Interactive, visual learning-based tool for hearing impaired children to improve language skills.

22_23-J 18

Project Proposal Report

B.W.E.K. Senarathna

B.Sc. (Hons) Degree in Information Technology Specialized in Data Science

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 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.

Name	**Student ID**	**Signature**
B.W.E.K. Senarathna	IT19142838	

Signature of the Supervisor
(Dr. Janaka Wijekoon)

…….

Date

..……

ABSTRACT

The sense of hearing is a fundamental need of every human being. Yet people are born with several informalities in the sense of hearing. This result in having different levels in their sense of hearing. For those with mild hearing loss the sense of hearing can be gained through consistent exposure to speaking environments. The cognitive skills and the ability to talk are tightly bounded with the sense of hearing. With this research we try to approach these children with mild hearing loss with a mobile based application where the child can interact with the application and learn to pronounce and spell words in the meantime improve his cognitive skills. The major focus is on increasing the vocabulary of hearing-impaired kids at early childhood and bring them to the level of normal children at the time of starting school.

We use Several Machine learning techniques to make the learning curve specific to each other. Initially object detection models using YOLO algorithm to capture the objects in the environment of the child as fast as possible and use these objects around his environment to practice the words. This provides an interactive learning and child learn through these words simply by involving in games.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Description
YOLO	You Only Look Once algorithm
CNN	Convolutional Neural Network
Mask R-CNN	Mask Region-Based Convolutional Neural Network
YOLOv2	You Only Look Once algorithms version 2
YOLOv3	You Only Look Once algorithms version 3
YOLOv4	You Only Look Once algorithms version 4
YOLOv5	You Only Look Once algorithms version 5
KT	Knowledge Transfer
CEHIC	Centre for Education of Hearing Impaired Children

1. Introduction

1.1 Background and Literature Survey

The sense of hearing is one of the fundamental needs of every person. When compared to all sense organs of an individual, the need of hearing is very essential for the coordination and the functioning of other sensory organs (speaking is based on hearing etc.). When considering the community with hearing impaired, the state of not having hearing sense is a challenge in their education. Researchers have shown that hearing impairment also affects reading and writing abilities because of low exposure to a natural language. To minimize those effects and to help these children with their education, an interactive solution should take place.

Improper parenting and lack of awareness also plays a major role where hearing impaired students falling behind in education because of not giving/referring the special education they need in their early childhood. Students with hearing impairment, who gain language skills in their early childhood show good improvement in reading writing and other cognitive tasks. The proposed application will be a student centric learning system with minimal involvement from the teachers or external parties. Here the students get an interactive environment where children can learn things from their surroundings and practice words/signs by themselves. Lessons will be presented to them in an attractive manner, and it should be simple to use. Within the learning process the level of student is monitored and the best course material is selected to the students. Also, there are measures to improve the verbal communication using lip exercises and speech therapy where students will be trained regularly to pronounce the language and to get the sense of a natural language.

Let us consider the following research which was focused on the development of the Portuguese sign language with the aid of a computer game [3]. The application was created to improve the sign language skills with the aid of a story created by the author and this story resembles items which they meet in their daily activities. The game basically creates a narrated story with key words and expects the child to reproduce them in the Portuguese sign language. The application is targeted for students of age 7 to 8 in school. The systematic setup contained a webcam and a web base game application. The implementation is based on a state machine model which detects the skeleton of the user

and take their positions based on the positioning of the hands and gestures on the x and y axis of the plain. The input captured through this model is fed to the game. The game was a 2D game developed using unity engine.

