Smart Scholarship Helper for Sri Lankan Grade 5 Scholarship Examination

Dhammika H. De Silva
Senior lecturer, Department of Computer Systems Engineering
Sri Lanka Institute of Information Technology Colombo, Sri Lanka

R. M. H. S. Gunasekara
Undergraduate, Department of Information Systems Engineering
Sri Lanka Institute of Information Technology Colombo, Sri Lanka

I. P. Abeygunawardana
Undergraduate, Department of Information Technology
Sri Lanka Institute of Information Technology Colombo, Sri Lanka

M. B. Peiris
Undergraduate, Department of Information Systems Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka

D. A. T. S. Dodangoda
Undergraduate, Department of Information Technology
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka

ABSTRACT
Education is a permanent resource for the development of a country. Education faces a so many challenges among the community. In this respect, this research is based on the educational problems which are faced by grade 5 Scholarship students. In Sri Lankan education system, Scholarship examination is the competitive examination for primary students. The proposed research is to build a platform for grade 5 Scholarship students to learn and get the correct knowledge about the Scholarship syllabus to pass the examination well. This is the smart helper for students that covers all the main areas of the Scholarship exam paper such as mathematics, environmental science and language base. The environmental science section is mainly based on the theory of object recognition using model training, mathematical section is based on the algorithm creation and language based section is using natural language processing. On the other hand, teachers can create a group of students where teachers can upload papers and homework. Students can answer them and check their answers with the answers in the system which are automatically generating and provide the student analyzing report using the data mining process. The key goal of this research is to maximize the student performance and knowledge for the Scholarship examination.

General Terms
ML - Machine Learning

AI – Artificial Intelligence

OCR – Object Character Recognition

NLP – Natural Language Processing

Keywords
Object Recognition, Algorithm Creation, Natural language processing, Machine Learning, Artificial Intelligence, Data Mining

1. INTRODUCTION
The Grade 5 Scholarship Examination is the most competitive Sri Lankan examination offered by the Department of Examinations of the Ministry of Education, Sri Lanka. Based on the results of the exam, students can be transfer to the prominent national schools. To get through this examination students have to engaged with more and more studies and revision tasks on examination papers (past papers, model papers, question related to the subject materials) .Students have to search for the papers and answers manually or consult a teacher and teacher or parents can get the student analyzing report. This is a huge time consuming task. In order to overcome of these identified problems, we have come up with this research project.

This mobile application platform is a practicable system, specially designed for the Scholarship examination. This system provides an extra feature comparing with others as a way to input the materials to teachers or any user. Students can get answers of the papers through the system. This application is based on three modules named as mathematics, environmental science and language conversion. In the system, past papers with answers are available and model papers and homework can be uploaded to the system and answers are automatically generated. Teachers can give the feedback easily for students after they complete those. Through the system students can answer for MCQ questions as well as written questions.

The proposed system is not limited to get the answers for past question papers stored in the system or uploaded model papers or homework. Students can get answers for other external questions by capturing, scanning or typing using the camera. Students can get solutions for mathematical questions in step by step answering, can get a clear identification of the objects in the images and can get answers for the language based questions. Finally considering this system into Scholarship examination, there is also a way to get the performance of the students by analyzing the task history.

2. BACKGROUND
The current Scholarship based apps are following only question basis methods. This is one of the main Scholarship examination help software focus on the primary students. The only way to answer the question is not an effective way to increase student knowledge. The proposed solution is to develop a platform that would provide the Scholarship knowledge areas for the student.

Students can get knowledge about syllabus with mathematics questions with step by step answers, object recognition questions, language conversions, model papers and answers, past papers and answers. Teachers can upload homework to the system and students can answer the questions. Results will be generated automatically. Therefore using this platform students can get proper knowledge and there is no issue about the knowledge.

2.1 Literature Review

2.1.1 Automatic Short Answer Marking

Computational linguistic techniques used for marking short free text responses automatically through this research work. They are dealing with answers ranging from a few words up to 5 lines for factual science questions that typically ask candidates to state, describe, suggest, explain, etc. and where there is an objective criterion for right and wrong. These questions are from an exam known as GCSE (General Certificate of Secondary Education).

