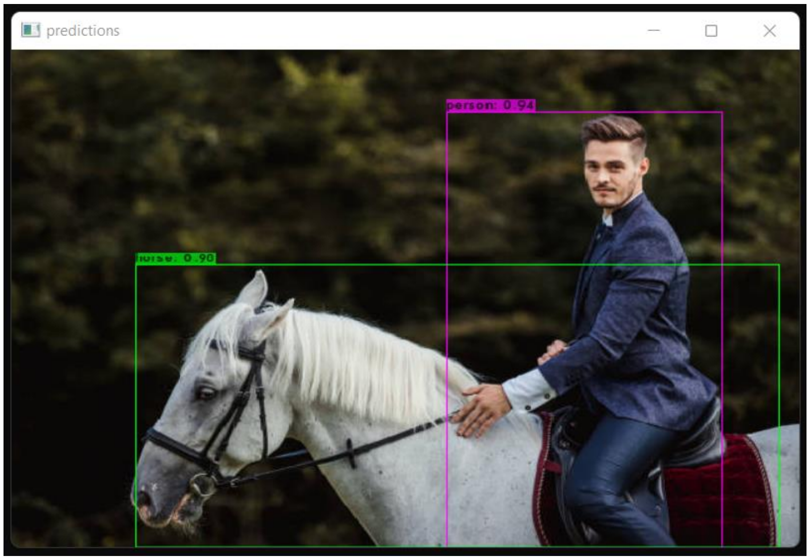
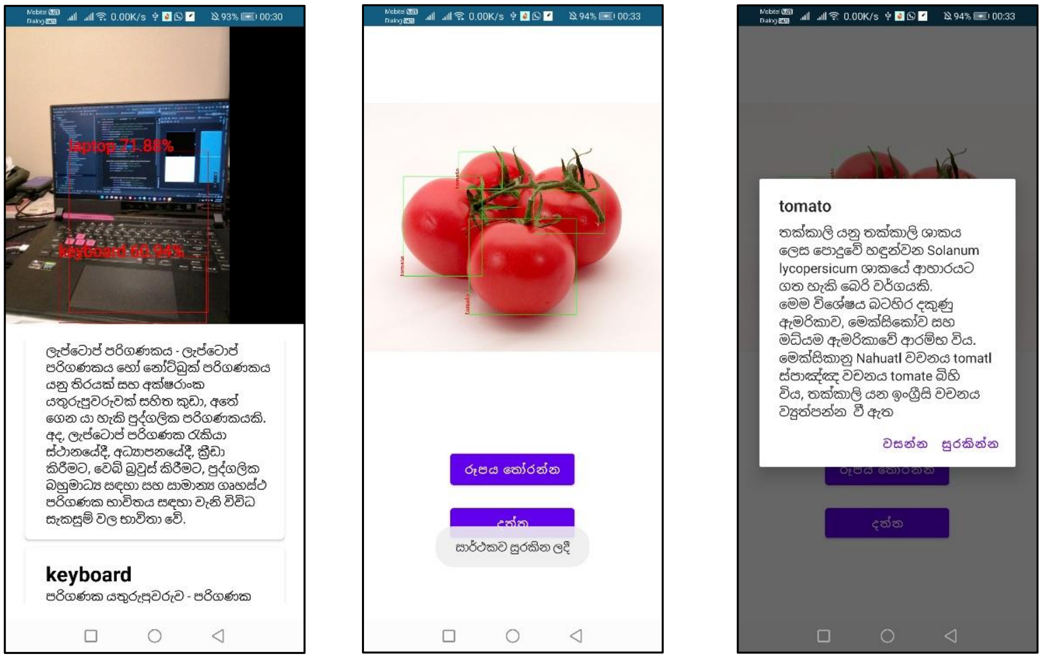


Figure 26 - Output of the tested image

3.1.2 Output of the Object detection and the description generator

The child needs to add/upload an image into the system. Using converted tflite model, the application will identify the objects in the image. After the objects are identified, the object names are then passed to generate Sinhala descriptions using Wikipedia API and Google API.

