IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Survey On Automated Answer Sheet Evaluation and Grading System

[1]Sejal Kaur Virdi, [2]Harshit Wadhwani, [3]Deepti Hegde, [4]Aditya Bhardwaj, [5]Sandhya N

Department of Artificial Intelligence and Machine Learning,
Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka

Abstract : In the dynamic landscape of education, the demand for innovation and efficiency has never been more critical. Our project endeavors to offer an effective solution to the time constraints and excessive efforts experienced by teachers and professors within the education system. The core objective is to liberate educators from the burdensome process of grading handwritten scripts, enabling them to channel their time and efforts toward refining pedagogical techniques and fostering a more engaging learning environment. Using techniques like image processing, OCR, NLP (embedding and similarity matching), and ML (Transformer) we can achieve automated evaluation and grading of handwritten scripts with great accuracy. The amalgamation of these technologies promises a transformative impact on the education sector, enhancing efficiency, objectivity, and consistency in the evaluation process.

***IndexTerms* - Machine Learning (Transformer), Natural Language Processing (NLP), Optical Character Recognition (OCR)**

I. INTRODUCTION

Correcting exam answer scripts is a demanding and time-consuming task, often presenting challenges in meeting deadlines [1]. Additionally, the accuracy of grading can be subject to scrutiny as it relies heavily on the evaluator's expertise, mood, and perspective. These factors contribute to inconsistencies and potential biases in the correction process [2].

The quarantine situation made educational institutions shift to online examinations, but these exams only consisted of one-word answers or MCQ which could be evaluated by the system. There exists no well-proposed system for subjective answer evaluations, therefore exams that need subjective answers to test students' conceptual knowledge do not exist. In major universities, answer scripts are scanned, and uploaded, and then assessors from different parts of the world assess them manually. This method is flawed due to several reasons like network bandwidth, loading time of answer scripts, constant stress to the eyes, human prone errors resulting in different grading of the same answer for different students [5].

II. BASIC CONCEPTS/ TECHNOLOGY USED

In the realm of answer script evaluation, diverse techniques play pivotal roles in comprehensively assessing student responses. The evaluation process unfolds through distinct phases, each leveraging sophisticated technologies:

OCR (Optical Character Recognition):

In the initial phase, both student answers and model answers undergo scanning, where Optical Character Recognition (OCR) tools are employed. This facilitates the extraction of textual content from scanned images, laying the foundation for subsequent analysis. The figure 1 below shows a basic demonstration of how text is extracted from handwritten answer scripts.

NLP (Natural Language Processing):

These multifaceted techniques collectively enhance the efficiency and accuracy of answer script evaluation. By seamlessly integrating OCR, NLP, machine learning, and similarity matching, educators can glean comprehensive insights into the quality and relevance of student responses. This synthesis of cutting-edge technologies underscores the evolution of answer script evaluation methodologies, contributing to the ongoing discourse in educational assessment practices.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Fig 1.2- Jaccard Similarity of text document A and B

Similarity Matching:

A crucial facet of the evaluation process involves measuring the similarity between student and model answers. Various similarity scores are employed, each contributing unique insights. Popular metrics include cosine similarity, Jaccard similarity, Pearson similarity, as well as assessments of synonym and semantic similarity as shown in Figure 1.1 and 1.2. Each metric yields distinct performance results depending on the nature of the answers being evaluated.

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Fig 1.1- Cosine Similarity of text document A and B

III. STUDY OF RELATED WORK:

DEMARCATION AND CONTEXT IDENTIFICATION:

Content identification is a multifaceted process crucial for extracting pertinent information from various data types, including text and images. Techniques such as YOLOv5 for text detection in natural scenes and a graph-based, unsupervised approach for precise context identification showcase the diverse methods employed. YOLOv5's two-stage approach excels in detecting and identifying language in scene images, while the unsupervised method utilizes a graph representation and community detection to discern context in unstructured text documents. These approaches offer effective solutions for tasks like information retrieval, document classification, and content filtering. Additional insights can be gained from resources like ScienceDirect Topics, IBM Watson, and Amazon Textract, which delve into the definitions, methods, and applications of content identification. Table 1.1 illustrates the comparative study.