They realized that standard syntactic and semantic analysis methods will be difficult, because have to look for different meanings of sentences like uses of a negation of negations, need for reasoning and making inferences, contradictory and inconsistent information. They tried to accurate with natural language processing however there also issues that sentences like unconventional expression for scientific knowledge, consistency across answers etc.

For information extraction process this research used Hidden Markov Model part-of-speech (HMM POS), Noun Phrase (NP) and Verb Group (VG) finite state machine (FSM) chunker techniques with different corpuses tuned by some their own modifications. Use some patterns to collapse all paraphrases into one by use of knowledge.

For marking the answers this research used machine learning algorithms with different methods such as Inductive Logic Programming (ILP), Decision Tree Learning (DTL) and Naive Bayesian learning (Nbayes). Neighbor classification also tested. Test the results with non-annotated data as well as annotated data using those techniques this research achieve some point of accuracy. As the conclusion this research said that Bayesian learning method is perform better than the Decision tree model and also when the pattern matching approaches accuracy level increased for some point. Though Automatic Short Answer Making, these methods are not accurate enough at present to be a real alternative to the hand- crafted, pattern matching approach and them furthermore investigating on pattern writing. [1]

2.1.2 Object recognition

Object identification is a vital part in the environmental science section of the paper because primary students have to identify the things in the surroundings. Object recognition is an important task in image processing and computer vision. It is concerned with determining the identity of an object being observed in an image [2].The recognition problem is considered as one of matching appearance rather than shape. The appearance of an object in a two dimensional image depends on its shape, reflectance properties; pose in the scene and the illumination conditions [3].In this scenario of object recognition, a model has to be trained using neural networks to identify the objects. In the last few years, deep convolutional neural networks have shown to be a reliable approach for image object detection and classification due to the relatively high accuracy and speed [4].So from this trained model, we can give the correct identity of the objects within a limit.

2.1.3 Mathematical equation detection and answer recognition

Mathematics is a challenging subject for most primary school students. Students need to pass scholarship exam with math during primary school. This study examined relations between children's conceptual understanding of mathematical equivalence and their procedures for solving equivalence problems (e.g., 3 + 4 + 5 = 3 + 9). Students in 4th and 5th grades completed assessments of their conceptual and procedural knowledge of equivalence, both before and after a brief lesson [5]. While recognition and retrieval of textual information is fairly mature, with wide-spread availability of optical character recognition and text-based search engines, recognition and retrieval of graphics such as images, figures, and mathematical expressions are in comparatively early stages of research. In math recognition there are four key problems (detecting expressions, detecting and classifying symbols, analyzing symbol layout, and constructing a representation of meaning) [6].

Automatic recognition of mathematical expressions is one of the key vehicles in the drive towards transcribing documents in scientific and engineering disciplines into electronic form. Note that some existing systems for offline recognition of mathematical expressions deal with data that contain both text lines and mathematical expressions in the same document Extraction of mathematical expressions and other similar objects (for example, diagrams, graphical drawings, chemical equations, etc.) from documents requires similar techniques. This indeed can be a large topic by itself [7]. Our survey, however, will only cover the recognition of mathematical expressions. It is hoped to get mathematical publications from question papers and to make the students ready to obtain correct answers for such mathematical statements.

2.1.4 Text recognition and Technical Review on Text Recognition in images

Store and search and use the multimedia activate the content based image and video/ audio analysis to usage of the media text recognition a wide application area [8]. Main purpose of this research is to improve image recognition system to make paper based historical documents to computer based information [9].

3. METHODOLOGY

3.1 Creating a machine learning model to categorize the question

Input the question and is has to be categorized to the relevant module, which are Image recognition, Mathematical equations and Language conversion. For implement this machine learning part almost used language is Python Language and technical that are used in is Natural Language Processing and model training.

3.2 Creating a homework uploading system and automatically results generated system

Create an artificial intelligence to provide the answers for uploaded homework and model papers. Students can answer the multiple answer questions by ticking and written question answers by typing and then answer upload the system and AI identify whether the answer is correct or incorrect. If the answer is incorrect, the correct answer will be provided and students can get the correct answer. For the implement Python language and natural language processing are used.