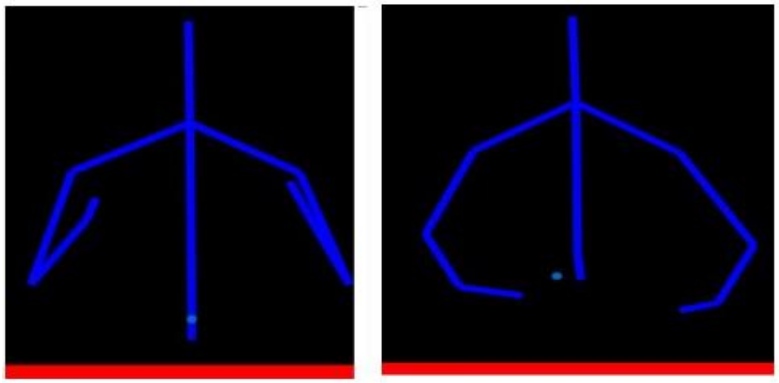
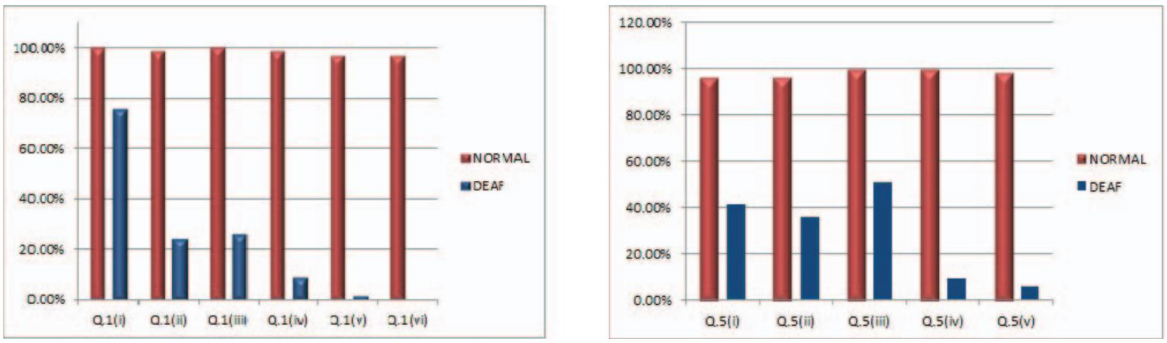


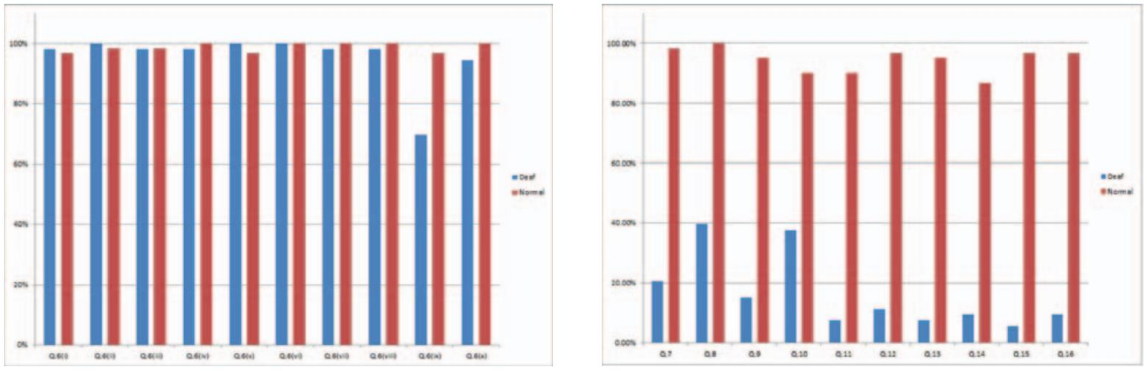
Fig 1.1.1. – Gesture recognitions through state machine model

It was also observed that the deaf children perform low in education due to not having the cognitive sense which normal students acquire at very early childhood. Many research have been conducted to give solutions for this issue in hearing impaired community. An app based unified approach to enhance the language and mathematical skills of hearing-impaired children was conducted by a team of Indian researchers based on the Indian sign language (ISL) [2]. The approach mainly focuses on developing the cognitive skills of hearing-impaired students. Here the learning curve was based on teaching the opposite words and contrast words in terms of a Indian Sign Language. The application focused on the context of the word also when making the child is engaging with the application. E.g.: White: the color of milk is white, Black: the color of hair is black. And comprehensive sentence making was also a part of this. In other perspective the children are given simple arithmetic questions to be solved. After simple arithmetic equations the students are given comprehensive arithmetic equations to be solved with the mobile application. The testing was conducted for both normal and hearing-impaired students of the same age group and 100% hearing impaired (98 students)



(a) (b)

Fig 1.1.2: (a) Complete sentence formation, (b) Comprehension, Normal hearing Red:Normal hearing & Blue:Deaf



(a) (b)

Fig 1.1.3: (a) represent Simple Arithmetic Calculations, (b) Comprehensive Arithmetic Calculations, Normal Vs Deaf Red:Normal hearing & Blue:Deaf

The above sample results show the statistics of how the hearing impaired and normal students performed in the mobile app test.

