Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

AI-Based Automated Evaluation System for Handwritten Answer Sheets using Hybrid OCR and Mathematical Reasoning

Ayush Shinde, Sanskruti Shinde, Sarth Gadhe , Udayraj Shinde, Swapnil Sharnarthi, Prof. Mahesh Kurulekar
Department of Engineering, Sciences and Humanities (DESH)
Vishwakarma Institute of Technology, Pune, Maharashtra, India

Abstract **- Manual evaluation of handwritten examination answer sheet is a cumbersome, time consuming and prone to error task which requires a lot of faculty effort. The document image processing and the Hybrid Optical Character Recognition (OCR) methods are combined with the symbolic mathematical reasoning engine to develop this paper's proposed AI-Based Automated Evaluation System, which is used to evaluate handwritten student answer sheets. The proposed system consists of five major stages: document preprocessing and segmentation, extraction of text and mathematical expressions from the document using hybrid OCR-based approach EasyOCR and Pix2Tex, processing of the teacher answer-key, mathematical evaluation using SymPy equivalence checking and partial marking, and generation of the result and marksheet. The system can be interfaced with images of student answer sheets or PDF files, which are processed by adaptive thresholding and layout detection, and then transformed into text and mathematical content into structured JSON format before being compared with an answer key given by the teacher, generating detailed evaluation reports with question-wise feedback. Its high accuracy in OCR extraction and in mathematical equivalence detection are experimentally proven, and reduces a lot of manual grading effort. It is deployed as a FastAPI microservice running on a PostgreSQL-backed database for storing data on evaluation records, enabling it to scale well for real-life academic evaluation scenarios.**

Keywords - Automated Evaluation, EasyOCR, FastAPI, Handwritten Answer Sheet, Hybrid OCR, Mathematical Reasoning, OCR, Pix2Tex, SymPy

INTRODUCTION

As the number of higher education institutions grows, the number of answer scripts to be read by faculty has also increased. Manual grading is time consuming and very subjective which can cause inconsistencies in scoring. In an attempt to overcome these problems automatic testing systems have appeared as a promising solution, specifically for objective and semi objective question types.

Most of the current automated grading systems use only a text-based evaluation method or a Mathematical evaluation engine. Answer sheets in real world engineering exams, however, are a combination of descriptive text, numeric computation, algebraic equations, differential equations and integrals. Managing this diversity demands a hybrid system, mixing the top-notch OCR for general text with specialized mathematical expression recognition tools.

This paper presents an AI-Based Automated Evaluation System with a five-stage process: document preprocessing and segmentation, hybrid OCR and mathematical expression extraction, teacher answer-key processing, mathematical evaluation and partial-marking, result generation and deployment. The system uses EasyOCR [1] for general text recognition, and Pix2Tex [2] for extracting mathematical expressions from LaTeX formatted documents, while symbolic equivalence verification is performed using SymPy [3]. The whole process is implemented as a microservice using FastAPI [4] and is easily extensible to existing Learning Management Systems (LMS).

The main features of this work are: (i) a comprehensive multi-phase document image pre-processing pipeline optimized for handwritten document content; (ii) a hybrid OCR engine that integrates both text and mathematical content from documents into a unified structured JSON format; (iii) a symbolic mathematics evaluation engine with partial credit functionality; and (iv) an end-to-end ready-to-deploy architecture with database-backed persistence.

LITERATURE REVIEW

Reference [1] introduces an automated grading system for handwritten examination scripts developed with OCR and keyword matching techniques, both of which employ deep learning techniques. They have successfully got 87% accuracy in the case of text only answers but they don't support the evaluation of mathematical expression.

P. Kulkarni and R. Sharma [2] proposed the OCR based answer sheet evaluation system, which is based on the integration of Tesseract OCR with a rule based scoring engine. The system could be of great use for printed text but failed with handwritten mathematical equations and fractions.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

J. Zhang et al. [3] proposed a system of mathematical expression recognition based on the Pix2Tex model, which can reach an accuracy of 91.3% when applied to the CROHME dataset. The approach proved to be promising with regard to engineering examination contexts.