3.3 Creating a student analyzing report

Data mining is used for the implementation. As the output of this part is generated the student analyzing report. It can be identified the student's weak modules and percentage of knowledge of each modules.

3.4 Creating an Object Recognition module

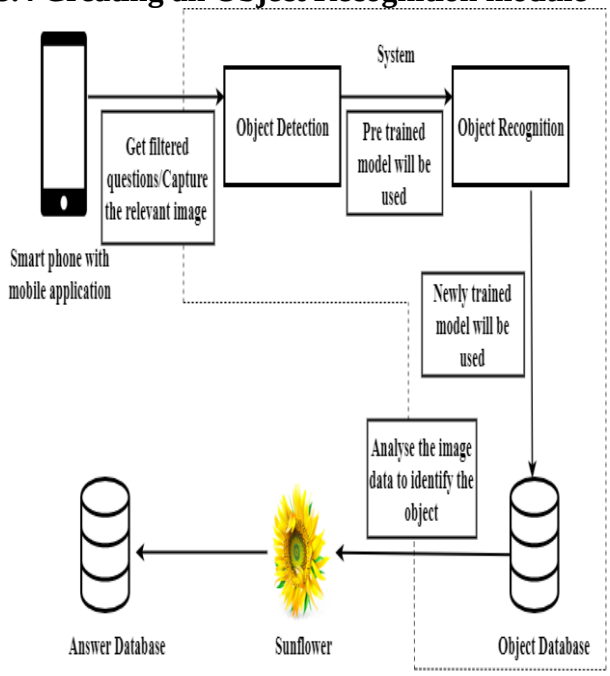

Figure 1: Architecture diagram of object recognition

In order to capture an image in a hard copy of a paper, back camera of the smart phone can be used and answers can be taken for the images in soft copies also. After capturing, images are uploaded to the system. The concept of neural networks to develop the proposed object recognition system. Neural network is an attempt to make a computer model. In here Node.js which is an open source, cross-platform for developing server-side, Tensorflow.js which is a software library or framework that can train and run deep neural networks for image recognition, Keras.js which is a high-level Python library run on top of TensorFlow framework and Numpy.js is a python package and a library consisting of multidimensional array objects are used to design the neural network and using Tensorflow and Keras a model is trained to identify the objects. This is trained by limiting the objects to the scholarship syllabus by two data sets, one set of images to train the network and other set of images to evaluate how accurately the network learned to classify images. This is the backend of the system and Visual studio code which is a source code editor is used to code the program. Android development is done using the Android Studio. Finally the identification of the image is displayed on the screen as the output.

3.5 Creating a Mathematical equation, statement detection and answer recognition module

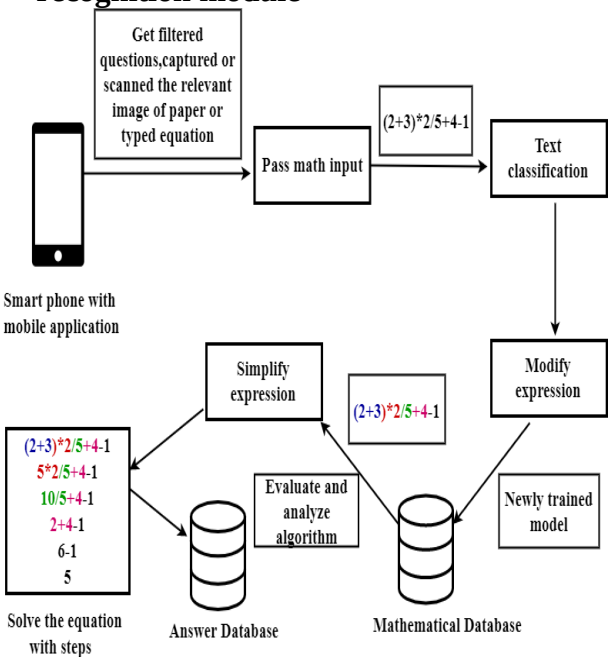

Figure 2: Architecture diagram of Mathematical solver

New optimized algorithms are developed to identify mathematical equations. It has been added to six basic mathematical expressions (Parenthesis, Exponents, Multiplication, Division, Addition and Subtraction). Evaluating algebraic expressions can be a simple process, but needs to follow an order of operations to get the right answer. The sequence details the order we follow to add, subtract, multiply, and divide. The order is:

P.E.D.M.A.S.

