iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VI **Month of publication:** June 2018

DOI: http://doi.org/10.22214/ijraset.2018.6266

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VI, June 2018- Available at www.ijraset.com

Automated Students Answer Scripts Evaluation System using Advanced Machine Learning Techniques

Ms. Prakruthi S T[1], Dr. Hanuman Kumar V[2]
[1]*Assistant Professor, Dept. of ISE, New Horizon College of Engineering, Bangalore*
[2]*Senior Associate Professor, Dept. of CSE, New Horizon College of Engineering, Bangalore*

Abstract: ***Increased numbers of students in Higher Education and the corresponding increase in time spent by staff on assessment has encouraged interest into how technology can assist in this area. Over the years, after conducting examinations, there comes a crucial task of paper checking and evaluation which has to be done precisely, in time and even-handed. Although this process has been computerized for multiple choice questions; for long answers, it still has to be done manually. As students are assessed based on their performance in examination, It is really required to take very good care in the evaluation process. Traditional ways of evaluation of answer scripts has some major defects which should be resolved to ensure efficient assessment of students' answer scripts.***

To ensure this we propose a new model using advanced machine learning techniques and methodologies. The process of generating new model for evaluation involves supervised learning method i.e., scanning the answer scripts of students, then handwritten words are recognized using Tesseract OCR with Python, then match the handwritten words with the keywords in adaptive database using python based code, then award the marks as per scheme of evaluation provided by the course coordinator. An authorized signatory can also be provided once all students' marks are consolidated. Secured email facility shall be provided to send the assessed marks to the designated personnel.

Keywords: OpenCV(Open Computer Vision), OCR(Optical Character Recognition), Tesseract with Python, Supervised Learning.

I. INTRODUCTION

Machine learning provides computers the ability to learn without being explicitly programmed .machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs[1].

There are two most widely used machine learning methods i.e., supervised learning and unsupervised learning. Supervised learning is commonly used in applications where historical data predicts likely future events. Unsupervised learning is used against data that has no historical labels.

In the recent past, there has been a significant growth in the use of the ML techniques in the area of higher education functionalities such as evaluating student performances [2] etc. In current evaluation system, evaluator has scheme of evaluation which is referred to cross examine the students' answers. This evaluation system we use has major drawbacks. Some of them are listed below

A. Time delay that is involved in current evaluation system. Usually manual evaluation by the evaluator takes longer time for evaluation. As a result the students get their results delayed [3].
B. When is there is lack of time and interest, evaluator may not evaluate the answer scripts up to the mark.
C. There are also chances of having biasness among different students in the evaluation system.
D. Manual Evaluation is more prone for totalling error, wrong marking of mistakes, etc.
E. Lot of time and money is spent for evaluation process

This paper proposes an innovative paper evaluating technique with the assistance of IT which will help to reduce professor's work and manual resources involved. The Proposed scheme also helps in validating manually checked paper scores with the use of this

system, quick feedback can be given to students. Also, documents will be stored in electronic form [4], so they can be referred to at any point of time which will help in revaluation process.

II. LITERATURE REVIEW

Great deal of relevant work has been performed in the field of automatic as well as semi-automatic evaluation of answer scripts. Handwritten character recognition is a field of research in artificial intelligence, computer vision, and pattern recognition. The handwriting recognition is playing a critical part in evaluating system.

Batuhan Balci ,Dan Saadati, Dan Shiferawan[5] explained classification of individual handwritten word so that handwritten text can be translated to a digital form. Two main approaches to accomplish this task: classifying words directly and character segmentation. Convolutional Neural Network (CNN) with various architectures is used to train a model that can accurately classify words and Long Short Term Memory networks (LSTM) with convolution is used to construct bounding boxes for each character. Later segmented characters to a CNN are passed for classification, and then reconstruct each word according to the results of classification and segmentation.

