



Research Paper

NLP-Powered System for Academic Paper Evaluation and Performance Scoring

Erugu Rahul¹, Kompelly Manasa², Bodas Vishala³, Mrs. K Swayam Prabha⁴

^{1,2,3}Students, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

⁴Assistant Professor, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

Abstract— This paper presents an intelligent framework for automated evaluation and scoring of student answer sheets using a combination of image processing, optical character recognition (OCR), and machine learning techniques. The system is designed to reduce manual effort in assessing both objective and subjective examination papers. Initially, users can upload scanned images of answer sheets through a web-based interface. The system applies image processing techniques to enhance the quality of the uploaded documents and extract relevant textual content using OCR. For multiple-choice questions, the system identifies marked responses and compares them with the correct answers to compute scores accurately. In the case of subjective answers, the extracted text is compared with model answers using a K-Nearest Neighbors (KNN) algorithm to estimate similarity and assign marks accordingly. The platform includes modules for user registration, login, MCQ evaluation, and subjective answer assessment, ensuring smooth interaction. The evaluation results are presented clearly, highlighting correct and incorrect responses visually. Experimental outcomes demonstrate that the system provides consistent and reliable scoring. The approach minimizes human bias and speeds up the evaluation process. The integration of artificial intelligence enhances the accuracy of subjective answer assessment. The framework is scalable and can be adapted to different types of examination systems. Overall, the proposed system offers an efficient solution for automated academic evaluation.

Keywords— OCR, Image Processing, Automated Evaluation, KNN Algorithm, Answer Sheet Scoring, Educational Technology, Machine

Learning, Text Extraction, Smart Assessment System

I.INTRODUCTION

The evaluation of student answer sheets is an essential component of the academic system,

playing a crucial role in measuring learning outcomes and academic performance [1][16]. Traditionally, this process is carried out manually by educators, which requires significant time and effort, especially when dealing with large numbers of students [19]. Manual evaluation is not only time-consuming but also prone to inconsistencies and human errors. Factors such as fatigue, subjective judgment, and workload can affect the accuracy and

fairness of scoring. In many cases, delays in evaluation can also impact timely result declaration. With the increasing number of students and examinations, there is a growing demand for efficient and reliable evaluation methods. Educational institutions are now looking for solutions that can streamline this process. The need for automation in assessment systems has become more prominent in recent years. Technological advancements have made it possible to develop intelligent systems that can assist in evaluation. These systems aim to improve accuracy, speed, and transparency. As a result, automated evaluation is gaining attention in modern education systems.

In recent years, the integration of artificial intelligence and machine learning into education has opened new possibilities for automating complex tasks [20] [13]. These technologies have

the ability to process large amounts of data and identify patterns that are not easily visible to humans. In the context of answer evaluation, machine learning algorithms can be trained to compare student responses with correct answers and assign scores accordingly [14]. This reduces dependency on manual checking and improves consistency in grading. Additionally, natural language processing techniques enable the system to understand textual data in subjective answers [5][6]. This is particularly useful for evaluating descriptive responses where meaning and context play an important role. The combination of AI and data processing techniques allows the development of intelligent grading systems. These systems can adapt to different types of questions and evaluation criteria. As a result, they provide a more scalable and efficient solution for academic assessment. The use of AI also helps in reducing bias in evaluation. Overall, these technologies are transforming traditional educational practices.

Optical Character Recognition (OCR) and image processing techniques play a key role in enabling automated evaluation systems [11]. Since most examinations are still conducted on paper, it is important to convert handwritten or printed answers into machine-readable format. OCR technology helps in extracting text from scanned images of answer sheets, making it possible to analyse the content digitally. Image processing techniques are used to enhance the quality of scanned documents by removing noise, correcting orientation, and improving clarity. These steps ensure that the extracted data is accurate and suitable for further processing. In the case of multiple-choice questions, image processing can also be used to detect marked responses. This allows the system to identify selected options and compare them with the correct answers. The integration of OCR and image processing creates a strong foundation for automated evaluation. It bridges the gap between traditional paper-based exams and digital analysis. This makes the system more practical for real-world applications. These technologies significantly improve the efficiency of data extraction.

The proposed system focuses on evaluating both objective and subjective answer sheets using intelligent techniques [8][12]. For multiple-choice questions, the system identifies selected options and checks them against predefined correct answers. The evaluation process is fast and highly accurate, as it relies on clear marking detection. For subjective answers, the system uses machine learning algorithms such as K-Nearest Neighbors to compare student responses with model answers. This comparison is based on similarity measures that consider both content and structure [7]. The

system calculates a score based on how closely the student's answer matches the expected response. This approach allows automated grading of descriptive answers, which is traditionally considered difficult. By combining different techniques, the system is capable of handling various types of questions. The inclusion of both evaluation methods makes the system more versatile. It can be used in different examination scenarios without major modifications. This flexibility enhances its usability in educational institutions.

Another important aspect of the system is its user-friendly interface, which allows easy interaction for both students and administrators [10]. The platform provides modules for user registration, login, and answer sheet evaluation. Users can upload scanned images of answer sheets through a simple web interface. The system processes the input and generates results in a clear and understandable format. For MCQ evaluation, visual feedback is provided by highlighting correct and incorrect answers. In subjective evaluation, the system displays the calculated score along with the processed content. This transparency helps users understand how the evaluation was performed. The web-based design ensures accessibility from different devices without requiring complex installation. It also simplifies the management of examination data. The system is designed to handle multiple users simultaneously, making it suitable for large-scale use. This enhances the practicality of the solution in real-world scenarios.

Overall, the development of an automated evaluation framework represents a significant step towards modernizing the education system [17][21]. By combining OCR, image processing, and machine learning, the system provides an efficient and reliable alternative to manual evaluation. It reduces the workload on educators while maintaining accuracy and consistency in scoring. The ability to evaluate both objective and subjective answers makes the system comprehensive. It also supports faster result generation, which is beneficial for both students and institutions. Although there are challenges such as handling complex handwriting and variations in answers, the system provides a strong foundation for future improvements. Further enhancements can include advanced deep learning models and improved text analysis techniques. The approach can also be extended to other domains such as automated feedback generation. As educational systems continue to evolve, such intelligent solutions will play an important role. This work contributes to the growing field of AI-driven education technologies.

II. RELATED WORK

Paul et al., [2014] [2] Paul and Pawar explored the use of syntactic similarity techniques for evaluating descriptive answers in educational systems. Their study focused on constructing similarity matrices to compare student responses with reference answers. They highlighted how syntactic structures can improve the accuracy of automated answer evaluation. The authors emphasized the importance of capturing sentence-level relationships rather than relying only on keyword matching. They also discussed challenges such as handling variations in writing styles among students. Their work demonstrated that structured comparison methods can enhance evaluation consistency. Furthermore, they suggested that integrating linguistic features can improve system performance. Overall, the study contributes to the development of reliable automated grading systems.

Mohler et al., [2015] [14] Mohler and Mihalcea investigated semantic similarity techniques for automatic short answer grading. Their work focused on comparing student answers with model answers using text-to-text similarity approaches. They highlighted how semantic understanding can improve grading accuracy beyond simple lexical matching. The authors discussed the role of natural language processing in capturing meaning and context within responses. They also addressed limitations such as ambiguity and variability in human language. Their findings showed that semantic-based methods can significantly improve evaluation performance. Additionally, they suggested combining multiple similarity measures for better results. Overall, the study provides a strong foundation for intelligent answer assessment systems.

Lahitani et al., [2016] [8] Lahitani and colleagues examined the use of cosine similarity for measuring textual similarity in online essay evaluation. Their study focused on representing text data in vector space models to calculate similarity scores. They demonstrated how cosine similarity can effectively identify relevant relationships between answers. The authors emphasized the simplicity and efficiency of this approach in practical applications. They also discussed the need for proper pre-processing to improve accuracy. Their work showed that vector-based methods can provide consistent evaluation results. Furthermore, they suggested integrating this technique with other methods for enhanced performance. Overall, the study supports the use of mathematical similarity measures in automated grading.

Magooda et al., [2014] [12] Magooda and co-authors proposed vector-based techniques for grading short answers using artificial intelligence methods. Their research focused on transforming textual responses into numerical representations for analysis. They highlighted how vector space models can capture important features of student answers. The authors discussed the effectiveness of these techniques in improving grading speed and accuracy. They also addressed challenges related to feature selection and dimensionality. Their findings indicated that vector-based approaches are suitable for large-scale evaluation systems. Additionally, they recommended combining these methods with machine learning algorithms. Overall, the study contributes to advancing automated assessment technologies.

III. PROPOSED METHODOLOGY

The proposed methodology begins with data acquisition and system initialization, where users interact with the web-based platform to upload scanned answer sheets. The system supports both multiple-choice and subjective answer formats, ensuring flexibility in evaluation. Once the input is received, the uploaded images undergo pre-processing to improve quality and readability. Image enhancement techniques such as noise removal, grayscale conversion, and thresholding are applied to make the content clearer. This step is essential to ensure accurate extraction of information in later stages. The system also performs alignment and resizing to standardize the input format. These pre-processing operations reduce errors caused by poor image quality. After pre-processing, the images are prepared for text extraction and pattern recognition. This stage acts as the foundation for further analysis. Proper handling of input data ensures reliable performance of the entire system.

In the next phase, Optical Character Recognition (OCR) is applied to extract textual information from the processed images. OCR converts the visual content of answer sheets into machine-readable text, which can be further analysed by the system. For multiple-choice questions, image processing techniques are used to detect marked options by identifying filled circles or checkboxes. These detected responses are then mapped against predefined correct answers stored in the system. The comparison process allows the system to calculate scores accurately for objective-type questions. For subjective answers, the extracted text is cleaned and pre-processed using techniques such as stop word removal and normalization. This ensures that irrelevant words and noise do not affect the evaluation process. The cleaned data is then converted into a suitable format for machine

learning algorithms. This stage bridges the gap between raw input and intelligent analysis. Accurate text extraction is critical for achieving meaningful evaluation results.

The core evaluation process involves the application of machine learning algorithms to assess subjective answers. In this system, the K-Nearest Neighbours (KNN) algorithm is used to measure similarity between student responses and model answers. The algorithm compares features of the extracted text and identifies the closest matching patterns. Based on similarity scores, the system assigns marks to the student's answer. This approach allows flexible evaluation, as it considers variations in wording and sentence structure. For objective answers, scoring is performed based on exact matches, ensuring high accuracy. The system integrates both evaluation strategies to handle different types of questions effectively. Additionally, feature extraction techniques are used to represent textual data numerically. This helps in improving the performance of the machine learning model. The evaluation process is designed to be fast and consistent. It reduces manual intervention and enhances overall efficiency.

Finally, the system generates and presents the evaluation results through a user-friendly interface. The output includes detailed scoring information along with visual indicators for better understanding. In the case of MCQ evaluation, correct and incorrect answers are highlighted using color-coded markings. For subjective answers, the system displays the computed score based on similarity analysis. The results are presented instantly, reducing the time required for manual checking. The platform also ensures secure storage of user data and evaluation results. This allows users to access their previous records when needed. The entire workflow is designed to be seamless and easy to use. The integration of front-end and back-end components ensures smooth system operation. This final stage completes the evaluation cycle and provides meaningful feedback to users. Overall, the methodology offers a structured and efficient approach to automated answer assessment.