The object detection can be done in several means so the need of a specific criteria to determine the best algorithm which fits for the situation. The object detection in the application is focusing on Realtime speed object detection. To achieve this, we need a fast algorithm. Yolo is a very fast algorithm since it has it only checks once the image [6]. Unlike Faster RCNN, YOLO detects the category and position of the regression target of the whole image by a single convolution neural network [9]. YOLO has different versions from v1-v5 (Latest versions V6 and V7 are not considered here) [7]

Main improvement measures of YOLO network from V1 to V5:

• YOLO: The grid division is responsible for detection, confidence loss;

• YOLO V2: Anchor with K-means added, two-stage training, full convolutional network.

• YOLO V3: Multi-scale detection by using FPN.

• YOLO V4: SPP, MISH activation function, data enhancement Mosaic/Mixup, GIOU (Generalized Intersection over Union) loss function.

• YOLO V5: Flexible control of model size, application of Hardswish activation function, and data enhancement.

Different upgrades can be seen with the increment of the versions. Based on the architecture the best and the performance the most suitable algorithm with required updates will be selected to run the algorithm in a mobile environment.

Table 1.1.1: Performance of YOLOv3, YOLOv4, and YOLOv5 (PC)

Label	YOLOv3 Average Precision	YOLOv4 Average Precision	YOLOv5l Average Precision
Small-Vehicle	29.25	39.62	44.8
Large-Vehicle	55.84	73.43	70.1
Plane	83.06	90.39	91.3
Storage-tank	44.69	61.52	63
Ship	71.19	82.67	78.6
Harbor	67.94	80.35	82.7
Ground-track-field	36.12	67.32	65.7
Soccer-ballfield	36.82	54.24	59.8
Tennis-court	87.30	92.57	92.7
Swimming-pool	39.76	57.57	65.4
baseball	61.35	76.62	75.8
roundabout	44.14	55.98	55.9
Basketball-court	37.79	63.04	64.5
bridge	26.65	42.41	50.1
helicopter	15.84	34.54	48.2

According to the above table it is obvious that the compared to YOLOv3 and YOLOv4, YOLOv5 shows high average precision [8]. Many subversions of YOLOv5 can be run on Android applications. There are many variations from these YOLO versions some are optimized to run in mobile environments.

1.2 Research Gap

Many tools with many different approaches have been introduced to the mitigate the gap between the learning curve of student and their studies. It is observed that the main form of communication among deaf children are based on the sign language. Since it is their principal form of communication among the hearing impaired [1] [2]. Based on this many applications and approaches has been taken to enhance the learning experience of each child by making them exposed to systems which help them in learning and practicing the sign language. Most of the teachers take traditional approaches to teach children.

In the research Which focused on teaching the child the Portuguese sign language sign language through means of games [3], the initial tracking of the character is done using state machines were resulting false positives. And it is discovered that the main component of teaching children the sign language is based on facial expressions [1]. So, this application does not take this into consideration. In other terms the game which involves in a game is based on a static story that is common to everyone. The stories needed for the game are preloaded to the application. No dynamic behavior in the process to make it more attractive to the user for a self-oriented learning

In the 2nd research regarding the app based unified approach to enhance the language and mathematical skills of hearing-impaired children [2] is considers the cognitive development of students in schools and these results show that the students show very weak performances according to the results. The cognitive abilities of kids should be exercised in their early childhood (age 3-4) so this shows that the training of cognitive skills after the cognitive development period shows in efficient results.

Technological Research Gap:

The object detection in child's environment can be done in several ways. We could use CNN models for the object detection when the speed of theses algorithms it is slow (Fast RCNN, Faster RCNN etc.) .As the best approach the YOLO algorithm can be taken into consideration. Here a tradeoff with accuracy to speed is taken. The mobile application detects the objects in the Realtime, and speed of detection is very important in that perspective. And on other perspective since we are involved in a simple object detection in the Childs surroundings and not involved in any complex detections.

1.3 Research Problem

Most of these implementations focus on the sense of the giving the children to understand and learn the sign language as fast as possible. Lack of cognitive development is one of the major issues that hearing impaired students face. This issue directly affects the education ladder of them. Although their brain function normally with all respective functionalities, the sense of hearing acts the main role acting as the bridge to the normal environment. The stimulus gained through the ear directly affects the memory of the student.

When considering existing solutions, it was observed that some applications directly translate the spoken words into the sign language. An issue was observed in this approach since the translation is done real time a time of 8 seconds is taken for the translation [1]. The time is huge and may change based on the complexity of the words. The need of a reward mechanism to make children involve in the app is also need and that is not observed in many applications which is already in the market.