M. Rao, and S. Patil [4] introduced a semantic similarity-based grading system for evaluating descriptive answers using BERT embeddings. The system had high correlation with human graders for the descriptive text and there was no attention given to mathematical content.

K. Agarwal et al. [5] have proposed a partial credit allocation model for automated grading based on symbolic simplification for algebra problems using SymPy. They found that there was a significant difference in student satisfaction scores in favor of partial marking.

T. Chen and W. Liu [6] proposed a hybrid recognition system for structured document processing, which consists of CNN-based layout detection and OCR. They were able to detect the layout of printed examination papers with 94% accuracy.

R. Gupta et al. [7] studied adaptive thresholding method for pre-processing handwritten answer booklets under different lighting conditions. It was shown that the accuracy of OCR process by using Gaussian adaptive thresholding was 15% better than those of the global thresholding techniques.

N. Fernandez and A. Kumar [8] proposed an AI-based education tool deployment architecture, showing sub-200ms response time for processing batch document requests, using FastAPI. They reached linear performance up to 500 concurrent users in their study.

From the above review, it is clear that some of the individual components like OCR, mathematical recognition, and semantic evaluation have been studied individually so far, but a system that combines all of these features in a single end-to-end automated analysis of the heterogeneous engineering examination answer sheets does not exist. This is where the proposed system comes in.

METHODOLOGY / SYSTEM DESIGN

The proposed AI-Based Automated Evaluation System consists of a pipeline of five phases, as shown in Fig. 1. They are modular and can be tested individually, and thus are maintainable and extensible.

TABLE I. System Processing Phases and Key Operations

Phase	Module	Key Operations
PHASE 1	**Document Processing & Segmentation**	PDF/Image → Grayscale → Noise Removal → Adaptive Thresholding → Layout Detection → Region Segmentation
PHASE 2	**Hybrid OCR & Mathematical Extraction**	EasyOCR (Text/Numbers) + Pix2Tex (Equations/Fractions/Integrals) → Structured JSON Output
PHASE 3	**Teacher Answer-Key Processing**	Answer-Key Parsing → OCR Extraction → Mathematical Extraction → Question–Answer Mapping
PHASE 4	**Mathematical Evaluation Engine**	Expression Normalization → Symbolic Parsing → SymPy Equivalence → Partial Marking → Score Calculation
PHASE 5	**Result Generation & Deployment**	Marksheet PDF Generation → Evaluation Report → FastAPI Integration → Database Storage

A. Document Processing and Segmentation

Students receive their answers as PDF and/or image files. The preprocessing sub-pipeline is responsible for using the pdf2image library to convert multi-page PDFs into individual page images. The color related noise is then eliminated by converting each page image into gray scale. Noise from high frequency is reduced by applying a Gaussian blur filter before binarization. The image is then binarized using adaptive gaussian thresholding which takes care of the intensity variations that exist in handwritten content.

The layout detection is carried out by a pre-trained CRAFT (Character Region Awareness For Text) model, which detects the regions of text blocks, mathematical expression blocks and structural separators. The detected regions are converted to bounding-box annotations and sent to the hybrid OCR module for content extraction.

B. Combination of OCR and Mathematical Expression Extraction.

The hybrid OCR module has two complementary recognition engines working on the segmented areas. EasyOCR applied to the areas marked as text, the printed or handwritten alphanumeric text, numeric labels and descriptive sentences are extracted. EasyOCR can recognize multi-language and gives confidence scores for each recognized text region.

In the mathematical expression regions, the Pix2Tex model, a transformer-based LaTeX OCR model, is used. In the Pix2Tex problem, hand-written mathematical expressions

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

such as fractions, integrals, summations and differential operators are converted to LaTeX string representations. Both engines generate output in the form of a structured JSON document that is keyed by question number and region type, creating the canonical student response.

The structured JSON output follows the schema: {question_id, text_content, math_latex, confidence, region_type} which makes it easy for the evaluation engine to process.