Parenthesis | Exponents | Division | Multiplication | Addition | Subtraction

1. Perform the operations inside the parenthesis first
2. Then exponents
3. Then division and multiplication, from left to right
4. Then addition and subtraction, from left to right

Implemented the algorithms to evaluate those equations using P.E.D.M.A.S order and used prologs to evaluate statement equation. Text Classification using Neural Networks, Optical Character Recognition (OCR) to identify the words and equation to reproduce the steps of the statement equation from Google API. To interpret a mathematical equation in answerable form, usually requires decoding the equation or captured picture. Character recognition is a procedure of converting images of handwritten, typewritten or printed text into text.

3.6 Creating a language conversion module

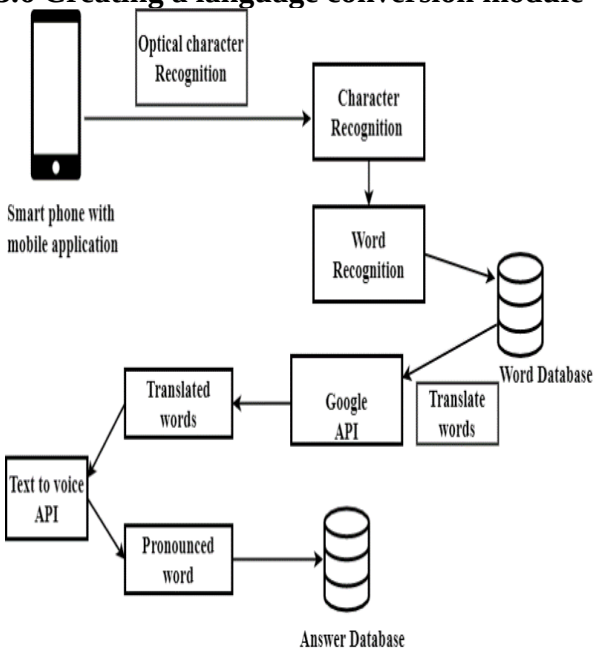

Figure 3: Architecture diagram of Language converter

The main objective of this part is to translate the words related to the 5th grade scholarship program. Language translations are made between English, Tamil and Sinhalese. The child only needs to enter the word and select the language, and then the translation into two other languages is done. Also children can enter the word as a keyboard input. Translations are done by the natural language processing when taking a picture. The app can take a photo of the word needed for translation. There is also a function to pronounce words in Sinhalese and English Tamil. In this part, use two API for the word translation process. The process of translation is done through Google translate API. Text to voice API is used for word pronunciation. Text classification is done using Optical Character Recognition (OCR) to identify the letters and words. The words are decoded from the captured image before the character recognition. The system allows recognizing the text that in handwritten, typewritten or printed form.

4. RESULTS

4.1 Question Categorization model

Question categorization model is used to identify the relevant module whether Language Conversion, Environmental Studies or Mathematical Equation. This model was trained with model questions (about 500 questions). The model was tested with 100 questions that relevant to the Grade 5 Scholarship Examination. The model was able to categorize them with an accuracy of 95%.

4.2 Object Recognition and Answer Generating model

In the environmental science section, object recognition is the main function as our research. In the scholarship exam paper there are two categories of questions relevant to the object recognition as finding the odd picture and finding the name of an object. This is done by training four neural networks as odd neural network(CNN 1),Animal neural network(CNN 2),Flower neural network(CNN 3) and Vehicle neural network(CNN 4).The accuracies of the neural networks are as following,

CNN 1- 95%

CNN 2-97%

CNN 3-96%

CNN 4-94%

The accuracy of the system is depend on two points as number of layers in the neural network and number of dense in the neural network.

4.3 Mathematical Equations and Answer Generating model

Using this model the students can get the answers with step by step for relevant mathematical equations. Problem-solving methods were not measured. Since the questions are multiple choices, students can use test-taking strategies such as checking every available choice to get the right answer. When generate the answer this model carry out the order of operations (BODMAS) for mathematical expressions. This model was trained with model questions and past paper questions which are related to the Grade 5 scholarship examination on selected criteria. The model was able to answering the relevant expressions with an accuracy of 97%.