Some students are born with mild hearing loss and these and due to lack of the exposure to a speaking environment. The major reason for this could be having deaf parents. It is proved that the plasticity of brain changes rapidly at early childhood (age 2-4) of children [5] and need the sense of speaking need to be stimulated at this stage and the need of stimulating that is vital. This exposure to sounds and making them speak is very important for mild hearing-impaired students. It is proven that the language acquisition is automatic if exposed to it regularly [4] and if missed at early ages it will not be gained fluently afterwards [5].

2. Objectives

2.1 Main objectives

The main objective of the implementations of the object detection is to get the child with his surrounding environment and make him more passionate about his surrounding and learn through his environment and increase his vocabulary before starting the school.

2.2 Specific objective

Implement Object detection algorithm

Using a specific procedure to implement the mechanism of the make the child interactively learn by capturing the objects around his environment and learn in a self-paced manner. The identification of objects should be done according to the level of understanding of the child.

Segmentation of labeled objects to highlight the objects to enhance user experience.

As an implementation of increasing the user experience by segmenting the objects and make it clickable for the kids to increase the user experience. The segmented images can be used for the enhancement of the user experience

Incorporating the identified objects with the relevant audio cues and presenting

The identified objects will be presented to the child to in terms of learning activity to make him learn through doing simple activity to grasp the knowledge related to the word. The object is presented to him in visual and audio means.

