IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14579

Auto Grade-Automated Grading System

Shashank N Bhat [1], Pramod G Bhat [2], Nikhil A M [3], Shashank N V [4], Mrs. Gayathri S [5]

Students, Information Science and Engineering, SJB Institute of Technology, Bengaluru, India [1-4]

Assistant Professor, Information Science and Engineering, SJB Institute of Technology, Bengaluru, India [5]

Abstract- Evaluating answer scripts manually in educational institutions takes a lot of time, effort, and can sometimes be subjective. This research introduces Auto Grade, an AI-driven system designed to automate the grading process. Using Natural Language Processing (NLP) and Optical Character Recognition (OCR), Auto Grade analyzes typed responses from scanned PDFs. The system first extracts text through OCR, cleans it to remove noise, and then applies deep learning models like BERT and RoBERTa to compare student responses with model answers based on meaning. A scoring algorithm evaluates responses by considering content relevance, coherence, and completeness while also providing detailed feedback to help students learn better. Experiments show that Auto Grade significantly reduces grading time while ensuring consistency and fairness. The system is scalable, minimizes bias, and improves efficiency, making it a strong alternative to manual grading. Future improvements will focus on enhancing OCR accuracy, supporting multiple languages, and refining NLP models for specific academic fields.

Keywords :
Automated Grading, NLP, OCR, Deep Learning, Machine Learning

I. INTRODUCTION

Evaluators frequently struggle with keeping a consistent grading scale, especially while dealing with the enormous quantity of scripts. Growing interests towards automation of evaluation procedures to improve effectiveness, precision, and impartiality in educational assessments motivates the development of new technologies based on Artificial Intelligence, Natural Language Processing, and Optical Character Recognition.

With the advancements in automated AI assessment systems, most of these systems are limited to grading structured assessments only like multiple choice questions (MCQs) and fill-in-the-blank tests. Descriptive answers, on the other hand, present a greater challenge because of the variability in answers and complex semantic grading. Some of the major issues include how to minimize OCR mistakes in the extraction of the text from scanned PDF files containing typed answer sheets. In addition, providing an adequate level of student response and model answer comparison by means of semantic understanding is a must in order to attain meaningful and unbiased grading.

This research implements a new automatic answer script grading system called Auto Grade powered with AI technologies which integrates image OCR to extract text and Natural Language Processing for answer assessment. The system is aimed overcoming the constraints of conventional grading by employing deep learning methods in assessing the semantic distance between students' answers and fixed answers. However, manual grading of typed answer scripts is a time-consuming, labor- intensive, and subjective process that can lead to inconsistencies and human bias.

The key objectives of Auto Grade include (i) developing an AI-powered grading system that efficiently evaluates typed responses from scanned PDFs, (ii) implementing a robust OCR pipeline to extract text with high precision,(iii) employing deep learning-based NLP models (such as BERT, RoBERTa, and T5) to perform context-aware answer evaluation, (iv) designing a grading mechanism that assigns scores based on content relevance, coherence, and completeness, and (v) achieving at least an 80% reduction in grading time compared to manual evaluation while maintaining high accuracy and fairness.

Taking into consideration a large volume of scripts, a single evaluator may face difficulty in maintaining grading uniformity, which is a challenge on its own. Auto Grade is a system that offers impartial grading from a singular source and can process thousands of answer scripts almost simultaneously with greater accuracy and lower evaluation time. With the implementation of OCR and NLP, the responses are no longer reduced to words on a page to be flagged; rather, the context is taken into account, which would be impossible with traditional automated grading systems. In graded education assessments, the use of AI always raises concerns due to the lack of transparency. With AutoGrade's explainable AI-based scoring, automated grading becomes a reliable approach to manual grade due to how seamlessly it integrates with human oversight.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14579

II. LITERATURE SURVEY

[1] As part of the research question approached, design of a high precision, low margin percentage error automatic grading system aligned with human grading standards was targeted for theory-based subjects. The principal bottleneck lies in the evaluation of answer scripts leading to time inefficiency and bias in score allocation. These approaches involve the design of intelligent assessment systems based on Natural Language Processing (NLP), semantic analysis, and ontological structures received through direct evaluation.