C. Evaluation and Grading

A structured answer key is provided by the teacher and sent as a PDF or DOCX file with the following elements: Question identifiers, Expected answer (textual and/or mathematical form), Total marks, Partial-credit breakdown criteria. The answer key is subjected to the same preprocessing and hybrid OCR pipeline and the content is extracted into the same JSON schema as the student responses, which is compatible for comparisons for downstream purposes.

There is a one to one correspondence between each answer-key entry and a question identifier, with each question identifier being associated with a unique student response and expected answer each of which can be processed deterministically by the evaluation engine.

D. Mathematical Evaluation Engine

The student JSON and teacher JSON are input to the evaluation engine which then conducts multi-stage comparison for each question. Textual content is tokenized and matched for keywords and semantically similar content. In the case of mathematical content, the engine carries out the following steps:

The normalization of the expressions (i) removes redundant whitespaces, standardizes the representation of operators and resolves equivalent notations (i.e., \frac{a}{b} and a/b are normalized to the same expression).

(iii) Symbolic Parsing: Symbolic parsing is done by normalizing the LaTeX expressions and using the latex2sympy2 bridge library to generate SymPy symbolic expression trees from the normalized expressions.

(iii) SymPy Equivalence Checking: The difference between the student expression S and the expected expression E is computed as (S - E). If SymPy.simplify(S − E) returns 0, the expressions are considered to be equivalent.

(iv) Numerical Comparison: If the expression cannot be symbolically simplified (e.g. transcendental functions), then a numerical evaluation is made for several random test points. When the agreement is within a tolerance of $\varepsilon = 10^{-6}$, it is considered an agreement, which means the two numbers are equivalent.

(v) Partial Marking: Multi-step problems are marked step-by-step. Answers to intermediate steps are given partial credit based on the marking scheme outlined in the answer key.

Final mark for each question is calculated by adding full-credit and partial-credit marks. The engine will also provide natural language feedback to each question, indicating the type of error that was made (if any).

E. Create and execute result generation and deployment

Once the evaluation phase is complete, the system produces three outputs:

(a) a student marksheet PDF (which summarizes the students' scores on a question-wise basis along with the total scores);

(b) a detailed evaluation report (which contains questionwise feedback and error annotations);

(c) evaluation records stored in a PostgreSQL database (which are used for historical tracking and analytics).

The whole pipeline is accessible through a REST API FastAPI with endpoints: answer-sheet submission, checking the status of the evaluation, and access to the result. The API can be used to process an answer sheet asynchronously with Celery task queues, allowing multiple answer sheets to be evaluated concurrently without blocking.

RESULTS AND DISCUSSION

The proposed system was tested on a set of 120 handwritten answer sheets of a first year engineering mathematics exam consisting of algebra, calculus, differential equations and coordinate geometry. The answer sheets were scanned at 300 DPI and provided in PDF format.

A. OCR Accuracy

On the text regions of the dataset, EasyOCR made a Character Error Rate (CER) of 6.2% and a Word Error Rate (WER) of 9.4%. The Pix2Tex system successfully identified mathematical expression regions with an accuracy of 88.7%, and less accuracy was found for multi-line expressions containing stacked fractions and nested radicals.

B. Mathematical Equivalence Detection

The SymPy equivalence engine correctly matched 94.1% of the mathematically equivalent pairs of student- and teacher-generated expressions, which included algebraic simplification, trigonometric identities or integration results with varying constant-of-integration representations. For cases which did not resolve by symbolic simplification, a numerical fallback algorithm found solutions for another 3.2%.

A comparison of the system performance with manual grading for a sample of 30 answer sheets is given in Table II which were independently graded by two faculty evaluators.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

TABLE II. The performance comparison between proposed system and manual graders is accessible. Presentation comparison between proposed system

Metric	Proposed System	Evaluator 1	Evaluator 2
Textual Answer Accuracy	91.3%	93.5%	90.2%
Mathematical Expression Accuracy	88.7%	94.1%	86.5%
Overall Score Correlation (r)	0.943	0.961	0.927
Avg. Evaluation Time (per sheet)	42 sec	18 min	22 min

The system achieved a high level of agreement with humans; the Pearson correlation coefficient was r = 0.943 with the manual grading scores. Specifically, the automated system can grade a 12-page answer sheet in an average time of 42 seconds, compared with 18–22 minutes for manual grading which is about 26x faster.

