Evaluation of Handwritten Answer Scripts Using Machine Learning Approaches

Ravikumar.M, Sampath Kumar.S and Shivakumar.G

Department of Computer Science, Kuvempu University, Shimoga, Karnataka, India.

ravi2142@yahoo.co.in, sampath1447@gmail.com.

Abstract: Learning is the fundamental aspect for human beings to gain knowledge. Educating students to learn and to give assessment to check their ability plays an important role. Examination is the key to evaluate he/she scored. students work is a central aspect of the teachers in term of evaluation. Though multiple criteria affect the assessing of student's work. Also, there are several time-consuming process that affects the departments like lot of answer scripts to evaluate, marking mistakes, errors in totalling. so, we are developing a robust tool to automate the short answer using machine learning techniques. There are two modules in the first module we use OCR to extract handwritten answer from the answer script and word is recognized. In the second module, handwritten answer is evaluation is done with this approach, computational time of manual processing can be reduced. Finally, the result gained is 90 percent.

Keywords: Handwritten answer, Segmentation, Recognition, Evaluation

1.INTRODUCTION

Automation in every aspect of life has emerged which is often used for today's society. The basic method to assess the learning capability of students is by answer in the exam written by students. It might be an accurate measurement if the system is automated to evaluate. Faculty must put entire effort into reading handwritten answers and evaluating all scripts in the current evaluation system. when evaluating, teachers has to be considerable deal with the scripts. The time needed to correct and to assign marks for the written answers depends on the number of answer scripts. An average teacher spends twenty to twenty-five minutes correcting answer scripts. Then upon calculation of 50 scripts the time spent for evaluation makes a concern. At present, grades of the students is given after the evaluation. The grading involves data entry of all marks to the concerned students in a class and the script has to stored safely. This process again

takes a considerable amount of time. A delay in manual evaluation would cause students discomfort, as it is a time-consuming process.

The grading of huge volumes of handwritten answer by the hand is a difficult procedure that is assured to be tensed with mistakes. In the education system incorporation of evaluation tool will help in teachers in many ways. By automating answers, large-scale evaluation will be more efficient. with growing of more needs and giving importance to education, educational institutions needs certain tool in terms of saving time of teachers and to give more time involved in teaching. Many educational institution across the globe use pen and paper method for most examination. This is what put the pressure on teachers to reduce the time on academics of students. The evaluation tool is needed with sole purpose of making the correction easier and less monotonous. To build a relevant and robust evaluation tool is difficult as it has to understand the students handwritten styles. The font of every student vary from each other. It is easy for the trained system if the handwritten style is considerable clear and neat and this will help students to work on writing quality or ability of improving feedback received from teachers. Though other similar tools are available which works on grammar, keywords matching similarities. The developed evaluation tool mainly works on handwritten answers. Answer script evaluation is crucial part of student assessment. In order to assess students, teachers take various methods, such as answering short questions, answering descriptive questions, and answering multiple choice questions. The implementation of evaluation tool in an education system is of multiple choice question and short answer. It is easy and less time consuming as compared with descriptive question answering and it takes more time to Evaluate.

$$f(x) = max\,(0,\,x)$$

In this case, the variable x is an input to the function. In our proposed algorithm, the length of the answer will also be considered as a parameter. The written answer length will be taken as an input from the teacher [5]. A teacher's input will be taken into account when determining the length of the written response. As a result, the paper is divided into the following sections: Section 2 contains a review of related works. Section 3 gives brief idea of proposed method and techniques used. The results & discussion of this research are presented in section 4. The conclusion is explained in section 5.

2. RELATED WORK

According to this review paper [1], subjective answer evaluation based on keywords is pro- posed. An application is developed to evaluate the descriptive answer which contains login module, information extraction module, weighting module and score generation module. All these above mentioned module formed together for the purpose of evaluation. The system is designed such that five different answers is provided for five users. Every answer is stored in database with meaning and keywords. Answers written in English language are only evaluated in the current system.