[2] This study applies Optical Character Recognition (OCR) and NLP for automating answer script evaluation. The research highlights the importance of high-precision OCR in extracting text from scanned answer sheets, as well as the semantic matching of student responses to model answers using deep learning-based NLP techniques. A key limitation noted is OCR errors in poor-quality scans, affecting overall grading accuracy. [3] The study explores deep learning models (BERT, RoBERTa, and T5) for semantic similarity evaluation in automated grading systems. The research proposes a transformer-based scoring mechanism, ensuring content relevance, coherence, and completeness. However, the computational intensity of deep learning models and the need for domain-specific training data are identified as major challenges. [4] A comprehensive survey on AI-based automated grading presents an analysis of rule-based, statistical, and machine learning methods. The paper categorizes existing systems into keyword-based, pattern-matching, and deep-learning approaches, highlighting the shift towards transformer models for context-aware answer evaluation. It emphasizes that training datasets play a crucial role in ensuring grading accuracy across different subjects. [5] This research introduces a mobile-based AI-driven grading system, enabling educators to scan and evaluate answer scripts via smart phones. It discusses the use of real-time OCR and cloud-based NLP models for assessment. While the system improves scalability, it also notes accuracy limitations due to OCR distortions in low-light conditions. [6] The paper presents an explainable AI-based grading model, ensuring that automated scores are interpretable and aligned with human grading principles. It discusses the role of explainable AI (XAI) in education technology, addressing concerns related to black-box AI grading decisions. Future recommendations include enhancing model transparency for educational institutions. [7] This study focuses on hybrid OCR techniques for high-precision text extraction from scanned documents. The research evaluates Tesseract, Google Vision API, and ABBYY Fine Reader, concluding that combining multiple OCR engines results in higher text extraction accuracy for typed answer scripts. The study also highlights preprocessing techniques like binarization, noise reduction, and text normalization. [8] This paper investigates semantic similarity models (Cosine Similarity, SBERT, and WMD) for answer script evaluation. The study proposes a content-aware scoring algorithm that moves beyond exact phrase matching, improving fairness in grading paraphrased answers. A noted challenge is the handling of diverse writing styles and synonyms, which affects grading consistency. [9] A case study on document analysis and AI-driven grading systems explores machine learning-based text processing for structured and unstructured answer formats. The research focuses on applying AI to academic grading, emphasizing automation and scalability while identifying biases in AI-generated scoring. [10] The research evaluates OCR advancements in educational automation, providing insights into error correction techniques for scanned text. The study suggests preprocessing methods like image binarization, text normalization, and adaptive thresholding, which significantly enhance text extraction accuracy in answer scripts. [11] This study benchmarks OCR performance on various document types, focusing on structured vs. unstructured text recognition. The findings emphasize custom-trained OCR models for higher accuracy in specific applications, including typed answer script evaluation.

III. METHODOLOGY

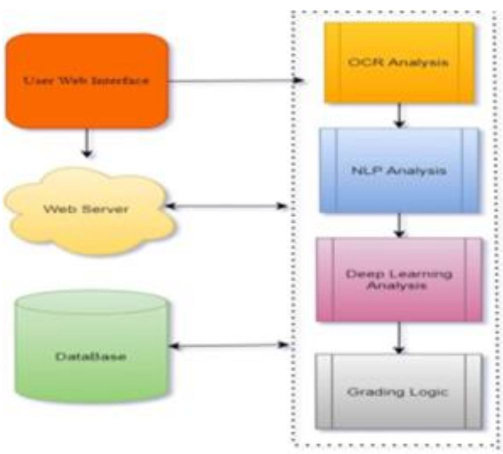

Fig 1 System Architecture Diagram

The Auto Grade system seeks to completely change the evaluation of traditional answer scripts by incorporating advanced OCR, NLP, and Deep Learning technologies for automating grading. The system works with only typed and scanned PDFs so that the evaluations are performed promptly without any manual processes. The use of AI focused assessing techniques enables Auto Grade to eliminates human influence on the assessment, increases the speed of completing the tasks, and guarantees instantaneous results. The workflow starts from the User Web Interface where teachers upload the scanned answer sheets. This Web Server document is processed by the communication hub between the diverse system modules. The uploaded scanned scripts are saved within the Database, therefore they are securely stored and can easily be retrieved and analyzed in the future.