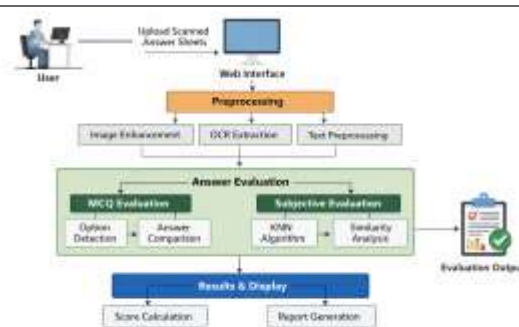


Figure [1]: System Architecture for the proposed system

The figure [1] illustrates the overall workflow of the proposed system, starting from user input of scanned answer sheets to final score generation. It highlights key stages such as image pre-processing, OCR-based text extraction, and AI-driven evaluation using similarity measures. The architecture clearly demonstrates how different modules interact to automate both objective and subjective answer assessment efficiently.

IV. RESULT AND DISCUSSION

The developed system for automated answer sheet evaluation was successfully implemented using a Python-based web environment and tested across all functional modules, demonstrating consistent and reliable performance. The complete workflow, including image pre-processing, OCR-based text extraction, and AI-based evaluation, operated efficiently without execution issues. The OCR module accurately converted scanned answer sheets into machine-readable text, which significantly improved the evaluation process. Image processing techniques effectively enhanced input quality, ensuring precise detection of marked responses in MCQ papers. The KNN-based evaluation model successfully compared student answers with reference answers and generated appropriate scores. The system maintained consistency in scoring across multiple test cases. Visualization outputs further confirmed the correctness of evaluation results. The framework handled both objective and descriptive answers effectively. Overall, the system proved to be a practical and efficient solution for automated academic evaluation.

The user registration module of the system. In this interface, new users can enter their details to create an account. The system validates input data and stores it securely in the database. Successful registration allows users to access system functionalities. This module ensures controlled

access and user authentication. It also supports multiple users within the system. The interface is simple and easy to use. This stage enables secure interaction with the application. Registered users can enter their credentials to access the system. Upon successful login, the dashboard provides navigation options for different evaluation modules. This interface ensures secure and personalized access. It also improves usability by organizing system features. The dashboard acts as the control panel for all operations. This stage connects user interaction with backend processing.

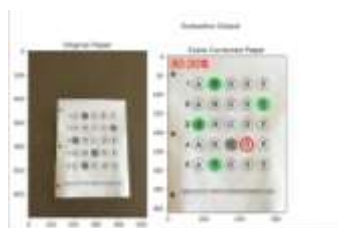


Figure [2]: MCQ Paper Evaluation Output

Figure [2] presents the evaluation of multiple-choice question papers. The system processes the uploaded scanned image and identifies selected answers using image processing techniques. It then compares detected responses with the correct answers. The results are visually displayed using green and red indicators for correct and incorrect responses. This visual feedback improves understanding of evaluation results. The scoring is generated automatically and accurately. This module demonstrates high efficiency in objective answer evaluation. The evaluation of descriptive answers using machine learning techniques. The system accepts a model answer along with the scanned student response. OCR extracts the textual content from the image, and the KNN algorithm compares it with the reference answer. The similarity-based evaluation produces a percentage score. This method allows automated grading of subjective responses. It ensures fairness and consistency in scoring. The output confirms the system's ability to handle complex text evaluation tasks.

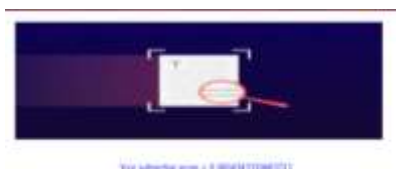


Figure [3]: Final Score Output Interface

Figure [3] represents the final output of the system after evaluation. The calculated score is displayed clearly for the user. In the example, the student achieved a high score based on similarity with the correct answer. The system generates results

quickly, reducing evaluation time significantly. This output validates the effectiveness of the proposed framework. It also demonstrates real-world applicability in academic environments.