Table 1.1- Comparison of Demarcation and Context Identification

Ref no.	Research Work/Paper	Author / Year	Techniques	Experiments/ Observations	Remarks
[14]	A Novel Text Recognition Scheme using Classification	M. Tamilselvi, G. Ramkumar (2022)	Classification, Digital Image Processing, Logical Text Classification Strategy, Text Recognition, Segmentation	Histogram equalization and line detection, smoothening filter.	Could not work well will blur images.
[15]	Text Detection and Language Identification in Natural Scene	R. S. Latha (2023)	Deep learning, Measurement, Text recognition, Neural networks, Object detection, Object recognition	This paper uses “you only look once” yolo method and offers prediction with just one single forward propagation.	No remarkable drawbacks found.
[16]	An Unsupervised Approach for Precise Context Identification	M. Mallek, S. Fournier (2020)	keyword frequencies, LDA, text mining, Jiang and Conrath similarity, TFIDF, Text Rank, averaged F1-score, precision.	Introduces a context identification approach of processing unstructured web documents without specialized algorithms	Improvements needed in documents of multiple context and extending approach to other languages as well.

[17]	Identification of image source using serial-number-based watermarking	Andjela Draganić, Milan Marić (2017)	Public Key Cryptography Signature (PKCS), Digital Watermarking, Compressive Sensing (CS), SN Extraction, Digital Signature	Introduces a watermarking technique using a 32-bit Public Key Cryptography Signature (PKCS) serial number, embedded into images to facilitate camera identification	While showing promising results, potential improvements and ongoing efforts

OPTICAL CHARACTER RECOGNITION:

Optical character recognition (OCR) is a technology that converts scanned images of text into editable and searchable digital text. OCR can be used for various purposes, such as digitizing books, documents, invoices, receipts, and handwritten notes. OCR can also be applied to natural scene images, such as street signs, billboards, and license plates. OCR involves several steps, such as image preprocessing, segmentation, feature extraction, classification, and post-processing.
The existing methods in this domain are related to the challenges and methods of OCR, such as cursive text recognition in natural scene images using deep convolutional recurrent neural network, handwritten OCR.
A comprehensive systematic literature review (SLR), and OCR for text recognition and its post-processing method as shown in table 1.2 demonstrates the potential and challenges of using OCR for various applications and languages, and provide insights and directions for future research in this domain. Figure 1.3 demonstrates the accuracy values for models

Table 1.2- Optical Character Recognition

[11]	Cursive Text Recognition in Natural Scene Images	A A. Chandio (2022)	preprocessing, segmentation, optical character recognition, cursive handwriting, hidden Markov model, search, graph, lexicon matching.	In this paper the handwriting analysis is done primarily by segmentation bade of height and width of characters.	Inefficient of illegible or unclear writing and writing of varying sizes and spelling errors
[12]	Handwritten Optical Character Recognition	J. Memon, M. Sami (2020)	extraction of features and discrimination/classification of these features (based on patterns)	Adaption of cluster computing and GPUs and better performance by deep learning architectures, which including RNN, CNN and LSTM.	No specific weak areas found
[13]	TEXT OCR AND POST-OCR CORRECTION	R. Avyodri, S. Lukas (2022)	Image acquisition, preprocessing, segmentation, feature extraction, classification, post processing	Lossy filter is applied for graph representation of characters and histogram equalization is done. Thinning and segmentation.	Could not manage skewed lines and deviates from distinctive orientation.

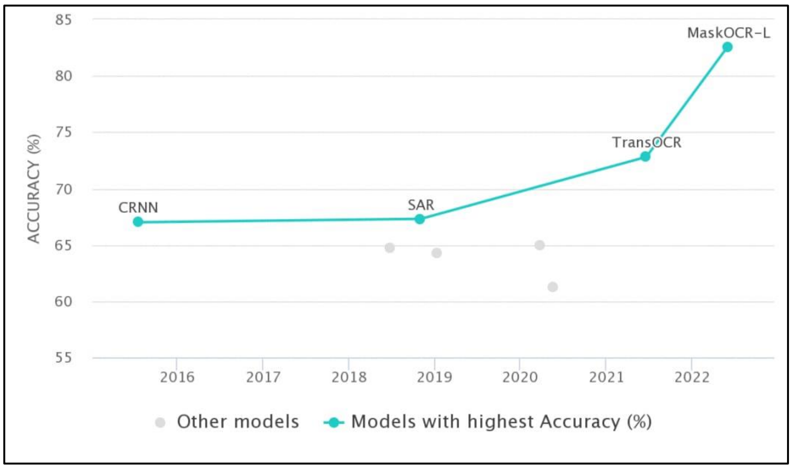

Fig 1.3- Comparison measure of accuracy values for various OCR models

NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING:

Natural Language Processing (NLP) is employed in tandem with OCR to preprocess the extracted text comprehensively. Various techniques, including stop word removal, lemmatization, tokenization, and stemming, contribute to refining the textual data for meaningful analysis. Figure 1.4 depicts the overall process of NLP as followed in various reference papers.