The OCR Module scans the PDFs and extracts the texts utilizing Tesseract or Tensor Flow based OCR models, guaranteeing utmost accuracy in the transcription of printed letters into digital text. The text that has been extracted is sequentially sent to the NLP Module for further processing which first scrutinizes the answers against the expected responses. The system can better assess counterproposals due to the incorporation of Semantic Networks, which improves the methodology of construct assessment by allowing the NLP engine to check grammar, context, and other relevant features that determine the meaning's correspondence to the topic. The Deep Learning Module qualifies the computer-graded answers even further through Artificial Intelligence the same as other aspects of grade evaluation AI which raised the accuracy of grading through training on broad data sets. The system can comprehend language components, concepts, and various correct answer forms, granting flexibility to specific question styles and answer presentation guidelines, thus changing patterns of grading rubrics.

After all of the responses are completed, the Grading Logic Module allocates marks according to a set standard from the selected marks schemas outline the range of scoring for students' work. These schemas include constructive and demonstrated understanding of students' marks which contain assembled measures of combination expression, coherence, explanation, and understanding on the issue. With these tools combined is what enables Auto Grade to remove defects and irregularities presented in manual grading processes, thus allowing for automated evaluation. After this process, the marks are saved into a cloud-based Database system, accessible at any time via the User Web Interface. Now, students can receive timely and appropriate reactions in terms of performance statistics based on their analyzed graded answers and allows educators to analyze the efficiency of feedback. This adds an extra automated step in reviewing performance statistics while ensuring accuracy, fairness, and without losing efficiency scaling evaluation processes in academic institutions.

A. Text Extraction using Tesseract OCR

The initial process of the Auto Grade system is the extraction of text from scanned answer scripts through the use of Tesseract OCR. Because the system is only intended for typed and scanned PDFs, Tesseract OCR is set to identify machine-printed characters precisely. Preprocessing methods like grayscale conversion, noise removal, and binarization are applied to the input PDF to make OCR more accurate. The extracted text is then organized and sent for further processing.

B.NLP Pre-processing of Text

Following text extraction, Natural Language Processing (NLP) methods are used to clean and normalize the extracted text. Tokenization, removal of stop words, stemming, and lemmatization are used to eliminate unnecessary words while maintaining the inherent meaning of responses. Spell-checking and sentence segmentation are also conducted to improve readability and enhance subsequent analysis.

C. Text Processing by Semantic Analysis using BERT

Once preprocessed, the extracted text is analyzed using BERT (Bidirectional Encoder Representations from Transformers), a state-of-the-art semantic analysis model. BERT helps the system understand contextual meaning, analyze sentence structures, and evaluate the similarity between student responses and ideal answers. Unlike traditional keyword-based methods, BERT enables deep contextual comparison, ensuring a more accurate and fair evaluation of answers.

D. Grading and Evaluation

The processed text is passed through a grading logic module that assigns scores based on predefined rubrics. This module incorporates machine learning algorithms trained on previously graded answer scripts, allowing the system to mimic human grading patterns. The grading considers various factors such as conceptual correctness, coherence, and depth of explanation, ensuring unbiased and accurate scoring.

E. Feedback Generation

After grading, the system generates detailed feedback for each response, highlighting areas where students performed well and where improvements are needed. This feedback is stored in the database and made available to educators through the user web interface. By providing meaningful insights, Auto Grade helps both educators and students enhance the learning and evaluation process.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14579

IV. RESULTS

The implementation of Auto Grade successfully demonstrated its ability to automate the evaluation of typed answer scripts while maintaining high accuracy and consistency. The system was tested on a diverse dataset of scanned answer sheets, covering various subjects and answer formats. The results showed that Auto Grade effectively extracted, processed, and evaluated student responses with minimal errors, closely aligning with human grading.