The below figure 25 shows the output of the identified object with the Sinhala description.



ලැප්ටොප් පරිගණකය - ලැප්ටොප් පරිගණකය හෝ නෝට්බුක් පරිගණකය යනු තිරයක් සහ අක්ෂරාංක යතුරුපුවරුවක් සහිත කුඩා, අතේ ගෙන යා හැකි පුද්ගලික පරිගණකයකි. අද, ලැප්ටොප් පරිගණක රැකියා ස්ථානයේදී, අධ්‍යාපනයේදී, ක්‍රීඩා කිරීමට, වෙබ් බ්‍රවුස් කිරීමට, පුද්ගලික බහුමාධ්‍ය සඳහා සහ සාමාන්‍ය ගෘහස්ථ පරිගණක භාවිතය සඳහා වැනි විවිධ සැකසුම් වල භාවිතා වේ.

keyboard

පරිගණක යතුරුපුවරුව - පරිගණක

සාර්ථකව සුරකින ලදී

tomato

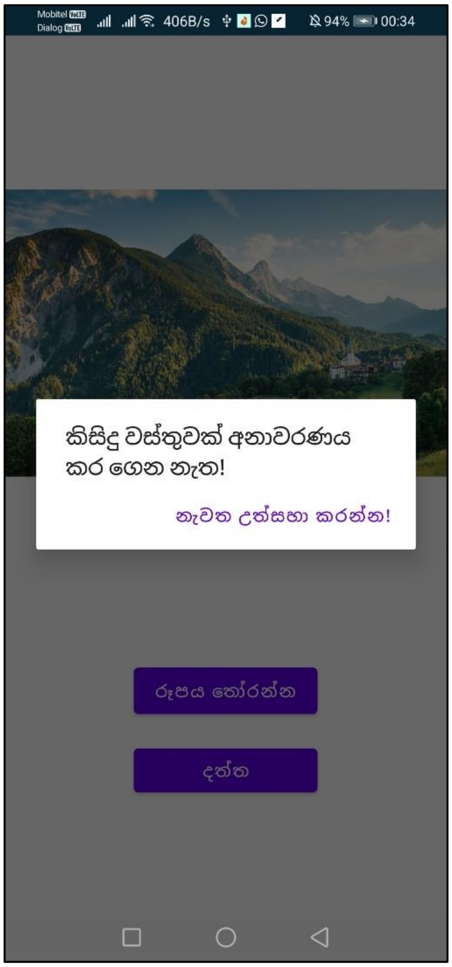
තක්කාලි යනු තක්කාලි ශාකය ලෙස පොදුවේ හඳුන්වන Solanum lycopersicum ශාකයේ ආහාරයට ගත හැකි බෙරි වර්ගයකි. මෙම විශේෂය බටහිර දකුණු ඇමරිකාව, මෙක්සිකෝව සහ මධ්‍යම ඇමරිකාවේ ආරම්භ විය. මෙක්සිකානු Nahuatl වචනය tomatl ස්පාඤ්ඤ වචනය tomate බිහි විය, තක්කාලි යන ඉංග්‍රීසි වචනය ව්‍යුත්පන්න වී ඇත

Figure 27 - Output of identified object with the Sinhala description.

3.1.3 Output of the non-identified object error message generation

The child needs to add/upload an image into the system. Using converted tflite model the application will identify the objects in the image which are included in the COCO dataset. If model does not identify the objects in the image the system will generate error message to child, saying that "There is no object identified, try again".

The below image shows output of the non-identified object error message generation.



කිසිදු වස්තුවක් අනාවරණය කර ගෙන නැත!

Figure 28 - Output of the non-identified object error message generation

3.1.4 Model Evaluation Results

In this study, the evaluation of the models is crucial. The system performed as expected under many testing conditions, including complicated backgrounds, environments with bright light, and environments with low light. The test functionality results are displayed in the table 9 below. Based on the accuracy obtained from the model evaluation, YOLOv4 model is selected for the object detection.

Test no	Test Scenario	No of test runs	Accuracy (%)
1	Dark light environment	10	83
2	Low light environment	10	83
3	Complex light environment	10	82

Table 8 - Model test result

3.2 Research Findings

The objective of this research is to create a system that will assist in enhancing slow children's learning capacities. Beside the object detection component of the application, other components display the output of the accuracy level of the child responses.

It is evident from the results of each phase that a variety of factors affect the system's accuracy. These include the quality and quantity of training samples, the image used as input, the model parameters, the accuracy requirement threshold, dataset size, and the precision of the object detection algorithm. These elements will be covered in the following chapter (3.3).

The system should be a web application or a mobile application, as per the specifications. Desktop applications are inappropriate for this use. Consequently, a mobile application is created.

- **Simple and user friendly**

  The primary user of the system – Children and parents

Interfaces are designed in a user-friendly manner. Therefore, children can easily understand how system functions.

The output view is also very clear. Each component has its own way of showing the accuracy level to child through the application.

- **Available with low cost**

  This application is a free application and everyone can use it.

- **Accuracy**

  The System provides a more accurate score for each object detected image. The accuracy of the system is checked by comparing the COCO dataset with the images that taken from the application. According to the test cases and the number of objects in the dataset, accuracy can be more than 75%.

- **Quick response**

  Using this system, child can identify different objects in a short period of time.

There are some difficulties occurring during the implementation of this application.

- Some objects would not be identified by the application.
- Only the objects in COCO dataset will be identified by the application and other object would not be identified by the application.

3.3 Discussion

This section describes the accuracy levels of the system and the accuracy of the algorithms used. This system's accuracy primarily depends on the object detection YOLOv4 model. The model's accuracy is determined by the quality and quantity of training samples, the input imagery, the model parameters, and the required dataset size. According to the existing article, the dataset size more important in object detection [26]. In this research, object detection is based on the COCO dataset. So, COCO dataset is used for training the model and testing.

The accuracy of the model was checked by creating a model using a different domain and using the same domain. The following table 10 will shows, few of the objects detection accuracy level after testing the with the YOLOv4 model.

Object name	Accuracy level (%)
Person	94%
Horse	90%
Dog	98%
Bicycle	92%
Truck	92%
Bowl	72%
Cat	93%
Train	94%
Bear	93.4%
Backpack	71.1%
Handbag	80.4%
Book	69.3%

Table 9 - Accuracy of the object detection result

Also, the reason for choosing YOLOv4 is, it is more accurate than new transformer based SWIN methods and other popular algorithms.

According to the research YOLOv4 has the best speed and accuracy rate [30]. The below diagram (Figure 27) shows the accuracy between YOLOv4 and the other algorithms.