NLP, which is a subfield of artificial intelligence that uses computational techniques to process and analyze natural language data, such as text and speech. Descriptive answers are answers that require the student to explain, analyze, or justify a topic in their own words. The papers propose different methods and models for automated evaluation of descriptive answers, using techniques such as feature extraction, text similarity, memory network, WordNet graphs, image processing, optical character recognition, text analysis, and fuzzy logic.
In this manner, several studies have shown to give the results as illustrated in table 1.3

Table 1.3- Comparison of NLP and ML

[1]	NLP-based Automatic Answer Script Evaluation	Rahman and Siddiqui (2018)	NLP techniques are used for summarization and extraction. The scores are awarded based on similarity measures computed using cosine, Jaccard, bigram, and synonyms.	Automatic evaluation of answer scripts was found very useful and often the assigned marks were the same as manually scored marks	The parameter weights used for grading are not computed automatically; they are assigned manually based on a survey. This is the main drawback.
[2]	An Intelligent System for Evaluation of Descriptive Answers	Bagaria et al. (2020)	Uses concept graphs, fuzzy string matching, grammar checking and other similarity metrics	The difference between the marks given by manual evaluation and proposed automatic evaluation was less for conceptual and specific type questions and greater for analytical questions.	Answers containing non-textual content like equations, diagrams and tables cannot be scored through this work.
[3]	A machine learning approach for automated evaluation of short answers using text similarity based on WordNet graphs.	Vij et al. (2020)	Uses WordNet graphs for structural and semantic dependencies.	RMSE was found to be 0.319	Fails in evaluating answer scripts with technical words and misspelled words.

[4]	Attention Is All You Need In Speech Separation	Cem Subakan; Mirco Ravanelli (2021)	Layered framework with a Multiattention layer that captures the relation between student and reference answers	Outperforms baselines like Logistic regression, Gradient boosted decision tree, TextCNN, Bitransformer, LSTM and Manhattan LSTM	No specific weak areas found
[5]	Deep Automated Text Scoring Model Based on Memory Network	Yang (2020)	Word weights to emphasize important words/phrases and KB comprising of standard answer are used to grade the student answers using deep learning LSTM networks.	The overall accuracy rate of this system is 83.37%	tf-idf may not be adequate for determining concentration. Single basic concept questions worked better than flow description questions
[6]	Automatic Evaluation of Descriptive Answers Using NLP and Machine Learning.	Prof. Sumedha P Rau (2022)	word choice, and various similarity measures, leveraging Natural Language Processing (NLP) for efficient evaluation. NLP also streamlines text summarization techniques, employing extractive and abstractive methods to condense lengthy articles,	Outperforms traditional text mapping algorithms	Abstractive techniques might introduce inaccuracies or change the original meaning, affecting the quality of the summary.
[7]	Automatic Subjective Answer Evaluation	Vijay Kumari, (2023)	Keyword matching, similarity matching, grammar check	The total score is the sum of the similarity, grammar/ language, and keyword scores	Requires Human graded answer sheets.
[8]	Evaluation of Handwritten Answer Scripts Using Machine Learning Approaches	Ravikumar.M (2023)	Handwritten answer, Segmentation, Recognition, Evaluation	The accuracy gained for model is up to 90 percent	Tackles only one-word answers
[9]	Answer script evaluator	Sachin B.S (2023)	The system performs extraction of keywords and this extracted text is then compared to the reference text using Latent similarity indexing.	Have a tolerance factor where in the comparison percentages can be rounded off to the next whole number or the next number which is divisible by 10. (For eg, if comparison percentage for a sentence is 73%, we around it off to 80 %.	Incapable of sophisticated grading. Heavily reliant on rounding off.
[10]	Towards Automated Evaluation of Handwritten Assessments	Vijay Rowtula (2019)	handwritten answers; self-supervised learning; deep learning	Scoring model based on a word spotting. Here, matching score between the keywords associated with the Textual Reference Answer (TRA) and Handwritten (HW) document are found.	Sensitive to input noise.
[18]	Influence of Word Normalization and Chi-Squared Feature	Haryanto, A.W., Mawardi, E.K (2018)	SVM, Stemming, Chi-square, Preprocessing, Feature Reduction	SVM with stemming and Chi-square outperforms lemmatization, achieving 95% precision, recall, and	exploring advanced feature reduction methods and optimizing lemmatization for enhanced text classification .