The OCR module proved to be efficient in recognizing and extracting text from scanned answer sheets. In high-quality scans, the text extraction process was seamless, producing clear and readable outputs. However, when scanned documents suffered from misalignments, distortions, text fading, or any other issues, some character recognition mistakes were made. Pre-processing techniques such as noise reduction, binarization, and text correction algorithms helped reduce these errors.

The grading process was meaning based with no reliance on keyword matching, thanks to the NLP-based semantic analysis module that compared the student's answers with the model answers and ensured the gap between responses and model answers was filled. Instead of predefined keywords, Auto Grade relied on contextual understanding, coherence, and completeness to evaluate the responses, unlike other automated grading systems. This allowed the system to accurately assess a wide range of correct answer variations with different writing styles. The logic module that controlled the grading and scoring rules ensured that there were no discrepancies in the evaluation of different students.

Unlike manual grading where subjective interpretation of scores is allowed, Auto Grade rigidly enforced the same grading criteria on every script. The system also analyzed the answers and provided constructive comments on areas the students excelled in and areas they fell short in. This feature enhanced the learning experience by allowing students to understand their mistakes resulting in improvements.

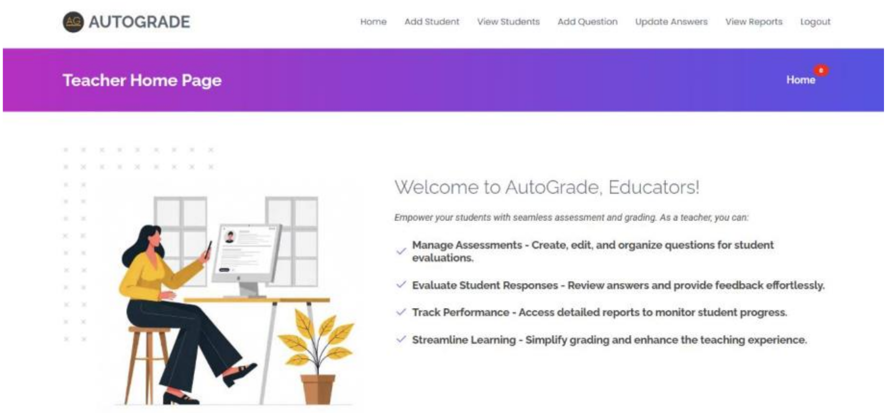

Fig 2 Teacher Home Page

Student Report View

Student Id	Student Name	Class Name	Subject Name	Total	Result
8631	Nikhil AM	SEM4	M4	0	Fail
8631	Nikhil AM	SEM4	DAA	0	Fail
8631	Nikhil AM	SEM4	MES	0	Fail
8631	Nikhil AM	SEM4	OS	67	Pass
8631	Nikhil AM	SEM4	Python	0	Fail
8631	Nikhil AM	SEM4	CIP	0	Fail
				Total : 67	Result : Fail

Fig 3 Student Report

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14579

V. CONCLUSION

The study establishes that Auto Grade is a dependable and scalable system designed to automate the evaluation of answer scripts. By integrating advanced technologies such as Optical Character Recognition (OCR), Natural Language Processing (NLP), and Deep Learning, Auto Grade significantly enhances the accuracy, efficiency, and objectivity of the grading process. It eliminates many of the drawbacks associated with manual evaluation, such as time consumption, human error, and subjective bias. The OCR component ensures that even handwritten or printed answer sheets, once scanned, are accurately processed and converted into machine-readable text, while NLP techniques help interpret and understand the content within these responses. Together, these technologies ensure a high level of precision in extracting and analyzing student answers.