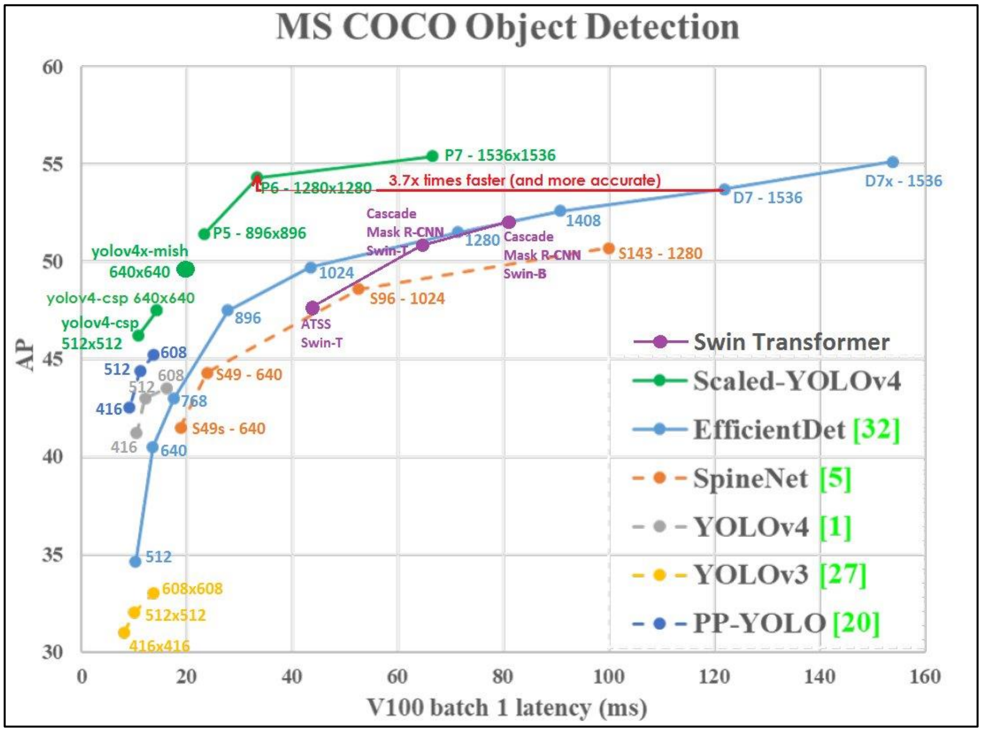


Figure 29 - Accuracy between YOLOv4 and the other algorithms

The accuracy of the system also depends on the parameters that are used to create the YOLOv4 model. To create the model, the parameters like batch=64, subdivisions=8, width=608, height=608, channels=3 and etc. are used. By changing the values of these parameters, the accuracy can be increased. Here the values used for parameters are,

- batch=64
- subdivisions=8
- width=608
- height=608
- channels=3

- momentum=0.949
- decay=0.0005
- angle=0
- saturation = 1.5
- exposure = 1.5
- hue=.1

- learning_rate=0.0013
- burn_in=1000
- max_batches = 500500
- policy=steps
- steps=400000,450000
- scales=.1,.1

- mosaic=1

Finally, the accuracy and the performance of the system are checked by inserting images with various objects. That means, images with different classes are inserted. The system will identify the objects that are included in the COCO dataset. So, this application will help the children to identify the objects in an accurate manner, within a short period of time.

3.4 Future Works

Future works will still have access to many features and experimentation. Future studies should pay closer attention to the system's accuracy. Since this accuracy level is below 90%. In the future, it is preferable to boost accuracy by more than 90%. It is essential to improve the model's accuracy in order to improve the system's accuracy. As a result, by adding more classes and enhancing the dataset, the model's accuracy can be raised. Currently, this system gives output in Sinhala language. Therefore, future research might apply to expand the system to other languages, as well for the children who speak different languages. To identify more objects with different classes, it is a must to collect a considerable size of the dataset related to other objects, beside the objects in COCO dataset. It is important that future research investigates the format of the input. This system takes the input in image format like png, jpg or jpeg. It is better to get the resume in PDF or another format. This is because the child can input any file format to the system for object detection. By considering the above-mentioned changes in the future, the best application can be built to overcome the issues present in the object detection system.

4. CONCLUSION

To summarize, this research work is about developing an application for slow learning children to improve their learning skills. With technology growing in the industry, everyone has started to use smart applications and modern technology for all purposes. When it comes to slow-learning children's education, most learning activities are done in physically. Now in the modern education system, most educational actives are based on E-Learning because children can learn from home without attending school. Since technology is advancing quickly along with time, every industry has moved online, with e-learning being one of the main ones. In the modern world, most application for e-learning cater to normal children. This is a common mistake that done by the modern community when there are slow-learning children who need more help compared to normal children. So, it is important to focus on these children as well for their better education in the future.

The main goal of this research is to develop an application to make the above process easy and smart. There are some applications available for this purpose in the market but not specifically for slow learners. However, this application identified the objects based on the model trained using the YOLOv4. Each object can get percentage-based values from this model. This system identified the objects by comparing the COCO dataset and the given image. The object detection model can predict items on diverse lighting conditions and backgrounds with multiple objects with the aid of TensorFlow Lite made training easier. The DL model is now trained just using the COCO dataset, and it works with an accuracy of 83%. After successfully identifying the object, the system will generate a Sinhala description about the identified object. If the system cannot identify the object, the application will generate an error message. This is the overall function of this system.