3. Methodology

3.1 System Architecture

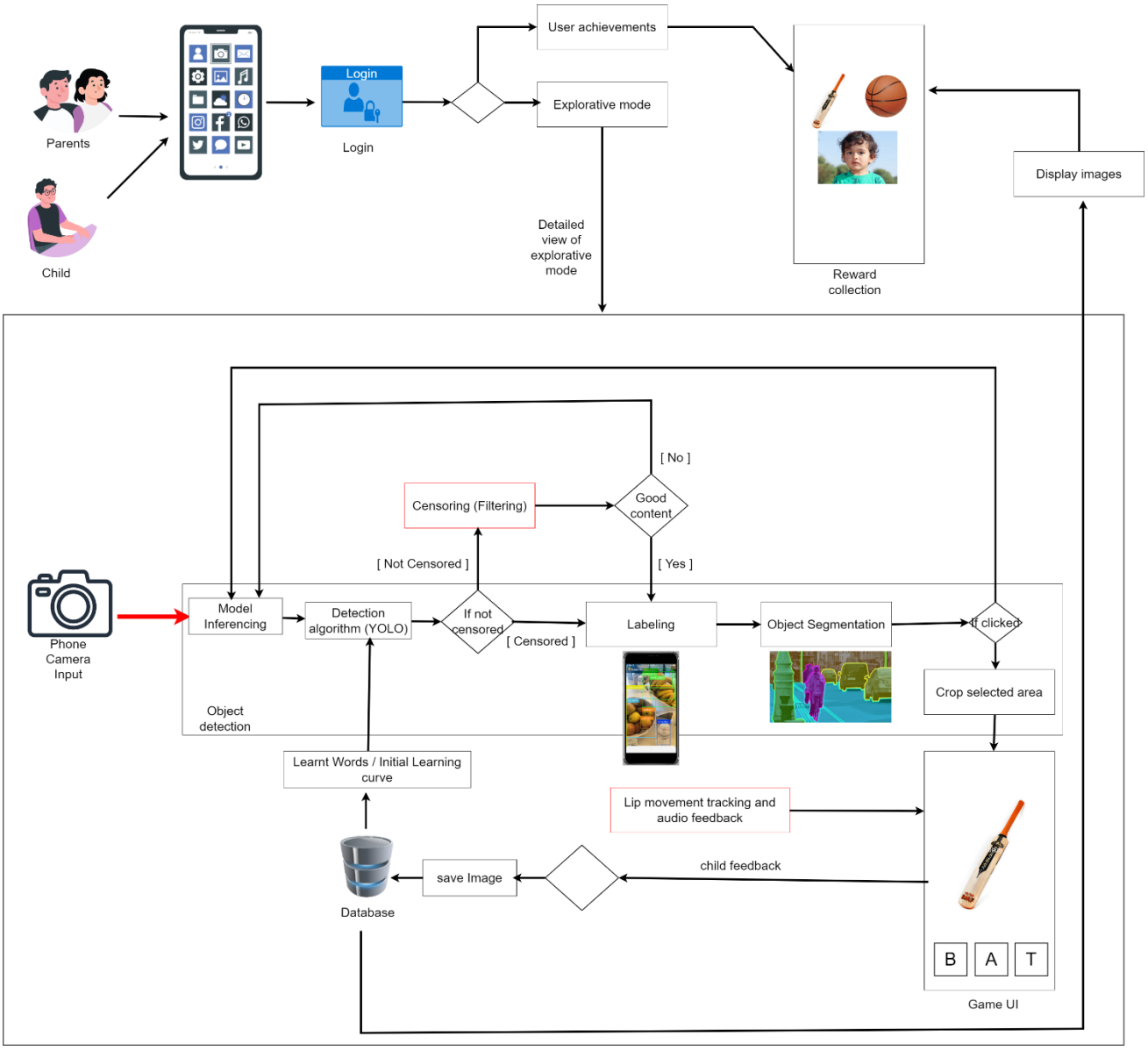


Fig 3.1.1: System architecture diagram for object detection

The proposed system will consist of several functional implementations including several aspects of improving the child learning skills and cognitive skills. Implementations related to the system will focus on making the hearing-impaired child learn from his own environment by interacting with it. The application takes a mobile based approach where the kid can easily use a mobile device with the inspection of the parents and parental supervision is considered as a must when using this application by the child. The mobile

based application consists of several mobile games and a explorative mode where the child can learn in terms of exploring his surrounding environment.

In explorative mode a video stream in the environment is captured by the child using the explorative mode camera. This footage will capture and identify the objects in the surrounding of the kid. The object identification is done with the help of You Only Look Once (YOLO) Algorithm. Based on the literature review many variants of YOLO algorithm was found and from them YOLOv5 will be used as a base version. In the object tacking multiple objects will be tracked in a single frame. The multiple bounding boxes for each object will separate the objects in the frame. The identified objects will be labeled using the students' level of understanding and a certain level of filtration is added upon the words. This filtration is done according to the level of students learning curve and in appropriate words are eliminated. If the child tackles many inappropriate contents for several times (maximum 3 times) the system will generate a message to the child to move to a different place.

The process of segmentation happens after the labeling of the objects. The segmentation will bound the object boundaries. Instance segmentation will happen at this point. A pixel wise mask will be used for these segmentations. Mask RCNN is used for this. This helps the student to click upon the objects at his preference. The clicked instance will show up the kid an enlarged image of the object he selected here. At this point the spelling of the object is taught to the child in terms of a simple game. UI based features will be used at this point to make it more interactive to the child. The words the words the child captures and learnt are added to the system database for further references. MongoDB will be used for the database.

3.2 Software Solution

The proposed system will be a mobile based application which enables the student to learn by himself. Using mobile phone, it will capture the environment of the child with the supervision of the parents and learn through his surrounding and update his vocabulary and cognitive skills. Several Games will assess the learning curve of the child and keep him track in a way he is fond about it.

3.3 Commercialization

The proposed mobile application will be mainly used by children who are hearing impaired, and this application can be used. This is very useful if the parents of the kids are also hearing impaired then the child can interact with the application and learn new words and speak. To get this sort of approach Knowledge transfer session regarding the application needs to be done to targeted audience. As initial approach KT sessions can be conducted for students and parents of children in the school of deaf in Ratmalana. Discussions are also conducted with the Centre for Education of Hearing Impaired Children (CEHIC) for having this application to be used by the students in the organization. The application can also be published in the PlayStore

4. Project Requirements

4.1 Functional Requirements

1. The system should be able to give the deaf child an interactive learning session when using the application. The child should be able to move the phone around the house and capture objects
2. A proper implementation of a reward system to keep the child track and motivate him to be using the application. The consistency of using the application is based on having a good reward mechanism.
3. Censor the child's environment and keep him track on the learning curve of the application and monitoring the learning environment
4. System should align with learning competencies of early childhood development of deaf children

4.2 Non-functional Requirements

1. Reliability – the application focuses on the learning curve of a children so the learning curve of the child should be reliable for deaf children. The object identification should happen in a reliable manner

2. Child-friendliness – the application should deal with early childhood development so the interfaces and should align to make an enjoyable learning experience for the kid
3. High performing – Several Machine learning-based models (YOLO, Mask RCNN) will be incorporated to the application, and these should be performing well in mobile and cloud environments
4. Accuracy – the Childs learning curve and the improving should be done accurately. The labeling of objects after detection and the segmentation of objects should happen in an accurate manner.