	Selection on (SVM)			95.05% accuracy with 80% feature reduction.	
[19]	Information extraction using natural language processing	Shariff, Umar (2022)	(NLP), Text Reduction, Unstructured Data Processing, Information Extraction, Clean Data Extraction.	Includes burgeoning volume of unstructured data, emphasizing sophisticated Information Extraction using NLP	Enhancing efficiency in text reduction and information extraction for growing volume of unstructured data remains a critical area of improvement.
[20]	Automated Evaluation of Student Answer Paper using ML Techniques.	Syeda Sameen Fatima, Ganga Sanuvala (2021)	OCR tool, Cosine Similarity, Logistic Regression, Naïve Bayes, Support Vector Machine.	This paper used different ML methods to find which one gives the best accuracy.	SVM proved to give good accuracy but lacked preprocessing.

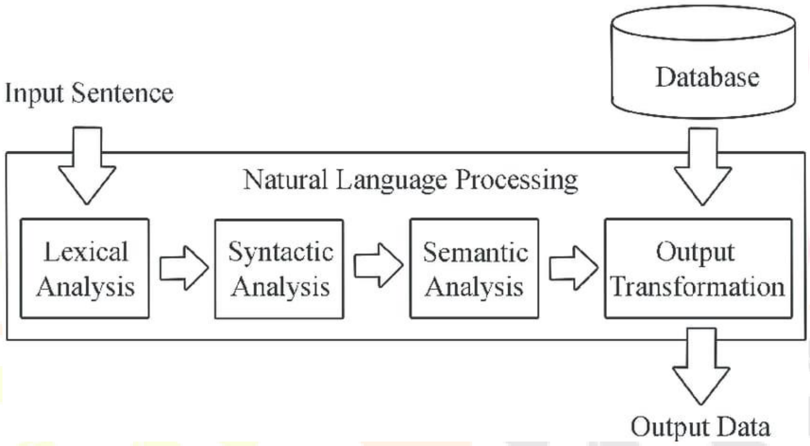

Fig 1.4- The general process of NLP

IV. CHALLENGES IN EXISTING SYSTEM

There are various challenges as understood by carrying out different surveys:

Handwriting Variability:
Handwriting styles can vary significantly among students, making it challenging to develop a accurate OCR models. Handling diverse handwriting requires robust techniques for feature extraction and pattern recognition.

Subjectivity in Grading:
The inherent subjectivity in grading, especially for open-ended questions, poses a challenge. Different educators may have varying perspectives on the quality of answers, making it difficult to establish a universally applicable grading standard.

Ambiguity in Responses:
Addressing ambiguity in student responses is a complex task. Open-ended questions may have multiple correct interpretations, and designing algorithms to capture the nuanced meaning of responses requires sophisticated natural language processing techniques.

Best Similarity matching:
Different metrics prove to be effective for different kinds of answers. Therefore, to find a suitable similarity degree among documents and selecting right features of the documents for similarity measuring is a challenge in itself.

Adapting to various question formats:
Adapting the system to evaluate different question formats, including essays, mathematical equations, diagrams, and code snippets, presents a challenge. Developing algorithms capable of comprehensively assessing diverse types of responses is crucial.

Answer demarcation:
To understand where one answer begins and where the next one starts is the main challenge where one needs to come up with different demarcation technique to understand the extent of an answer.

V. CONCLUSION

In conclusion, the literature survey illuminates the dynamic landscape of "Automated Answer Script Evaluation and Grading," showcasing the diverse methodologies and technologies that have transformed traditional assessment practices. The fusion of Optical Character Recognition (OCR), Natural Language Processing (NLP), and Machine Learning (ML) has paved the way for a paradigm shift in how educators approach the evaluation of student responses.

The integration of OCR ensures efficient text extraction from scanned answer scripts, while NLP techniques, including stop word removal, lemmatization, and tokenization, refine the extracted text for meaningful analysis. Machine Learning emerges as a powerful tool, encompassing supervised and unsupervised models, deep learning networks, and attention mechanisms, providing a robust framework for assigning weights to words and determining similarity scores.

The exploration of various similarity matching metrics, such as cosine similarity, Jaccard similarity, and semantic similarity, further enriches the evaluation process, offering educators nuanced insights into the quality and relevance of student answers. As the literature survey unfolds, it becomes evident that automated evaluation systems not only expedite the grading process but also enhance its objectivity, consistency, and adaptability to diverse

question formats.In the era of automated answer script evaluation, educators are empowered to focus on refining teaching methodologies, fostering an environment of continuous improvement. The synthesis of these technologies represents a transformative leap towards an education system that is not only efficient but also capable of providing insightful, personalized feedback to students.