4.4 Language Translation and Answer Generating model

A language translation is done between English, Tamil and Sinhala languages. Translations process can be done by keyboard entering, capturing and scanning. The system allows recognizing the text those in handwritten, typewritten or printed forms. Printed words can get translated by capturing the image. The model was able to answering the relevant questions with an accuracy of 90%.

Table 1: Accuracy Level of Results

Name of the Module	Accuracy
Question Categorization model	95%
Object Recognition and Answer Generating model	95.5%
Mathematical Equations and Answer Generating model	97%
Language Translation and Answer Generating model	90%

5. CONCLUSION

To get through this grade 5 Scholarship examination students have to engage with more and more studies and revision tasks on examination papers (past papers, model papers, question related to the subject materials) .So they have to search for the papers and answers manually or consult a teacher and teacher or parents can get the student analyzing report. This is a huge time consuming task. To overcome these identified problems, come up with this research project. This mobile application platform is a practicable system, specially designed for the Scholarship examination. This system provides an extra feature comparing with others as a way to input the materials to teachers or any user. And also students can get answers the papers through the system. This application is based on three modules called Mathematics, Environmental science and Language conversion. In the system, Past papers with answers are available and model papers and homework can be uploaded to the system and answers are automatically generated.

Teachers can give the feedback easily for students after they complete those. And also in the system student can answer for MCQ questions as well as written questions. The system is not limited to get the answers for question papers stored in the system or the uploading model papers or homework. Students can also get answer other external questions by capturing, scanning or typing. And students can get solutions for mathematics questions in step by step answering, can get a clear identification of the objects in the images and can get answers for the language based questions. Finally considering this system into Scholarship examination, there is also a way to get the performance of the students can be get through the student analyzing report.

6. REFERENCES

[1] Stephen G. Pulman, Jana Z. Sukkarieh," Automatic Short Answer Marking", Centre for Linguistics and Philology, Centre for Linguistics and Philology, Walton St., Oxford, OX1 2HG, UK,2012

[2] Reetu Awasthi, "Techniques for Object Recognition in Images and Multi-Object Detection". International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 4, April 2013.

[3] Besl,P.J,and Jain,R.C.1985. "Three-Dimensional Object Recognition,"ACM Computing Surveys,Vol.17.

[4] Sherrah, J. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery.

[5] Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? Journal of Educational Psychology, 91(1), 175-189.

[6] Richard Zanibbi, Dorothea Blostein. (2012).Recognition nd retrieval of mathematical expressions.International Journal on Document Analysis and Recognition (IJDAR),15,331–357. https://link.springer.com/article/10.1007/s10032-011-0174-4

[7] K.-F. Chan, D.-Y. Yeung: Mathematical expression recognition: a survey

[8] https://www.researchgate.net/publication/37432860 A Survey of_Text Detection_and Recognition_in_Images and_VideosA SURVEY OF TEXT DETECTION AND RECOGNITION IN IMAGES AND VIDEOS Datong Chen, Juergen Luettin, Kim Shearer

[9] https://www.aclweb.org/anthology/W05-0202 Automatic Short Answer Marking Stephen G. Pulman Jana Z. Sukkarieh

[10] Schantz, H.F. (1982). History of OCR, optical character recognition. Recognition Technologies Users Association.

[11] Wang, S., Uchida, S., Liwicki, M., & Feng, Y. (2013).Part-based methods for handwritten digit recognition. Frontiers of Computer Science, 514-525

[12] Behr, M., Erlwanger, S., & Nichols, E. (1980). How Children View the Equals Sign. Mathematics Teaching, 92, 13-18.

[13] Cai, J., & Moyer, J. (2008). Developing Algebraic Thinking in Earlier Grades: Someinsights from International Comparative Studies. In C. Greenes, & R. Rubenstein (Eds.), Algebra and Algebraic Thinking in School Mathematics (70th Yearbook of the National Council of Teachers of Mathematics, pp.169-180). Reston, VA: NCTM.