Analysis of results clearly mentioned that the semi-automated evaluation works well than other two modules. It has been proposed that an MCQ form with one-word answers can be used for automated evaluation [2]., different techniques of examination process which provides quick and reliable evaluation are discussed. Blooms taxonomy represents Different levels of complexity in increasing order from the bottom to the top The weight of each level should be maintained when evaluating different levels of knowledge. For better implementation, the paper concludes with the latest technology involved. A brief survey on techniques used for evaluation is discussed [3], natural language processing is used for machine translation, text summarization etc. latent semantic analysis technique based on mathematical model, which helps in evaluating the quality of student's answer. The next technique used is BLEU. This system is used for evaluating answer in English and Spanish language. Here standard answer is compared with student answer. Finally, marks are assigned based on the correct answers.

Real-time datasets are collected from students. The datasets of students are categorized into brilliant, average and better. The answers are stored in text file. The keywords are converted into vector using hyperspace analog to language algorithm. A comparison is made between SOM results and clustering methods. Better performance is achieved with SOM. Accuracy is calculated based on answer. we also find some more approaches for answer evaluation [7, 11, 14, 15]

3. PROPOSED METHOD

In this section, we discuss the proposed method for evaluation of answer script

A few works have been done on handwritten recognition with standard dataset. Dataset is collected from students handwritten answers. Question paper is formulated in such a way that each answer script contains 10 question that is in one-word answer. In the step 1,

students written Script are scanned and preprocessed. Preprocessing techniques contains removing of shadow or dark spots, ink smudges, printer scales while scanning etc. it is necessary to process the answer script to get necessary data. It helps in saving of memory required to store the answer script data. In order to get rid of variations in characters' page layout fitment is done to standard size. When the answer scripts are normalized to one size. Few techniques have been done to before moving to next step i.e. skew is a rotation transformation that brings the word orientation back to parallel to the horizontal. Slant is corrected which attempts to make all the vertical strokes erect. In the step 3, based on the baseline by applying horizontal line detection method length of the baseline is calculated and coordinates x, y, w, h values are applied to subtract the horizontal line which is present below the handwritten answers. The next step is training of model answers, the segmented words are trained by passing ottakshara and without ottakshara characters. Now using CNN words are trained and stored. if the extracted features and trained characters' matches with the in- put, the handwritten word is recognized in the step 5. The recognized handwritten answer is converted to printed text by comparing with trained labels and stored in the python flask web. Finally, based on the factors like word recognition and written handwritten answers evaluation is done, marks are assigned. Based on handwritten answer in the script and model answers in the database. The application will pro- vide marks in the range of 5 marks for each correct answer. Function to calculate percentage of marks for handwritten answer matched is given in the equation (1).