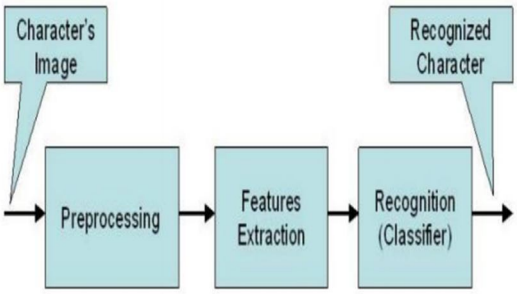


Fig1. Structure of character recognition in an image

Sameeksha Barve[6] explained how to generate optical character recognition system using Artificial Neural Networks(ANNs). The ANN is trained using the Back Propagation algorithm. Intelligent word recognition is employed in cursive handwriting conversion.

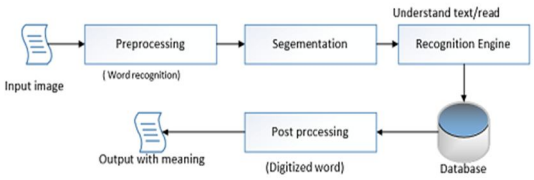


Fig. 2. Algorithm structure for character recognition

Shraddha Ghogare, Chetan Mahajan, Preeti Mulay[4] explained overview of an approach which allows users to interact with paper documents, books and answer sheets etc. and to evaluate performance based on keywords harmonized with a predefined adaptive database, by using electronic devices like Desktops, Laptops, Xerox machines, Mobile phones and Tablet Computers. The application would consist of query terms, error-grams, and correction rules to create search-able keywords to be matched with text processed from documents.

III. BASICS OF MACHINE LEARNING TECHNIQUES

A. Supervised Learning

Supervised learning algorithms are trained using labeled examples, such as an input where the desired output is known. For example, a piece of equipment could have data points labeled either “F” (failed) or “R” (runs). The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.

B. OpenCV(Open Computer Vision)

OpenCV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by Willow Garage then Itseez (which was later acquired by Intel). The library is cross-platform and free for use under the open-source BSD license. OpenCV supports the deep learning frameworks TensorFlow, Torch/PyTorch and Caffe

C. OCR(Optical Character Recognition)

OCR (Optical Character Recognition) also called Optical Character Reader is a system that provides a full alphanumeric recognition of printed or handwritten characters at electronic speed by simply scanning the form. More recently, the term Intelligent Character Recognition (ICR) has been used to describe the process of interpreting image data, in particular alphanumeric text. Function of OCR Forms containing characters images can be scanned through scanner and then recognition engine of the OCR system interpret the images and turn images of handwritten or printed characters into ASCII data (machine-readable characters). Therefore, OCR allows users to quickly automate data capture from forms, eliminate keystrokes to reduce data entry costs and still maintain the high level of accuracy required in forms processing applications. Features of OCR The technology provides a complete form processing and +documents capture solution. Usually, OCR uses a modular architecture that is open, scaleable and workflow controlled. It includes forms definition, scanning, image pre-processing, and recognition capabilities. What is ICR Intelligent Character Recognition (ICR) is the module of OCR that has the ability to turn images of hand written or printed characters into ASCII data. Sometimes OCR is known as ICR.

D. Tesseract with Python: pytesseract

It will recognize and read the text present in images. It can read all image types - png, jpeg, gif, tiff, bmp etc. It's widely used to process everything from scanned documents.Tesseract works best when there is a (very) clean segmentation of the foreground text from the background. In practice, it can be extremely challenging to guarantee these types of segmentations.

IV. PROPOSED SYSTEM

Main Objective of the proposed system is to evaluate the students' examinations answer scripts.