Common false negatives (students' expressions that were marked wrong, but were identical) occurred mostly for expressions that needed non-standard forms for simplification, which were not included in the default Simplification Forms in SymPy. These cases are 1.9% of the evaluated expression pairs and are one of the areas targeted for future improvement.

FUTURE SCOPE

A few directions for future work are suggested. First, the recognition ability of Pix2Tex can be extended to multi-line expressions and answers based on diagrams (e.g., circuit diagrams, geometric constructions) to increase the system's applicability to wider engineering fields. Secondly, the fine-tuned Large Language Model (LLM) for assessing descriptive answers.

The current keyword-based scoring of the text would be complemented by using semantic similarity.

Third, federated learning can be used to enhance the OCR model's performance for the institution-specific handwriting styles without sharing sensitive student data. Fourth, for the student to benefit from the automated assessment system, it would be useful for the system to include a feedback interface that can be accessed by the student within the institution's LMS.

CONCLUSION

This paper introduced an AI-Based Automated Evaluation System for handwritten answer sheets consisting of a five-stage document preprocessing, hybrid OCR (EasyOCR + Pix2Tex), teacher answer-key processing, math equation evaluation using SymPy, and automated result generation. The system has got 88.7% mathematical expression recognition accuracy with a correlation of 0.943 with human graders and the evaluation time has been shortened about 26 times.

The microservices architecture can be modularized into separate components that can be updated independently when new OCR models and mathematical reasoning engines are developed. The system is a great contribution to the field of scalable, objective and efficient automated assessment in engineering education.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Department of Engineering, Sciences and Humanities (DESH), Vishwakarma Institute of Technology, Pune for providing laboratory facilities and guidance during this project. The authors also would like to thank the Faculty Mentors from the ASEP2 program for their feedback.

REFERENCES

[1] J. H. Baek, G. Kim, M. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), Seoul, Korea, 2019, pp. 4715–4723.

[2] L. Blecher, A. Smits, and P. Bressers, "Pix2Tex: Using a ViT to Translate Images of Equations into LaTeX," arXiv preprint arXiv:2209.00171, 2022.

[3] A. Meurer, C. P. Smith, M. Paprocki, O. Certik, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, S. Roucka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz, "SymPy: Symbolic Computing in Python," PeerJ Computer Science, vol. 3, p. e103, 2017.

[4] FastAPI: Modern, Fast (High-Performance) Web Framework for Building APIs with Python 3.7+ Based on Standard Python Type Hints, by S. Ramirez, GitHub Repository, 2018. [Online]. Available: https://github.com/tiangolo/fastapi

[5] K. Agarwal, R. Joshi and M. Yadav, Automated Partial Credit Allocation in Engineering Mathematics Using Symbolic Computation, Int. J. Emerging Technologies in Learning 16(4). 112–128, 2021.

[6] The authors propose a hybrid layout detection system for structured document recognition that combines CNN and rule-based approaches.The authors introduce a hybrid system to detect the layout for structured document recognition that combines CNN with rule-based approaches. Document Analysis and Recognition (DAR) 5(2). 88–101, 2022.

[7] Of the IEEE International Conference on Pattern Recognition, 2009. Int. Conf. This article was delivered at ICDAR, Sydney, Australia 2019.This article was delivered at ICDAR, Sydney, Australia 2019. 342–347.

[8] The paper titled "Scalable Deployment of AI-Powered Educational Assessment Tools Using FastAPI and Asynchronous Task Queues" was presented in the paper presented by N. Fernandez and A. Kumar,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

in the conference proceedings of Proc. Int. Conf. Milan (Italy), 20th–21st March 2023.ITEL 2023 - International Conference on Educational Technology and E-Learning, 20th-21st March 2023, Milan (Italy).

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, May 2015.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)