As we delve deeper into this evolving field, it is imperative to acknowledge the challenges, such as handwriting variability, subjectivity in grading, and scalability issues, and work collectively to address them. Through this literature survey, we navigate the complex landscape of automated evaluation, recognizing its potential to redefine educational assessment, drive efficiency, and ultimately contribute to the continuous enhancement of teaching and learning practices. The journey towards automated answer script evaluation is an ongoing one, promising a future where educators and students alike benefit from a more streamlined, objective, and impactful grading process.

VI. REFERENCES

[1] Rahman, Md & Siddiqui, Fazlul. (2018). NLP-based Automatic Answer Script Evaluation. 4. 35-42.
[2] Bagaria, Vinal & Badve, Mohit & Beldar, Manasi & Ghane, Sunil. (2020). An IntelligentSystem for Evaluation of Descriptive Answers. 19-24.10.1109/ICISS49785.2020.9316110.
[3] Vij, Sonakshi & Tayal, Devendra & Jain, Amita. (2020). A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on WordNet Graphs. Wireless Personal Communications. 111. 10.1007/s11277-019-06913-x.
[4] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi and J. Zhong, "Attention Is All You Need In Speech Separation," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 21-25, doi: 10.1109/ICASSP39728.2021.9413901.
[5] S. Yang, "Deep Automated Text Scoring Model Based on Memory Network," 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL).
[6] Prof. Sumedha P Raut, Siddhesh D Chaudhari, Varun B Waghole, Pruthviraj U Jadhav, "Automatic Evaluation of Descriptive Answers Using NLP and Machine
Learning" in IJARSCT vol 2, issue 1,March 2022.
[7] Kumari V., Godbole P. and Sharma Y. (2023). Automatic Subjective Answer Evaluation. In Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods – Volume 1: ICPRAM, ISBN 978-989-758-626-2, pages 289-295. DOI: 10.5220/0011656000003411
[8] Ravikumar.M, Sampath Kumar.S and Shivakumar.G (2023) "Evaluation of Handwritten Answer Scripts Using Machine Learning Approaches", Turkish Journal of Computer and Mathematics Education (TURCOMAT), 9(1), pp. 620–633. doi: 10.17762/turcomat. v9i1.14041.
[9] Journal, Ijariit. "Answer Script Evaluator." International Journal of Advance Research, Ideas and Innovations in Technology, 2019.
[10] V. Rowtula, S. R. Oota and J. C.V, "Towards Automated Evaluation of Handwritten Assessments," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 2019, pp. 426-433, doi: 10.1109/ICDAR.2019.00075.
[11] A. A. Chandio, M. Asikuzzaman, M. R. Pickering and M. Leghari, "Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network," in IEEE Access, vol. 10, pp. 10062-10078, 2022, doi: 10.1109/ACCESS.2022.3144844.
[12] J. Memon, M. Sami, R. A. Khan and M. Uddin, " Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," in IEEE Access, vol. 8, pp. 142642-142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
[13] R. Avyodri, S. Lukas and H. Tjahyadi, "Optical Character Recognition (OCR) for Text Recognition and its Post-Processing Method: A Literature Review," 2022 1st International Conference on Technology Innovation and Its Applications (ICTIIA), Tangerang, Indonesia, 2022, pp. 1-6, doi: 10.1109/ICTIIA54654.2022.9935961.

[14] M. Tamilselvi, G. Ramkumar, G. Anitha, P. Nirmala and S. Ramesh, "A Novel Text Recognition Scheme using Classification Assisted Digital Image Processing Strategy," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752542.

[15] R. S. Latha et al., "Text Detection and Language Identification in Natural Scene Images using YOLOv5," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-7, doi: 10.1109/ICCCI56745.2023.101284
[16] M. Mallek, S. Fournier, "An Unsupervised Approach for Precise Context Identification from Unstructured Text Documents," 2020 IEEE pp. 821-826, doi: 10.1109/ICTAI50040.2020.001.
[17] A. Draganić, M. Marić, I. Orović and S. Stanković, "Identification of image source using serial-number-based watermarking under Compressive Sensing conditions," 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2017,1227-1232, doi: 10.23919/MIPRO.2017.7973611.
[18] Haryanto, A.W. and Mawardi, E.K., 2018, September. Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification. In 2018 International Seminar on Application for Technology of Information and Communication (pp. 229-233). IEEE.
[19] Shariff, Umar & . N, Srikanth & Reji, Varun. (2022). Information extraction using natural language processing. International journal of scientific research in engineering and management. 06. 10.55041/ijsrem13271.
[20] G. Sanuvala and s. S. Fatima, "a study of automated evaluation of student's examination paper using machine learning techniques," 2021 international conference on computing, communication, and intelligent systems (icccis), greater Noida, India, , 2021.