4.3 System Requirements

The system should function in the state of provisioning all the resources and getting its maximum usage. The following technologies will be used for the implementation

Mobile application – React Native used for the creation of the client mobile application with the use of it native modules and components the smooth responsive user interfaces can be created

Database – MongoDB will be used for data storage since it is a document DB with abilities for scalability.

Object detection and Image Segmentation - – Open CV or TensorFlow will be used for the running and testing of YOLO algorithm and Mask RCNN models

Code editor – VScode will be used as the primary code editor.

4.4 User requirements

Hearing impaired child

- Increasing the child's vocabulary to a level where his learning curve can reach to the level of a normal child. Learning through his environment to learn new words. The child needs to get exposed to the environment as much as possible. The child learning with his surrounding environment will enhance his learning with visual and physical exposure for relevant objects

Parents/ Teacher

- Getting their hearing-impaired kid learn a certain amount of vocabulary when learning before staring elementary education. Identifying the names of the objects in the environment with also increasing the cognitive skills. Ability to monitor the learning environment of the child. Ensure the child is exposed to relevant content when capturing the images from the environment.

Requirements Gathering

Collecting information for initial investigations

Had visit to the School of Deaf in Ratmalana and had the initial discussion with the teacher regarding the key areas to be focused when implementing a system for hearing impaired students and discussion about the initial understanding about the learning curve of the student. Furthermore, discussions related to teaching techniques used by teacher were also considered into discussions. As results of these discussions understood the key areas are the development of cognitive skills and the vocabulary of the students at early childhood will enhance the learning experience of the students.

Furthermore, involved in discissions with Audiology unit of the school and had information related to speech testing of early childhood development. And there understood since a mobile is accessible to many people the approach of a mobile application was taken at this point.

Collecting data

The development of the object detection model is based on the child living environment. Any dataset related to object detection can be used for the training purposes of the application.

4.5 Wireframes

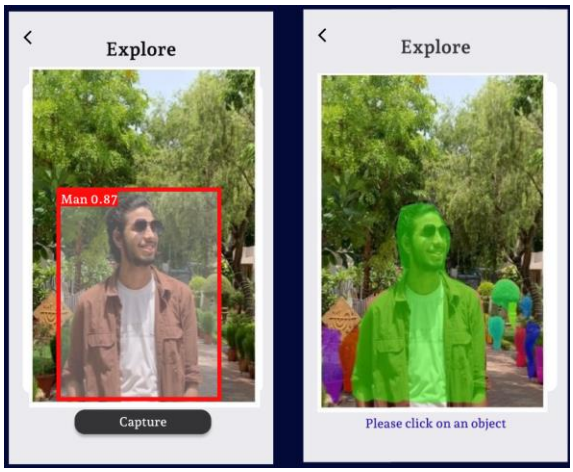


Fig 4.5.1: wire frames for the explorative mode object detection and image segmentation