To train a model, the COCO dataset is used. The object detection model in the research is had been designed, implemented, and tested successfully. Data can be crawled from job websites like Monster, using a crawling mechanism. This crawled corpus then undergoes text preprocessing and finally, preprocessed corpus is used to train the model. The Word2Vec method is selected from among other methods to train the model. When the recruiter inserts the resumes of the applicants and the job requirements into the system, it preprocesses the resumes and requirements and gets top tokens. Then tokens from the resumes are compared with tokens from the job requirements. The similarity between these two tokens is calculated by using the word embedding model. The system calculates similarity scores for each token and finally total the scores and finally output the final score. Then system ranks the resumes based on the scores. This is the overall function of this system.

Therefore, it is expected that slow-learning children will benefit from this application and more children will find this enlightening for their studies in the near future.

REFERENCES

[1] "Brain Development - First Things First," *First Things First*, Jul. 04, 2022. [Online]. Available: https://www.firstthingsfirst.org/early-childhood-matters/brain-development/#:~:text=90%25%20of%20Brain%20Growth%20Happens,full%20grown%20%E2%80%93%20by%20age%205. [Accessed: Sep. 07, 2022]

[2] Sabina Martyn, "In Post-Conflict Sri Lanka, Language is Essential for Reconciliation." https://asiafoundation.org/2013/01/16/in-post-conflict-sri-lanka-language-is-essential-for-reconciliation/#:~:text=Sri%20Lanka's%20population%20is%20comprised,Malays%2C%20Burghers%2C%20and%20others. (Accessed Sep. 7, 2022).

[3] F. Liu, Y. Liu, S. Lin, W. Guo, F. Xu, and B. Zhang, "Fast Recognition Method for Tomatoes under Complex Environments Based on Improved YOLO," Transactions of the Chinese Society for Agricultural Machiner, vol. 51, no. 06, pp. 229-237, 2020.

[4] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[5] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[8] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: single shot multibox detector," European conference on computer vision, pp. 21-37, 2016.

[10] W. Yijing, Y. Yi, W. Xue-fen, C. Jian and L. Xinyun, "Fig Fruit Recognition Method Based on YOLO v4 Deep Learning," 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2021, pp. 303-306, doi: 10.1109/ECTI-CON51831.2021.9454904.

[11] P. Meghana, S. Sagar Imambi, P. Sivateja, K. Sairam "Image Recognition for

Automatic Number Plate Surveillance” International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume: 08, Issue: 04, February 2019.

[12] R. Matheson, “Machine-learning system tackles speech and object recognition, all at once”. MIT News | Massachusetts Institute of Technology. https://news.mit.edu/machine-learning-image-object-recognition-0918 (Accessed 22 August 2022).

[13] A. Kojima , T. Tamura , K. Fukunaga , Natural language description of human activities from video images based on concept hierarchy of actions, Int. Comput. Vis. 50 (2002) 171–184 .

[14] P. Hede , P. Moellic , J. Bourgeoys , M. Joint , C. Thomas ,Automatic generation of natural language descriptions for images, in: Proceedings of the Recherche Dinformation Assistee Par Ordinateur, 2004.

[15] J. Donahue , Y. Jia , O. Vinyals , J. Hoffman , N. Zhang , E. Tzeng , T. Darrell , DeCAF: a deep convolutional activation feature for generic visual recognition, in: Proceedings of The Thirty First International Conference on Machine Learning, 2014, pp. 647–655.

[16]A. Farhadi , M. Hejrati , M.A. Sadeghi , P. Young , C. Rashtchian , J. Hocken- maier , D. Forsyth , Every picture tells a story: Generating sentences from images, in: Proceedings of the European Conference on Computer Vision,,2010, pp. 15–29.

[17] M. Hodosh , P. Young , J. Hockenmaier , Framing image description as a rank- ing task: data, models and evaluation metrics, J. Artif. Intell. Res. 47 (2013) 853–899 .