$$MAE = \frac{\sum^{n}_{i=0}}{n} |\ y_i - \hat{y}_j| \qquad (1)$$

Calculating Mean Absolute Error for above mentioned. Mean absolute error is calculated between the marks calculated by our proposed algorithm and the evaluated samples that were evaluated by the teachers. It is calculated by taking difference between marks calculated by our proposed algorithm and the marks calculated by teachers evaluated answer scripts. There is a possibility that difference can also be negative so absolute values are only considered while considering the sum of all differences as shown in above formulae for MAE.
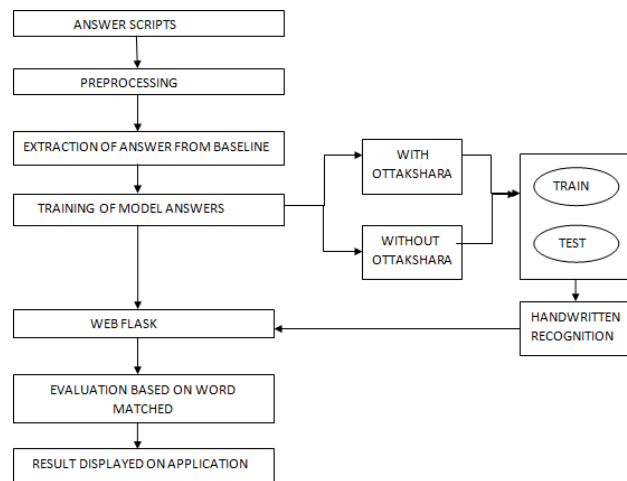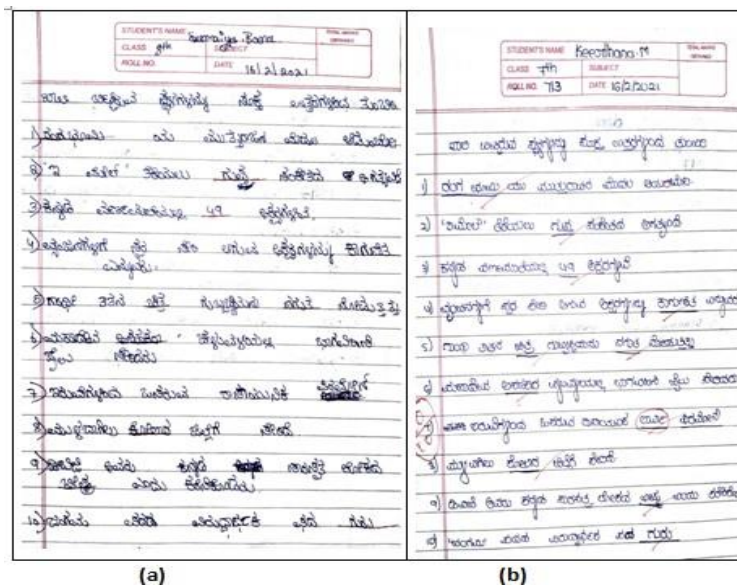
Fig. 1. Various Stages involved in the Evaluation process.

Function to calculate the number of marks scored for Handwritten Answer Matched (HAM) and is given in equation 2.

$$HAM = \frac{MaxMarks \ X \ Answer \ Matched}{Total \ Answer}. \qquad (2)$$

Fig. 2. Sample of Handwritten Answer script (a) with ottakshara and (b) without ottakshara.

4.RESULTS AND DISCUSSION

For the purpose of experimentation, we have collected real-time handwritten answer scripts from different government schools. we categorized dataset into two types i.e. 1. fully handwritten dataset where question and answer both are in the form of handwritten style, where it is further differentiated into two types with ottakshara scripts and without ottakshara scripts. 2. printed and handwritten dataset.

Here, combination of questions is in the form of printed and baseline is provided to write the answer above the horizontal line and answers is in a handwritten for- mat written by different students. which consists of around 500 scripts of Kannada language. To demonstrate the efficacy of the proposed method, teachers must evaluate answer scripts with handwritten answers from more students and recognize them with OCR. Our research work aims at evaluation of one-word handwritten answer scripts. Most of the handwritten words are present with ottakshara with one or more touching characters. The extraction of answers from baseline and to segment

the character individually is a challenging task. SVM and CNN is used for word recognition. The obtained word recognition accuracy is 98.6%. The current manual evaluation takes about 2-3 minutes to evaluate one-word answer for single answer script. The proposed system takes about 1 minute to evaluate the answer. The built system helps in less time consuming to evaluate an answer. It is efficient as compared with manual answer evaluation system. Teachers don't have to spend time and effort evaluating answers anymore because the tool does it for them. By using the evaluation tool results of the students will also be given in a short period with less effort. Manpower like storing of answer scripts, keeping the marks record of each register number and keeping it regularly can be left off. Where the tool stores the total marks of each student obtained.

4.1 Handwritten Answer with Ottakshara

Answer scripts containing kannada handwritten words with ottakshara is collected for experimentation, the answer script 1 contains 200 extracted answers of same word in the whole dataset and answer recognized answer is 185 words out of 200. From the table 1 the highest accuracy gained is 98.6 percent and least percentage is 81.3 percent of answer script 7.