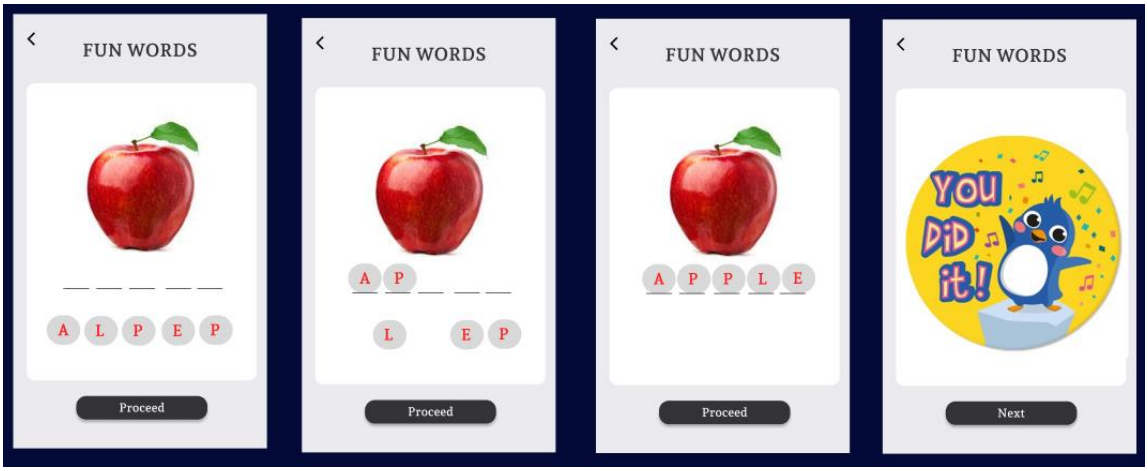


Fig 4.5.2: wire frames for the explorative mode game

4.6 Use Case Diagram

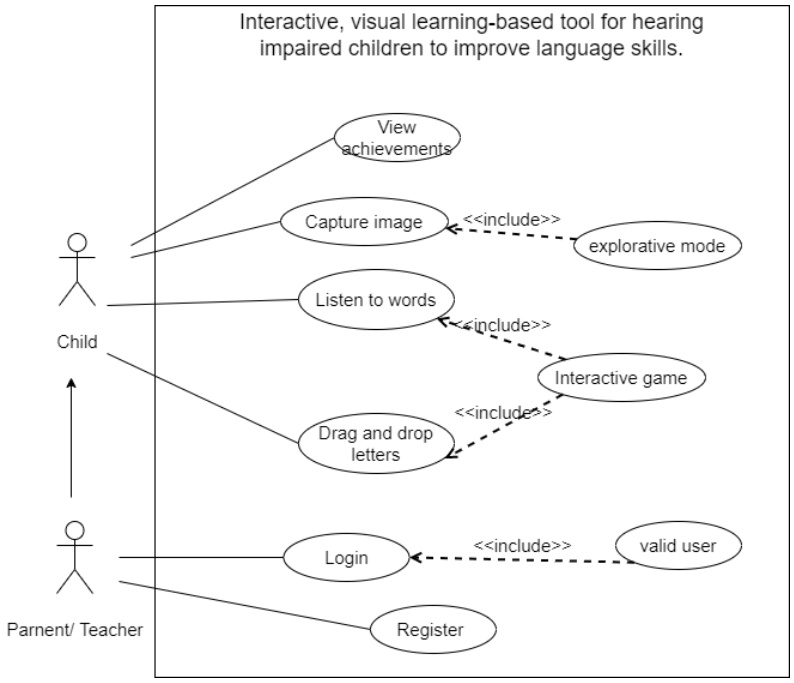


Fig 4.6.1: use case diagram

5. Gannt Chart

Table 5.1: Gannt chart

	2022							2023					
	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN
Topic Selection													
Initial Supervior Discussion													
Idea refining (Brainstorming)													
Requirement Analysis													
Literature Review													
System Overview Diagram													
Topic Assesment													
Individual Component diagram													
Proposal document													
Design Wireframes													
Data collection													
ML model developing													
Testing													
Bug Fixing													
Research paper													
Final reports													