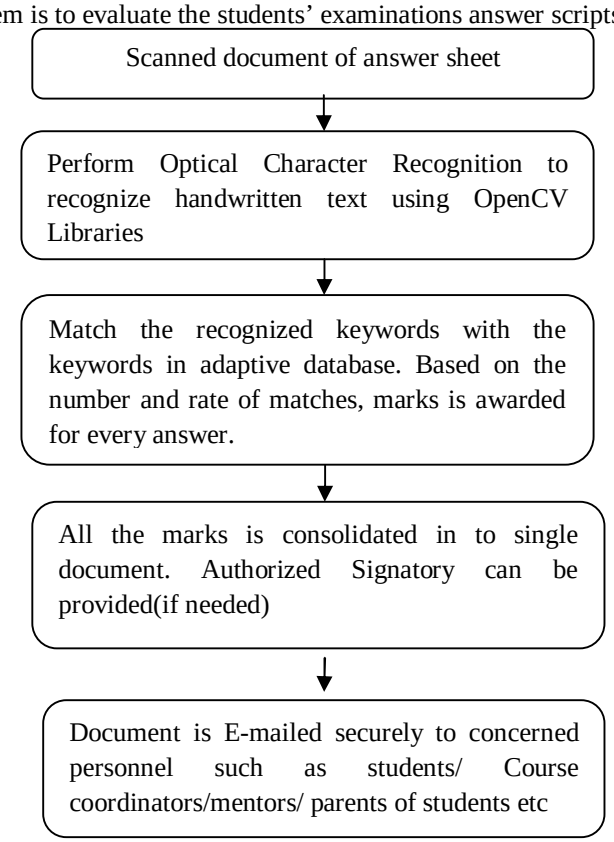


Fig 3: Detailed Design of Automated Students' Assessment using machine learning techniques

The Proposed System works as follows. Firstly Answer scripts are scanned using scanner. OpenCV is used for reading the scanned image. As a part of OCR, image is first converted to greyscale and then OpenCV applies two morphological operations on the converted greyscale image i.e., dilation and erosion to remove noise, Isolation of individual elements and joining disparate elements in an image and finding of intensity bumps or holes in an image. Then perform Image segmentation by using threshold function to get image in black and white background. Threshold function of OpenCV separates out the regions of an image corresponding to objects which we want to analyze. This separation is based on the variation of intensity between the object pixels and the background pixels. This thresholding is a critical step in Optical Character Recognition (OCR).

Tesseract, A Google's OCR Library is applied for precise character and word identification from the image. A programming language python is used in this process.

The keyword recognizer software then asks the user for evaluation scheme and pattern and fetches respective answers from the adaptive database. Then it finds a match between written keywords and database. Based on the number and rate of matches, marks is awarded for every answer.

Later, all students' marks are consolidated in to a document and emailed securely to concerned personnel such as students/ Course coordinators/mentors/ parents of students etc for further reference.

This proposed method overcomes major drawbacks of manual evaluation system As the system keeps on generating more electronic records, the storage need will increase subsequent and The evaluators need to be trained in order to make use of the system efficiently.

VI. CONCLUSION

This paper proposed a design of the system where in machine learning generated model can be utilized for automated evaluation of students' answer scripts. It also consolidates all students' marks and triggers email to concerned personnel such as students/ Course coordinators/mentors/ parents of students etc. The proposed method when applied, can result in time and cost savings. This can also help in decreasing manual reports involved. Both students and professors can be befitted as there is no human intervention in evaluation of answer scripts. The Proposed scheme also helps in validating manually checked paper scores.

VII. FUTURE WORK

A. Cloud computing can be used to mitigate storage and database requirements considering its various advantages both to end users and businesses of all sizes.

B. Incorporating synonyms of keywords dynamically and use of parallel algorithms to fasten the process can be done in future.

C. Also, cognitive computing can be employed to make computer program understand the concept instead of mere keywords

D. This automated evaluation system shall be enhanced with web interface so that the system can be accessed globally through any platform.

A Model for dual evaluation shall be designed for evaluating answer scripts twice and comparing the results. So that evaluation is better validated.

REFERENCES

[1] Pariwat Ongsulee," Artificial Intelligence, Machine Learning and Deep Learning", in 2017 Fifteenth International Conference on ICT and Knowledge Engineering, 978-1-5386-2117-2/17/$31.00 ©2017 IEE