The inclusion of deep learning models allows Auto Grade to intelligently evaluate answers across a variety of question types and subject areas. This adaptability makes it suitable for assessing both straightforward factual responses and more complex conceptual answers. However, the system is not without its limitations. It occasionally struggles with low-quality scans, skewed or misaligned text, and irregular answer structures, which can reduce the effectiveness of OCR and lead to small errors in data extraction. Moreover, open-ended or creative responses that require subjective interpretation pose a challenge for the semantic capabilities of the current deep learning models, often requiring large volumes of training data to improve reliability.

To address these issues, future improvements should focus on enhancing the robustness of OCR to better handle poor-quality or distorted input. Supporting multilingual evaluation would also expand the system's usability across different regions and educational boards. Introducing adaptive learning models could enable the system to learn from past evaluations, thereby improving its grading logic and flexibility over time. Additionally, incorporating real-time feedback mechanisms could help students understand their mistakes instantly, allowing for a more interactive learning experience.Lastly, integrating Auto Grade with widely used educational platforms like Learning Management Systems (LMS), examination portals, and digital classrooms would further streamline academic workflows. Such integration would allow for smooth data transfer, automated report generation, and performance analytics, providing educators with actionable insights and students with timely, personalized feedback. This would transform Auto Grade from just a grading tool into a comprehensive assessment ecosystem, contributing to a smarter, faster, and more transparent education system.

REFERENCES

[1] F. A. Rawther, R. Rasaldeen, I. R. Jaison, and S. M. Fernandez, "A survey on automating answer-sheet evaluation using AI techniques," International Journal of Engineering Research & Applications (IJERA), vol. 4, no. 2, pp. 1–10, 2023.
[2] B. A. Masa, "Integrating machine learning and NLP efficient retrieval of characters in Pali script preservation," Junikhyat Journal, vol. 24, no. 1, pp. 112–120, Apr. 2024.
[3] B. A. Masa, "Development of an artificial intelligence-based solution for document processing automation using ML and NLP," University of Bologna Thesis, 2022.
[4] A. Chawla, A. Gupta, and K. S. Shushrutha, "Intelligent information retrieval: Techniques for character recognition and structured data extraction," Journal of Emerging Technologies, vol. 5, no. 3, pp. 55–68, 2022.
[5] A. S. Abdullah, S. Geetha, A. B. A. Aziz, and U. Mishra, "Design of an automated model for inspecting and evaluating answer scripts using NLP and deep learning," Alexandria Engineering Journal, Elsevier, 2024.
[6] M. Amutha and G. P. R. G., "OCR-based document parsing and evaluation using AI," in 2024 IEEE International Conference on AI & NLP, 2024, pp. 145–152.
[7] S. V. Mahadevkar, S. Patil, and K. Kotecha, "Exploring AI-driven approaches for unstructured document analysis and future horizons," Journal of Big Data, Springer, vol. 11, no. 2, pp. 234–250, 2024.
[8] M. Bardvall and I. Hassle, "Automating invoice recognition: A comparative study of large language models and OCR/ML technologies," DiVA Portal, 2024.
[9] Y. Qiu, V. C. Duvvuri, P. Yadavalli, and N. Prasad, "AI-based automated essay grading system," in Proceedings of the 2024 ACM International Conference on Educational Technologies, 2024, pp. 78–85.
[10] D. Baviskar, S. Ahirrao, V. Potdar, and K. Kotecha, "Efficient automated processing of unstructured documents using AI," IEEE Access, vol. 12, pp. 14567–14580, 2023.
[11] P. Sil, P. Chaudhuri, and B. Raman, "Can AI assistance aid in the grading of answer sheets?," arXiv preprint arXiv:2408.12870, 2024.
[12] A. M. Naeem, A. Gad, M. Alhalabi, and M. Yaghi, "AI-based mobile paper grading: Trends and challenges," in IEEE Future Internet of Things Conference (FIoT), 2023, pp. 320–328.
[13] P. Nyein, "Implementing automated testing strategy for legal decision understanding AI," Theseus Digital Repository, 2023.