Table 1. Handwritten Kannada Words with Ottakshara Characters.

Answer Script	Number of words	Recognized Answer	Recognition Rate
1	200	195	97.5%
2	160	158	98.75%
3	150	148	98.66%
4	150	140	93.33%
5	150	146	97.33%
6	150	147	98.0%
7	150	147	98.00%
8	150	146	97.33%
9	160	156	97.05%
10	150	148	98.66%

Table 1 illustrates the answer script with ottakshara present in words. here, second column explains the total number of same words present in the answer script. The third column shows the result of total number of answers recognized with number of words in first column and at last the average recognition rate of particular answer is mentioned.

Table 2. Recognition Accuracy of Handwritten Kannada Words without Ottakshara Characters.

Answer Script	Number of words	Recognized Answer	Recognition Rate
1	200	185	92%
2	160	150	93.45%
3	150	142	94.66%
4	150	125	83.33%
5	150	128	85.33%
6	150	148	98.6%
7	150	122	81.3%
8	150	122	81.3%
9	160	150	93.45%
10	150	125	83.33%

4.2 Handwritten Answer Without Ottakshara

Answer scripts containing kannada handwritten words without ottakshara is collected for experimentation, the answer script 1 contains 160 extracted answers of same word in the whole dataset and answer is recognized. Recognized words is 158. From the table 2 the highest accuracy gained is 98.75 percent and least percentage is 97.05 percent of answer script 9.

Table 2 shows that recognition accuracy of kannada answers without ottakshara characters. in the step 1, number of handwritten answers is extracted and values are mentioned in column 2. In the third column answer is recognized based on extracted number of words. Finally, recognition rate of particular answer is labelled.
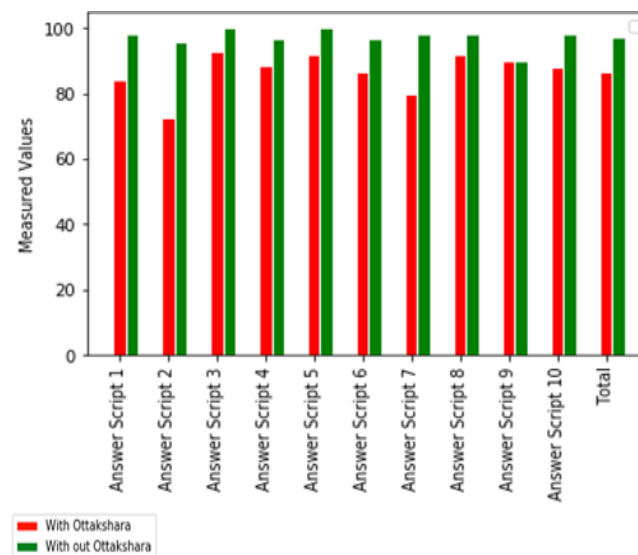

Fig. 4. Graphical representation for Measured Values of Ottakshara and without ottakshara Answers.

Figure 4 shows that the measured values of kannada handwritten answers with ottakshara and without ottakshara scripts results. It is observed that the scripts without ottakshara shows high accuracy of 98.66% result. when compared with ottakshara words.

Table 3. Evaluation Parameters of Answer Script.

Answer Script	Number of Answers	Number of Answers Matched
1	10	08
2	09	09
3	10	08
4	10	08

5	09	09
6	08	08
7	10	10
8	10	10
9	09	08
10	10	10

4.3 Handwritten Answer with evaluation parameters

For the purpose of evaluation of answers, the extracted answers are compared with trained answers which are already in web flask. The answers are matched with given answers and parameters of results is displayed.

Table 3 illustrates that the parameters like number of answers, number of an- swers matched, the machine calculate the marks to be allotted to a particular answer.

The bar graph 5 shows that the samples ranging from answer scripts 1 to 10. The number of answers matched by proposed system is less in comparison to the teacher's evaluation. The answers ranging from 1,5,7,8 almost coincide with each other which shows high accuracy. For answer script like 1,3,4,9 are doesn't match noticeable result. so, there is a difference in number of answer matched.
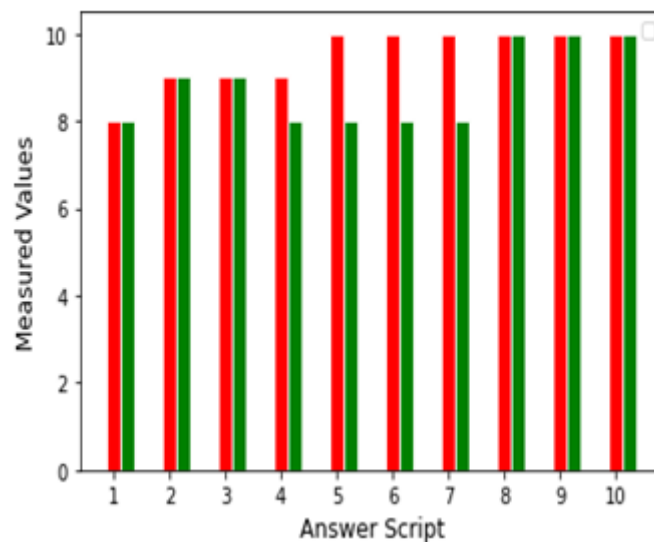

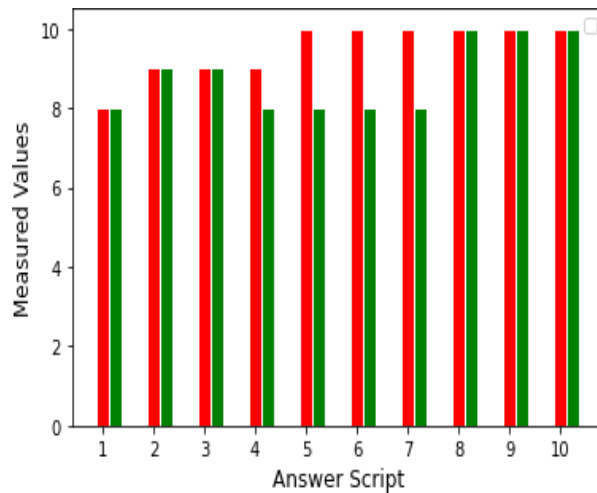Fig. 5. Graph Depicting Comparison for Number of Answers and Answers

Fig. 6. Bar Graph Depicting Comparison of Manual and System Allotted Marks.

The graph 6 is depicting bar graph which shows the evaluation comparison between manual evaluated by teachers and machine allocated marks. In this graph there are 20 bars, where 10 red bars represent manually given marks and another 10 green channel bars represents system allocated marks.

The figure 7 shows the evaluation result for 10 answer script.
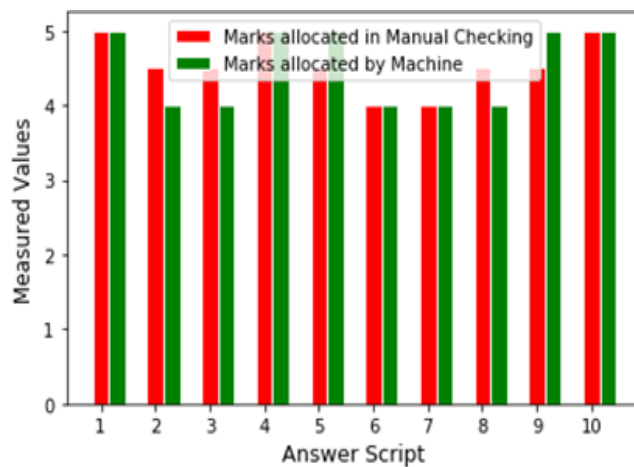
In the first column status of the answer is mentioned based on whether answer is written for the particular question or not. In the second column, marks obtained for written handwritten answer is allocated by system. Third column represents the time complexity of each script to process whole document and to give the output.