[18] Y. Yang , C.L. Teo , H. Daume , Y. Aloimono , Corpus-guided sentence generation of natural images, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 4 4 4–454 .

[19] R. Socher , A. Karpathy , Q.V. Le , C.D. Manning , A.Y. Ng , Grounded compositional semantics for finding and describing images with sentences, TACL 2 (2014) 207–218 .

[20] O. Vinyals , A. Toshev , S. Bengio , D. Erhan , S3how and tell: a neural image caption generator, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3156–3164 .

[21]Q. You , H. Jin , Z. Wang , C. Fang , J. Luo , Image captioning with semantic attention, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4651–4659

[22]N. Komal Kumar, D. Vigneswari, A. Mohan, K. Laxman, J. Yuvaraj, “Detection and Recognition of Objects in Image Caption Generator System: A Deep Learning Approach.”, 8728-516 ©2019 IEEE.

[23] Aghasi Poghosyan, Aghasi Poghosyan, “Long Short-Term Memory with Read-only Unit in Neural Image Caption Generator.” 8312-163 ©2017 IEEE.

[24] Mirza Muhammad Ali Baig, Mian Ihtisham Shah, Muhammad Abdullah Wajahat, Nauman Zafar and Omar Arif, “Image Caption Generator with Novel Object injection”, 8615-810 ©2018 IEEE.

[25] Dong-Jin Kim, Donggeun Yoo, Bonggeun Sim, In So Kweon, “Sentence Learning on Deep Convolutional Networks for Image Caption Generation”, 7625-747 ©2016 IEEE.

[26] Hao Dong, Jingqing Zhang, Douglas McIlwraith, Yike Guo, “I2T2I: LEARNING TEXT TO IMAGE SYNTHESIS WITH TEXTUAL DATA AUGMENTATION” , 8296-635 ©2017 IEEE.

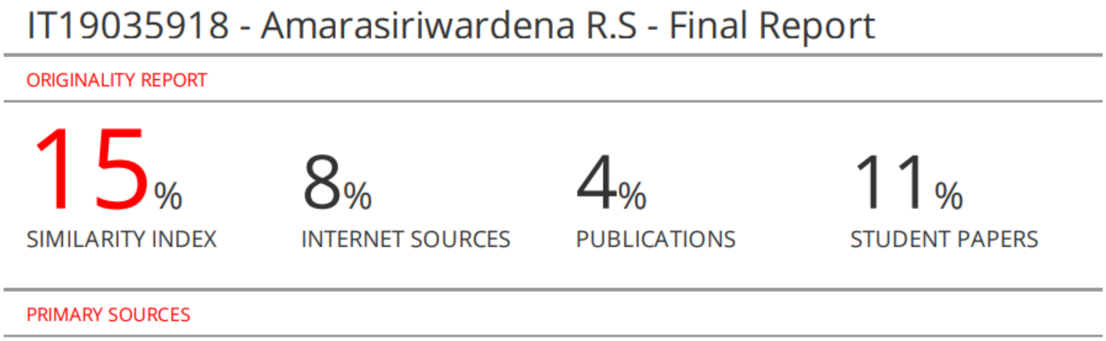
[27] M. Tatariants, “Object Detection with Small Datasets: How Much Data is Enough?,” *MobiDev*, Sep. 10, 2021. [Online]. Available: https://mobidev.biz/blog/object-detection-small-datasets-use-cases-machine-learning. [Accessed: Sep. 14, 2022]

[28] “COCO - Common Objects in Context,” *COCO - Common Objects in Context*. [Online]. Available: https://cocodataset.org/#home. [Accessed: Sep. 14, 2022]

[29] T. Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context | SpringerLink,” *Microsoft COCO: Common Objects in Context | SpringerLink*. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48. [Accessed: Sep. 14, 2022]

APPENDICES

Plagiarism Report



IT19035918 - Amarasiriwardena R.S - Final Report

ORIGINALITY REPORT

15%	8%	4%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES