WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Automated Answer Sheet Evaluation Using OCR and NLP

B. Santhosh, A. Sagar, B. Manikanta, A. Abhiram, Ch. Mani Bhargavi

Department of Computer Science & Engineering GMRIT, Rajam, Andhra Pradesh, India

ABSTRACT —

Grading examinations in such a traditional way can be laborious and maybe subject to inconsistencies since it relies on human checking and reading of handwritten answer sheets, streamlining the process and avoiding subjective errors will help alleviate the unfair grading problem. This project proposes an automated evaluation system, with an OCR (Optical Character Recognition) component to extract words from handwritten answer sheets, and an NLP (Natural Language Processing) model component to read the text and assess or score the submissions based on semantic relevance and accuracy. The OCR pipeline of the project is driven by EasyOCR, which is a versatile and effective method to detect and read a variety of handwriting styles. After the text is extracted, the text will go through several NLP methods, including: semantic analysis, keyword matching, and scoring based on similarity to measure the quality and correctness of answers regarding ideal model answers. The proposed system also has a web-based user-interface to allow educators to easily upload the answer sheets, view automated grading results, along with displaying the customizable scoring parameters as needed. Furthermore, this solution will reduce and time and effort for grading, it may be more objective and fair it has the potential to be employed within a range of educational and training contexts that have a scalable architecture or adaptable

Keywords— *Automated evaluation, Optical Character Recognition (OCR), Natural Language Processing (NLP), semantic analysis, EasyOCR, grading automation, unbiased evaluation, similarity scoring.*

1. Introduction

Grading an exam the old-fashioned way entails heavy-dutywork and corresponds to biases because it employs human grading to scan through handwritten texts. With human grading comes the malady of subjectivity and biases create room for even gross injustices. This work appreciates this dilemma and proposes a speedy mechanism of evaluation by means of Optical Character Recognition (OCR), which pulls out text from handwritten answers, and Natural Language Processing (NLP), which evaluates and scores them based on semantical relevance and correctness, Henceforth, this speeds up the assessment process while improving parameters of objectivity and fairness in academic grading. The model developed for the optical character recognition (OCR) process by using EasyOCR performs well in various handwriting style detection and recognition ensuring reliable and accurate text extraction. After extraction, the text undergoes various stages of NLP processing which include semantic evaluation, keyword matching, and similarity scoring to rank the correctness and quality of answers against model responses prepared in advance. The accompaniment for the system is a web-based application wherein educators can upload answer sheets and obtain scoring results automatically, allowing educators to modify scoring criteria according to requirement. The interface system enables teachers to participate in evaluation by transparent and flexible means while drastically downscaling the manual workload involved in it. Automated scoring can therefore be favored against traditional methods delivering maximally fast, fair, and efficient solutions. The architecture is scalable and flexible and can be deployed across multiple educational and training environment thereby yielding sustained benefits for academic appearance as well as for professional evaluation.

2. Literature Survey

Several advanced techniques in image processing, machine learning, and natural language processing (NLP) have been employed for automating the evaluation of handwritten answer scripts. [1] Bukhari et al. introduced an OCR-based approach to extract handwritten text, employing CNNs and LSTMs to enhance the accuracy of character recognition in noisy document images. [2] Singh et al. explored the use of BiLSTM models for recognizing and classifying handwritten digits in answer scripts, showing improvements over traditional methods by capturing both forward and backward context in the text. [3] Jain and Shukla applied a hybrid deep learning approach, combining CNNs for feature extraction with LSTMs for sequence learning, which achieved an impressive accuracy in extracting text from scanned handwritten pages. [4] Kaur et al. proposed an OCR system for answer sheet evaluation that incorporated pre-processing techniques like skew correction and noise removal, alongside feature extraction using traditional methods and deep learning techniques for final recognition. [5] Gokhale et al. extended the idea of using neural networks for answer script evaluation by incorporating fuzzy logic, thus enabling a more flexible grading system that accounts for partial correct answers. [6] Ahmed et al. demonstrated the application of a CNN-based model to evaluate handwritten text, presenting a system that combines text recognition and semantic analysis for automatic grading. [7] Kumar and Soni evaluated handwritten answers using a CRNN model, showcasing its efficiency in handling varied handwriting styles and formats in scanned answer sheets. [8] Yadav and Mishra focused on keyword extraction from handwritten texts using BERT-based models, followed by semantic analysis, which helped in understanding the context and meaning of the answers to enhance grading accuracy. [9] Choudhary and Garg proposed a framework that combined image processing techniques with machine learning algorithms to detect and evaluate handwriting in real-time, using a combination of OCR and feature extraction methods. [10] Gupta and Agrawal applied a deep learning architecture consisting of CNNs and BiLSTM networks to evaluate handwritten responses, using domain-specific training data for improved recognition accuracy. [11] Sharma et al. investigated the use of an LSTM network, combined with a semantic parsing module, for handling subjective answer evaluation, showing improvements in understanding context beyond simple keyword matching. [12] Patil et al. proposed an integrated system that combined multiple NLP techniques and OCR for more accurate grading of handwritten answers, especially in open-ended questions. [13] Soni and Sharma introduced a grading system based on machine learning classifiers, including Random Forests and SVM, to categorize answers as correct, incorrect, or partially correct, using features extracted from the handwritten text. [14] Mishra and Tripathi developed a hybrid grading system that utilized both deep learning and rule-based methods to evaluate handwritten answer scripts, emphasizing the integration of contextual understanding in scoring. [15] Gupta et al. applied fuzzy logic to handle ambiguous answers, proposing a more robust system for evaluating handwritten responses, which is capable of adjusting the grading scale based on subjective judgment.

3. Methodology

1. System Architecture Diagram

The following diagram illustrates the workflow of the model:

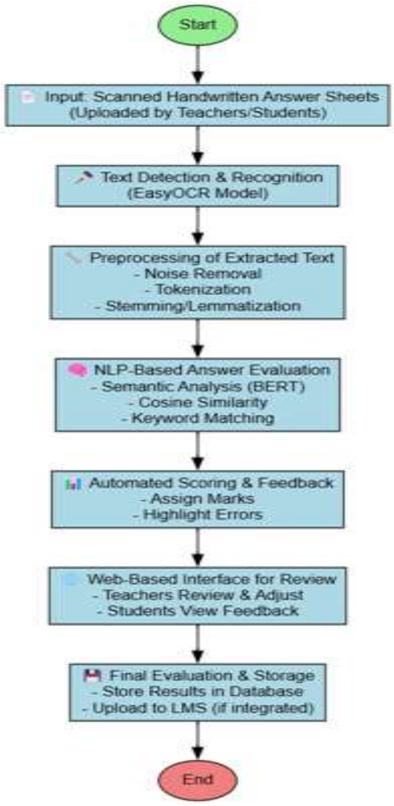

Fig 3.1: Workflow of Automated Answer Sheet Evaluation

Dataset

For this project, the dataset comprises scanned images of handwritten answer sheets submitted by students. Scanned images obtained directly from physical answer sheets, rather than pre-labeled files or datasets. Therefore, the images vary in quality and resolution, often as a result of student differences in handwriting style. As such, the image datasets are susceptible to noise, blurry registration, or distortions. The scanned images are pre-processed for quality improvements, such as noise reduction and alignment. The core focus is obtaining the handwritten text from the scanned images using Optical Character Recognition (OCR) methods that are pre-processed by way of Natural Language Processing (NLP) methods for later evaluation and analysis. This dataset is important for training the model for automatic student response recognition and assessment.

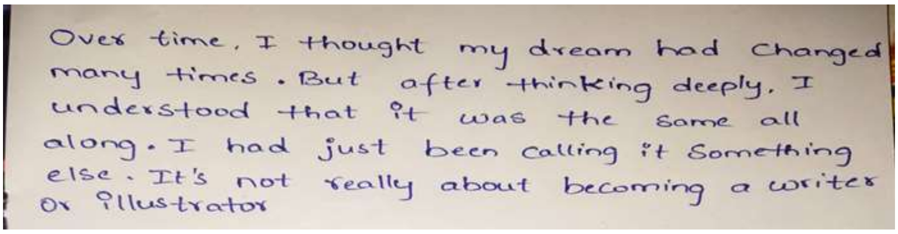

Fig 3.2: Scanned answer sheet collected from the student

Model Architecture

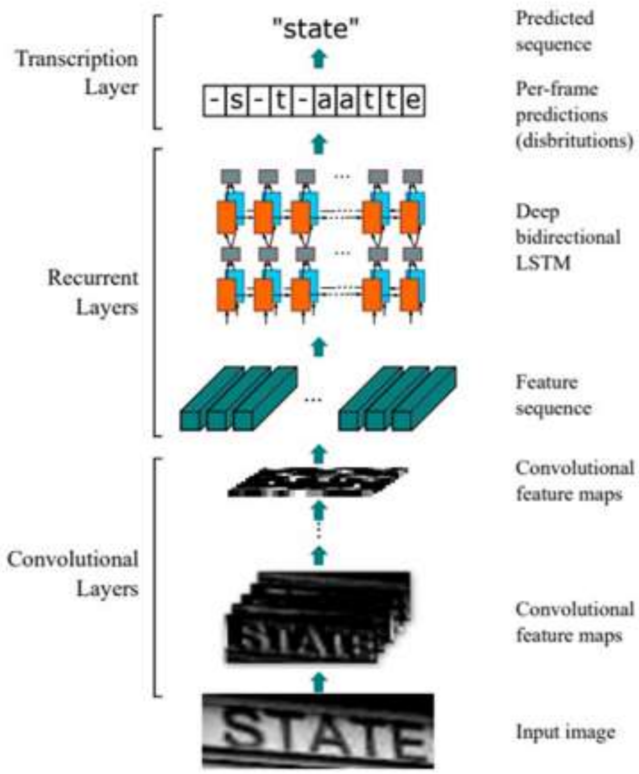

Fig 3.3: Architecture Diagram of the paper evaluation system using EasyOCR.

Extraction of Text from Handwritten Scripts Using EasyOCR and Semantic Answer Evaluation

In our project, we put EasyOCR to work to pull out handwritten text from scanned answer sheets. EasyOCR is a tool based on deep learning that can handle both handwritten and printed text in many languages. It uses CNN and RNN models to spot text without having to break it up by hand. After getting the text, we used Natural Language Processing (NLP) methods to compare what the student wrote with the right answer in a way that makes sense. We first cleaned up the text by breaking sentences into words and changing words to their basic form. This made sure everything was the same. To grasp what the answers meant, we turned to BERT, a strong language model that looks at context and how sentences are built by reading both ways. At the end, we worked out how similar the student's answer was to the model answer using cosine similarity. This gives a score showing how close the meanings are. This way of doing things lets us grade even when students say things.

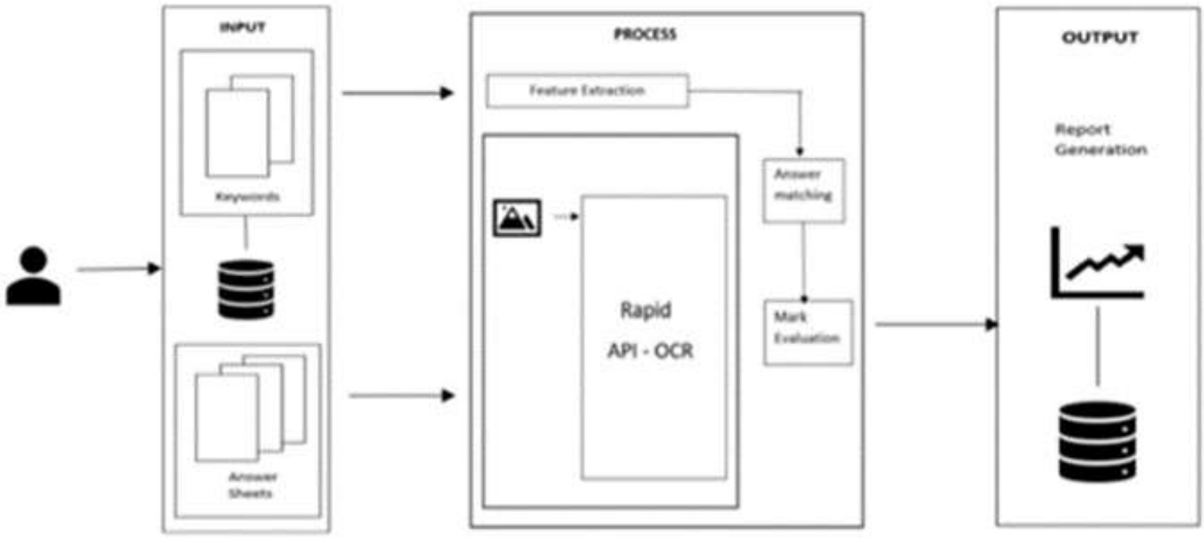

Fig 3.4: The architecture of EasyOCR Network

Implementation Details

Python serves as the foundation for this project relying on key tools like EasyOCR to recognize handwritten text and NLP libraries such as NLTK and transformers to analyze and compare text. The team uses EasyOCR, which is based on PyTorch, to pull handwritten content straight from scanned answer sheet images. After extracting the text, the project applies NLP methods including tokenization, lemmatization, and semantic similarity analysis with BERT to assess student responses. To grasp the meaning of the response beyond exact word matches, the system measures the similarity between student answers and model answers using cosine similarity. Users can access the system through a simple web application built with Flask. This setup allows them to upload scanned answer sheets and get evaluation results right away.

Front-End Technologies

ReactJS plays a key role in building a user-friendly interface on the front end. Its component-based design helps create an interactive experience. The interface features an option for users such as teachers or graders, to upload scanned images of handwritten answer sheets for grading. When someone uploads an image, ReactJS sends the information to the Node.js backend through an API request. The backend then takes charge of processing the image. It uses EasyOCR and NLP methods to pull out and grade the text. After this process, it sends the results back, which show up on the user's screen. The combination of ReactJS and Node.js results in a smooth, quick, and adaptable web app. This setup makes it easier to grade handwritten answers for schools and other educational settings.

2. Evaluation metrics

The evaluation of the proposed system proves that it works efficiently for recognizing handwritten text and grading student responses. After text extraction through OCR, the system applies semantic similarity techniques for comparing students' responses with the model answer. This technique recognizes the meaning behind concepts rather than merely surface-level word matches; thus, it is appropriate for marking answers presented in different styles.

The system generates a score in the similarity range 0 to 1; the following intervals are used for awarding marks based on the similarity score:

- similarity score 0.85–1.00: full marks
- similarity score 0.70–0.84: 75–90 percent
- similarity score 0.50–0.69: 50–70 percent
- similarity score below 0.50: these could even be awarded with partial to no marks depending on relevance.

The above range-based scoring system makes the evaluation fair and consistent all across the board since different students may express the same ideas in very different terms. This thereby enhances the overall output of the system: in providing reliable and valid results in evaluating handwritten answer sheets.

5. Results and Discussions

Student Samples	**Similarity Score**	**Evaluation Result**
Student 1	95.83%	Correct

Student 2	86.00%	Correct
Student 3	75.56%	Partial
Student 4	25.90%	Incorrect

Table 5.1: Similarity metrics of Answer Sheets of different Students.

Fig 5.2: Output of the proposed model

6. Conclusion:

This paper presents an intelligent, effective system for automatic evaluation of handwritten answer sheets using Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques. The system is based on EasyOCR for extracting text from scanned answer sheets and further on advanced preprocessing of obtained texts, tokenization, and lemmatization. The semantic evaluation itself is conducted by a transformer-based model, BERT, which uses cosine similarity scoring to evaluate the relevance and contextual precision between student answers and predetermined model answers. The present technique is more contextual regarding evaluation compared to its predecessor, as it does not focus only on keyword matching.

The results of implementation are promising and new because it involves the use of OCR and deep NLP models that promise automation of manual grading. Its architecture is modular and scalable to accommodate future improvements, such as integration of other NLP metrics, grammar checking, and dynamic scoring rubrics according to types of questions. Future plans will see this system integrated into academic repositories, thus saving time for educators in generating evaluations and offering consistent results free from bias. Expansion innovatively accommodates more languages and handwriting styles, with adaptability for wider use in educational systems.

To sum up, our proposed OCR-NLP-based framework for the evaluation of handwritten answers offers an efficient, effective and intelligent solution which will be useful in educational evaluation in modern times. Future enhancements or integrations, including AI driven feedback, real time and mobile based evaluations, will further improve its usability and impact in the academic sector.

Application and Use Cases:

The evaluation of handwritten answer sheets can be automated in educational institutions with the help of this project. The system can immediately generate evaluation scores by extracting text from scanned copies of answer scripts, comparing it to model answers, and calculating the scores. This minimizes human error in manual inspection and saves time. Additionally, it can be applied to tasks that require a rapid and equitable evaluation of subjective responses, competitive assessments, and online exams. The technology allows teachers concentrate more on feedback and progress than manual correction and guarantees a consistent.

References:

[1] Johri, E., Dedhia, N., Bohra, K., Chandak, P., & Adhikari, H. (2021, May). ASSESS-automated subjective answer evaluation using semantic learning. In Proceedings of the 4th international conference on advances in science & technology (ICAST2021).

[2] Rahaman, M. A., & Mahmud, H. (2022). Automated evaluation of handwritten answer script using deep learning approach. Transactions on Machine Learning and Artificial Intelligence, 10(4).

[3] Rowtula, V., Oota, S. R., & Jawahar, C. V. (2019, September). Towards automated evaluation of handwritten assessments. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 426-433). IEEE.

[4] Kulkarni, M., Adhav, G., Wadile, K., Chavan, R., & Deshmukh, V. (2024). Digital Handwritten Answer Sheet Evaluation System.

[5] Q. -D. Nguyen, N. -M. Phan, P. Krömer and D. -A. Le, "An Efficient Unsupervised Approach for OCR Error Correction of Vietnamese OCR Text," in IEEE Access, vol. 11, pp. 58406-58421, 2023, doi: 10.1109/ACCESS.2023.3283340.

[6] Chen, X., Cong, P., & Lv, S. (2022). A long-text classification method of Chinese news based on BERT and CNN. IEEE Access, 10, 34046-34057.

[7] Prabakaran, N., Kannadasan, R., Krishnamoorthy, A., & Kakani, V. (2023). A Bidirectional LSTM approach for written script auto evaluation using keywords-based pattern matching. Natural Language Processing Journal, 5, 100033.

[8] Najam, R., & Faizullah, S. (2023). Analysis of recent deep learning techniques for Arabic handwritten-text OCR and post-OCR correction. Applied Sciences, 13(13), 7568.

[9] Nandwalkar, B., Pardeshi, S., Shahade, M., & Awate, A. (2023). Descriptive Handwritten Paper Grading System using NLP and Fuzzy Logic. International Journal of Performability Engineering, 19(4), 273.

[10] Patil, R., Kulkarni, A. A., Ghatage, R., Endait, S., Kale, G., & Joshi, R. (2024). Automated Assessment of Multimodal Answer Sheets in the STEM domain. arXiv preprint arXiv:2409.15749.

[11] Agbemuko, D., Okokpujie, I., Salami, M., & Tartibu, L. K. (2024). Automated Data Extraction and Character Recognition for Handwritten Test Scripts Using Image Processing and Convolutional Neural Networks. *Nigerian Journal of Technological Development*, *21*(4), 97-115

[12] Rahman, M. M., & Akter, F. (2019). An automated approach for answer script evaluation using natural language processing. IJCSET (www. ijcset. net), 9, 39-47.

[13] Abdullah, A. S., Geetha, S., Aziz, A. A., & Mishra, U. (2024). Design of automated model for inspecting and evaluating handwritten answer scripts: A pedagogical approach with NLP and deep learning. Alexandria Engineering Journal, 108, 764-788.

[14] Rahman, M. M., & Siddiqui, F. H. (2020). NLP-based automatic answer script evaluation (Doctoral dissertation, DUET Journal).

[15] Bashir, M. F., Arshad, H., Javed, A. R., Kryvinska, N., & Band, S. S. (2021). Subjective answers evaluation using machine learning and natural language processing. IEEE Access, 9, 158972-158983.