4.4 Assigning scores for manual checking and system generated

In this section, answer scripts of manually evaluated marks are stored and same answer scripts are passed to system to do evaluation and to assign marks. Results of manually

evaluated and machine generated is compared. which shows near to each other. So that, we conclude from table 4 evaluation done manually and system allocated represents close results for allocating marks for a particular answer.

Table 4. Comparison Between Manual Checking and Proposed System.

Answer Script	marks allocated in manually checking	Marks allocated by Machine
1	05	05
2	4.5	04
3	4.5	04
4	05	05
5	4.5	05
6	04	04
7	04	04
8	4.5	04
9	4.5	05
10	05	05

According to table 4, the results are compared. The marks are calculated for 10 answer script. In this table, second column evaluate the scores based on manual checking by teachers and third column represents the marks according to the machine. We can observe the difference between the teacher's evaluation and machine evaluation are very close. The teachers can give marks in fraction value for simple mistakes, but the machine is programmed to assign marks for exact value.

Fig. 7. Results of Answer scripts by system.

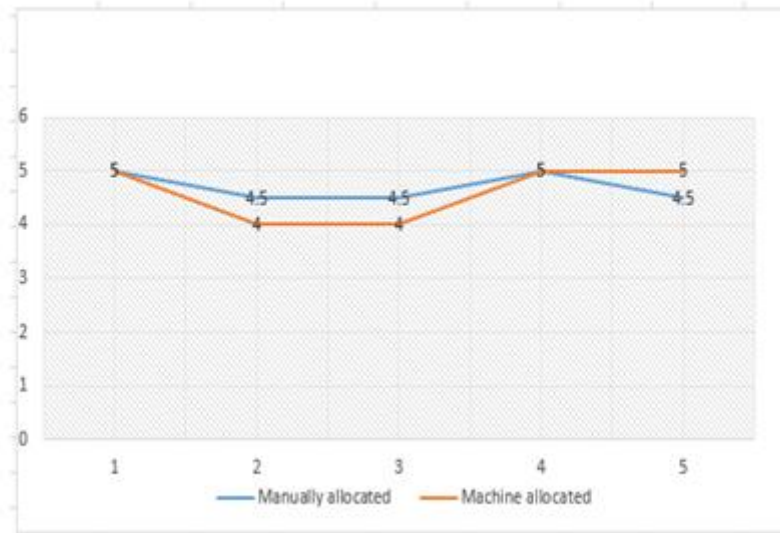| QUESTION NO	ANSWER STATUS	MARKS OBTAINED	TIME COMPLEXITY
1	NO	0	29.38 sec
2	YES	2	21.9 sec
3	NO	0	40.95 sec
4	YES	2	45.41 sec
5	NO	0	22.65 sec
6	YES	2	30.26 sec
7	YES	2	29.78 sec
8	YES	0	36.06 sec
9	NO	0	21.49 sec
10	YES	2	20.49 sec

Fig. 8. Comparison of Manual and System Allotted Marks.

The figure 8 shows the results of the answer scripts evaluated by teachers and output given by machine for same answer scripts. By observing, we conclude that by comparing both evaluation i.e. manually and system generated, the results are compromising and proposed system gives best result.

Fig. 9. Match The Following.

5.CONCLUSION

The goal of the study is to evaluate kannada one-word answer script automatically and assign marks for the written answers. we have developed a machine learning approach for

scoring marks. The accuracy gained for model is up to 90 percent. In future, the proposed system can be trained for descriptive answers as well as different languages and can be adopted in schools.

REFERENCES

1. Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande," S u b j e c t i v e Answer Evalua-tion System",Volume 7 Issue No.4 International Journal of Engineering Science and Computing, Pp.10457-10459, 2017.

2. Dharma Reddy Tetali, Dr. Kiran Kumar G, Lakshmi Ramana," Python Tool For Eval-uation Of Subjective Answers (APTESA)", International Journal of Mechanical Engineering and Technology (IJMET), Volume 8, Pp. 247-255, 2017.