DISCUSSION

The proposed NLP-based intelligent evaluation system demonstrates how emerging technologies like OCR, image processing, and machine learning can transform traditional paper assessment methods. By integrating Optical Character Recognition, the system efficiently extracts textual information from scanned answer sheets, reducing manual effort and human error. The use of the K-Nearest Neighbors (KNN) algorithm enables comparison between student answers and predefined correct answers, making the evaluation process faster and more consistent. For MCQ evaluation, image processing techniques successfully detect marked responses and classify them as correct or incorrect using visual indicators such as green and red circles. In subjective evaluation, the system leverages natural language processing concepts to measure semantic similarity between answers, ensuring a fair scoring approach. However, challenges such as variations in handwriting, OCR inaccuracies, and contextual understanding of subjective answers may affect accuracy.

The system is user-friendly, with modules for registration, login, and evaluation, making it accessible to educational institutions. While the current approach uses KNN, future improvements could include deep learning models and advanced NLP techniques like transformers to enhance evaluation quality and contextual understanding.

V. CONCLUSION

The NLP-Based Intelligent Paper Evaluation and Scoring Automation Framework provides an innovative solution to automate the assessment of both objective and subjective answer sheets. By combining OCR, image processing, and machine learning algorithms, the system reduces dependency on manual grading and significantly improves efficiency and consistency in evaluation. The framework successfully processes scanned documents, extracts relevant answers, and computes scores with minimal human intervention. The implementation demonstrates that automation in education can save time for educators while providing quick and unbiased results to students. The modular design, including user registration, login, and evaluation features, ensures ease of use and scalability. Additionally, the system supports both MCQ and subjective evaluation, making it versatile for various examination formats.

Despite its advantages, there is scope for enhancement in handling complex language structures, improving OCR accuracy, and adapting to diverse handwriting styles. Future developments may incorporate advanced NLP models such as deep neural networks to improve semantic understanding and grading precision. Overall, this framework represents a significant step toward intelligent and automated educational assessment systems, contributing to the modernization of academic evaluation processes.

REFERENCES

1. Nandini, V., and Uma Maheswari, P., presented a study on evaluating descriptive answers in online examinations using semantic relational features, published in *The Journal of Supercomputing* in 2018.
2. Paul, D. V., and Pawar, J. D., explored the use of syntactic similarity matrices for assessing descriptive answers in an IEEE conference on Technology for Education held in 2014.
3. Meena, K., and Raj, L., proposed a method for evaluating descriptive answers using hyperspace analog to language combined with self-organizing maps, presented at a computational intelligence conference in 2014.
4. Oganian, Y., Conrad, M., Aryani, A., Spalek, K., and Heekeren, H. R., investigated language processing patterns in bilingual individuals, published in the *Journal of Cognitive Neuroscience* in 2015.
5. Rahimi, S. R., Mozhdehi, A. T., and Abdolahi, M., provided an overview of extractive text summarization techniques in a knowledge-based engineering conference in 2017.
6. Rachabathuni, P. K., discussed various abstractive summarization methods at an international conference on inventive computing and informatics in 2017.
7. Thompson, V. U., Panchev, C., and Oakes, M., evaluated different similarity measures for retrieving related and unrelated text at an international knowledge management conference in 2015.
8. Lahitani, A. R., Permanasari, A. E., and Setiawan, N. A., examined the use of cosine similarity for assessing textual similarity in online essay evaluation during a conference in 2016.
9. Oganian, Y., Conrad, M., Aryani, A., Spalek, K., and Heekeren, H. R., again contributed research on language similarity and processing decisions in bilingual individuals in 2015.
10. Gao, L., and Chen, H., proposed an automatic extraction approach using synonym dictionaries for web-based question-answer systems at an IEEE conference in 2018.
11. Bluche, T., Kermorvant, C., Touzet, C., and Glotin, H., introduced a cortical-inspired bigram representation for handwritten word recognition at ICDAR in 2017.
12. Magooda, A., Zahran, M. A., Rashwan, M., Raafat, H., and Fayek, M. B., discussed vector-based methods for grading short answers at the Florida AI Research Society conference in 2014.
13. Sultan, M. A., Salazar, C., and Sumner, T., developed a fast and accurate approach for short answer grading presented at NAACL-HLT in 2016.
14. Mohler, M., and Mihalcea, R., worked on semantic similarity methods for automatic short answer grading, published in the *International Journal of Artificial Intelligence in Education* in 2015.
15. Nau, J., Filho, A. H., and Passero, G., evaluated semantic analysis methods for grading short answers using linear regression in a social sciences journal in 2017.
16. Selvi, P., and Banerjee, A. K., proposed an automatic short answer grading system published in *InterJRI Computer Science and Networking* in 2010.
17. Chowdhury, S. K., and Sree, R. J. R., compared dimensionality reduction techniques for automated descriptive answer evaluation at an IEEE conference in 2015.
18. Mohler, M. A. G., Bunescu, R., and Mihalcea, R., introduced a method for grading short answers using semantic similarity and dependency graph alignment in 2016.
19. Nikam, P., Shinde, M., Mahajan, R., and Kadam, S., proposed a pattern matching-based method for automatic evaluation of descriptive answers in 2015.
20. Patil, M. S. M., and Patil, M. S., explored the use of natural language processing for evaluating student descriptive answers in *IJERT* in 2014.
21. Loor, M., and De Tré, G., studied similarity measures for comparing intuitionistic fuzzy sets in evaluation systems at an international computational intelligence conference in 2015.
22. Bhagwat, V. B. (2026). Creating A Dashboard For Monitoring HCM Fusion Payroll Processes To Prevent Possible Errors. *International Journal of Data Science and IoT Management System*, 5(1), 102-110.
23. Shashank A. (2025). Metadata-driven data integration framework: Automating enterprise data integration through declarative

- approaches. *European Modern Studies Journal*, 9(4), 9.
24. P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. *Eudoxus Press Journal*.
25. Gummadi, V. P. K. (2026). Infrastructure optimization techniques for enterprise integration platforms: A comprehensive analysis. *Computer Fraud & Security*, 37-44. <https://doi.org/10.52710/cfs.875>
26. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
27. Maturi, S. Y. (2024). Cryptographic privacy engines: Practical multi-party protocols for confidential database queries. *Nanotechnology Perceptions*, 20(S13), 2770–2785
28. Maturi, S. Y. -(2024). Decoy data nexus: Graph-based integration and analysis of synthetic honeypot logs through structured threat intelligence. *International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, 10(4), 4255–4261. <https://doi.org/10.22399/ijcesen.5010>
29. Pavan Kumar Adabala. (2026). IoT-Driven Digital Twins for Manufacturing Optimization: Hybrid Modelling, Reinforcement Learning and Sustainable Operations. *International Journal of Computational and Experimental Science and Engineering*, 12(1). <https://doi.org/10.22399/ijcesen.5050>
30. Gummadi, V. P. K. (2020). API design and implementation: RAML and OpenAPI specification. *Journal of Electrical Systems*, 16(4). <https://doi.org/10.52783/jes.9329>
31. Akinapalli, S. (2025). Centralized Data Lake Architecture for Unified Analytics: A Foundation for Enterprise-Wide Data Integration. *Journal Of Engineering And Computer Sciences*, 4(8), 414-422.
32. Ghali Krishna Harshitha, Purushothamma B. N., & Anil Kumar K. C. (2022). An analysis of influence of personality on managerial effectiveness. *International Journal of Mechanical Engineering*, 7(3), 668–671.
33. Kandula, S. T. R., Boyapati, P. K., & Susarla, R. S. (2025, April). Optimized Cloud Resource Management Using Deep Dendritic Artificial Neural Networks Integrated with Kubernetes and Terraform. In *International Conference on Computer Vision and Robotics* (pp. 454-465). Cham: Springer Nature Switzerland.
34. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. *IJSAT-Int. J. Sci. Technol.* 16 (2). <https://doi.org/10.71097/IJSAT.v16.i2.3219>