5.1 Work breakdown chart

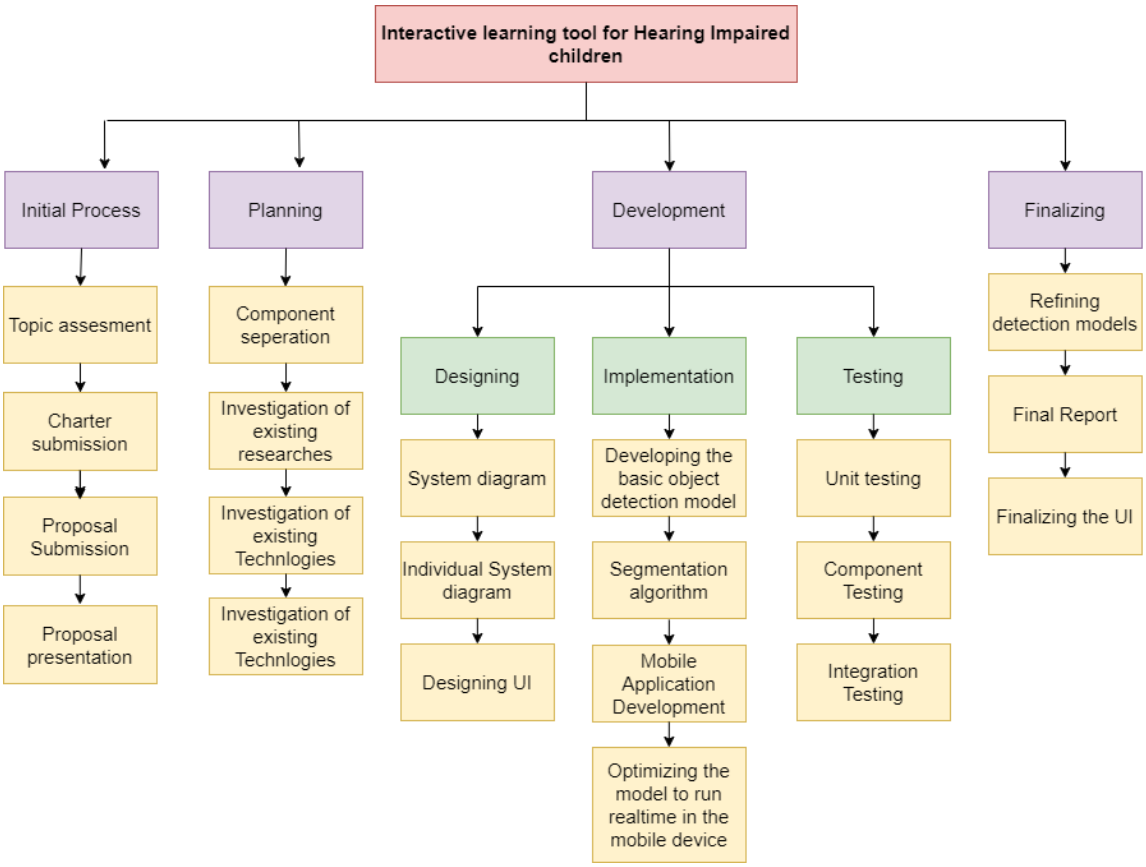


Fig 5.1.1: work breakdown chart

6. BUDGET AND BUGET JUSTIFICATION

Table 6.1: Average budget with costs

Requirement	Average cost
Travel cost (Field Visit)	15000
cost for Mongo DB (monthly)	20712.57
Hosting the app in PlayStore	9084.46
Azure Cloud services	15000
Average Total cost	59797.03

References

[1] Paulo Martinsa, Henrique Rodrigues, Tânia Rochaa, Manuela Franciscoc, Leonel Morgadoa, “Accessible options for Deaf people in e-Learning platforms:technology solutions for Sign Language translation” in *6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2015)*

[2] Sandipa Roy, Arpan K Maiti, Indira Ghosh, Indranil Chatterjee, Gopal K Basak, Kuntal Ghosh, “An app based unified approach to enhance language comprehension and mathematical reasoning ability of the hearing-impaired using contrast words”

[3] Filomena Soares, João Sena Esteves, Vitor Carvalho, Gil Lopes, Fábio Barbosa, Patrícia Ribeiro, “Development of a serious game for Portuguese Sign Language” in *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT*)

[4] M. Hall, W. Hall, and N. Caselli, "Deaf children need language, not (just) speech",First Language, vol. 39, no. 4, pp. 367-395, 2019.

[5] Wake, M et al. “Hearing impairment: a population study of age at diagnosis, severity, and language outcomes at 7-8 years.” Archives of disease in childhood vol. 90, no. 3, pp 238-44, 2015

[6] Joseph Redmon∗, Santosh Divvala∗†, Ross Girshick, Ali Farhadi∗†, “You Only Look Once: Unified, Real-Time Object Detection”, in *2016 IEEE Conference on Computer Vision and Pattern Recognition*

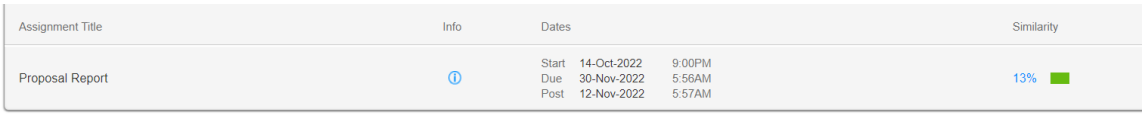
[7] Peiyuan Jiang, Daji Ergu*, Fangyao Liu, Ying Cai, Bo Ma, “A Review of Yolo Algorithm Developments”, in *The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021)*

[8] Upesh Nepal and Hossein Eslamiat, “Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs”,

[9] Xiuli Li[1,2], Yi Qin[1], Fujie Wang[1], Feng Guo[1], John T. W.Yeow , “Pitaya detection in orchards using the MobileNet-YOLO model”, in *39th Chinese Control Conference July 27-29, 2020, Shenyang, China*

Appendix

Turn it in report



Assignment Title	Info	Dates	Similarity
Proposal Report	ⓘ	Start 14-Oct-2022 9:00PM Due 30-Nov-2022 5:56AM Post 12-Nov-2022 5:57AM	13%