3. Vimal P. Parmar, C K Kumbharana," Analysis of different examination patterns having question answer formulation, evaluation techniques and comparison of mcq type with one-word answer for automated online examination" international journal of scientific and research publications, volume 6, issue 3, pp.459- 463, 2016.

4. Meena,K, Lawrance Raj," Evaluation of the Descriptive type answers using Hyper- space Analog to Language and Self-organizing Map", EEE International Conference on Computational Intelligence and Computing Research,2014.

5. Sheeba Praveen," An Approach to Evaluate Subjective Questions for Online Ex-aminationSystem", International Journal of Innovative Research in Computer and Communication Engineering,Vol. 2, Pp. 6410-6413, 2014.

6. Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande," S u b j e c t i v e Answer Evalua-tion System",Volume 7 Issue No.4 International Journal of Engineering Science and Computing,Pp.10457-10459, 2017.

7. Asmitha dhokrat, gite hanumant r, c. namrata Mahender, Automated Answering for Subjective Examination", International Journal of Computer Applications.VOL.56. Pp.14-17, 2012.

8. Pranali Nikam, Mayuri Shinde, Rajashree Mahajan and Shashikala Kadam, Auto- matic Evaluationof Descriptive Answer Using Pattern Matching Algorithm", Interna- tional Journal of Computer Sciences and Engineering, Pp.69-70, 2015.

9. Nutchanat Sattayakawee," Test Scoring for Non-Optical Grid Answer Sheet Based on ProjectionProfile Method", International Journal of Information and Education Tech-nology, Vol.3, No. 2, Pp.273-277, 2013. http://www.rsc.org/dose/title of subordinate document. Cited 15Jan 1999.

10. Dhananjay Kulkarni, Ankit Thakur, Jitendra Kshirsagar, Y. Ravi Raju4, " Automatic OMR Answer Sheet Evaluation using Efficient Reliable OCR System", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 3, Pp.688-690, 2017.

11. Akhilesh Patil, Madhusudhan Naik, P.H. Ghare, "FPGA Implementation For OMR AnswerSheet Scanning Using State Machine And IR Sensors", International Journal of Electrical,Electronics and Data Communication, Volume-4,Issue-11, Pp.15-20, 2016.

12. Shweta M. Patil, Sonal Pati"," Evaluating Student Descriptive Answers Using Natu-ral LanguageProcessing", International Journal of Engineering Research Technology (IJERT). Vol.3 Issue 3, Pp. 1716-1718, 2014.

13. Ms. Paden Rinchen," Comparative Study of Techniques used for Automatic Evalua- tion of Free Text Answer", International Journal on Recent and Innovation Trends in Computing and Communication. Pp no 3839-3842, 2014.

14. Divya Patel1, Shaikhji Zaid2, "Efficient System For Evaluation Of Omr Sheet A Survey", international journal of advanced research in engineering, science and man-agement, Pp no01-08, 2017.

15. K. Ntirogiannis, B. Gatos and I. Pratikakis,"An Objective Evaluation Methodology for HandwrittenImage Document Binarization Techniques", Pp.217-224, 2008.

16. Vaishali Malpe, Shalini Bhatia, "Evaluation of Students' Answer Scripts Using Soft ComputingTechniques", International Journal of Modern Engineering Research (IJMER), Vol.2,Issue.3, pp-1280-1289, 2012.

17. Tilani Gunawardena, Nishara Pathirana, Medhavi Lokuhetti, Roshan Ragel, and Sam-path Deegalla, "Performance Evaluation Techniques for an Automatic Question An- swering System",International Journal of Machine Learning and Computing, Vol. 5, No. 4, Pp.294-300,2015.

18. Andrea Andrenucci, Eriks Sneiders, "Automated Question Answering: Review of the Main Approaches", Proceedings of the Third International Conference on Information Technology and Applications.Pp.514-519, 2005.