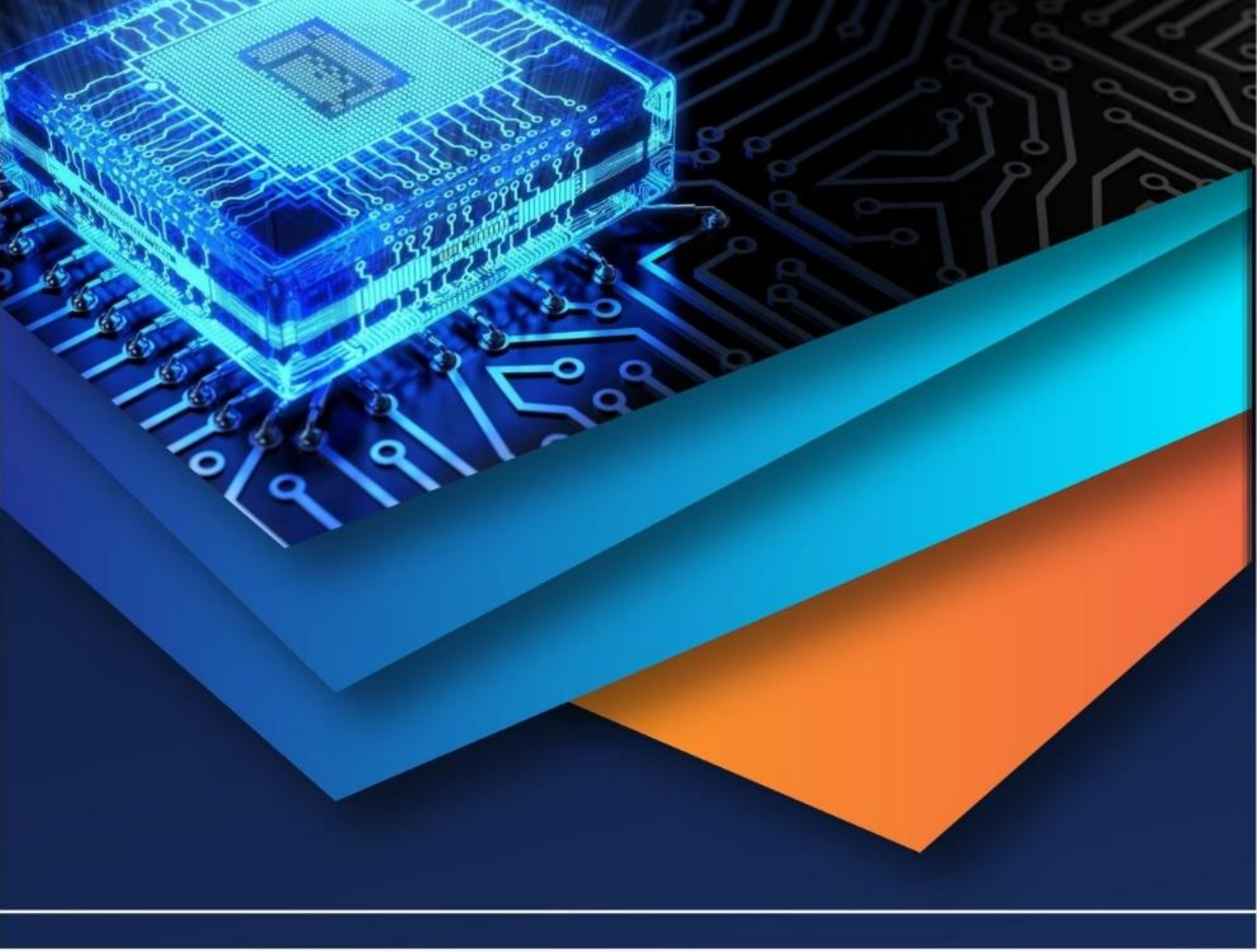
[2] V.O. Oladokun, A.T. Adebanjo, O.E. Charles-owaba, "Predicting student's academic performance using artificial neural network: A case study of an engineering course", In Pacific Journal of Science and Technology, vol. 9, no. 1, pp. 72-79, 2008

[3] anupriya Choudhury1, Kartikeya Jain2, Lakshya Aggarwal3, Ayushi Gupta4, Garv Saxena5, "Computerized Paper Evaluation Using Neural Network" in International Journal of IEEE, 978-1-5386-0514-1/17/$31.00 ©2017 IEE

[4] Shraddha Ghogare, Chetan Mahajan, Preeti Mulay, "Automation Related to Professor Evaluation" in International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 978-1-4673-9223-5/15/$31.00 c 2015 IEE

[5] Batuhan Balci ,Dan Saadati, Dan Shiferawan, " Handwritten Text Recognition using Deep Learning",http://cs231n.stanford.edu/reports/2017/pdfs/810.pd

[6] Sameeksha Barve, "Optical Character Recognition Using Artificial Neural Network" in International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 4, June 2012 pg-131-133



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY