International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Automated Paper Evaluation System for Subjective Handwritten Answers

P. Farooq Khan[1], A. Sri Harinath Yadav[2], K. Mohan Chandu[3], P. Vishnu Vardhan[4], N. Sashi Kumar[5], Y. Dasaratha Rami Reddy[6]
[1]UG Student, Department of Computer Science & Engineering Artificial Intelligence, CBIT, Proddatur, YSR, AP
[2] UG Student, Department of Computer Science & Engineering Artificial Intelligence, CBIT, Proddatur, YSR, AP
[3] UG Student, Department of Computer Science & Engineering Artificial Intelligence, CBIT, Proddatur, YSR, AP
[4] UG Student, Department of Computer Science & Engineering Artificial Intelligence, CBIT, Proddatur, YSR, AP
[5] UG Student, Department of Computer Science & Engineering Artificial Intelligence, CBIT, Proddatur, YSR, AP
[6]Professor, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
Corresponding Author E-mail: khanfarooq86360@gmail.com

Abstract

The evaluation of descriptive answers in a handwritten format is a tedious and non-standardized process in academic institutions. The correction of answers written in a handwritten format is a laborious process for educators, and it often results in a varying score for students based on personal interpretation. With the increasing number of students, it has become a necessity to have a standardized and objective evaluation system.

The objective of this project is to create an "Automated Paper Evaluation System" designed to assess students' handwritten subjective responses through the application of Artificial Intelligence methodologies. Optical Character Recognition (OCR) techniques are utilized to extract handwritten answers from scanned answer sheets. Subsequently, semantic similarity is determined by contrasting the extracted answers with the model answers, leveraging transformer models like BERT and Sentence Transformers.

This system evaluates the conceptual alignment between student responses and the model answer, as opposed to a keyword-based comparison.

It supports multiple modes of evaluation, batch processing of answer scripts, and score analysis. The proposed system reduces the time consumed for evaluating answer scripts and eliminates personal bias and inconsistency.

Keywords: Automated Paper Evaluation, Optical Character Recognition (OCR), Semantic Similarity, Transformer Models (BERT), Handwritten Answer Assessment, Artificial Intelligence Evaluation System

Introduction

Assessment constitutes a cornerstone of educational practice. Although objective assessments lend themselves to automated evaluation through the application of predetermined responses, the evaluation of descriptive or subjective answers typically necessitates manual review within the majority of educational settings.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Manual grading, however, introduces a number of significant difficulties:

• Substantial time investment

• Variability in scoring among different evaluators

• Prolonged processing times for results

• Challenges in scaling to accommodate large student populations

• The potential for human bias in the interpretation of responses

Recent developments in Natural Language Processing (NLP) and deep learning have facilitated the semantic evaluation of text, as opposed to a purely syntactic approach. Transformer-based models, exemplified by BERT, facilitate a contextual comprehension of sentences, thereby rendering the automated evaluation of descriptive answers a practical possibility.

Literature Review

The transformer model introduced by Vaswani et al. (2017) was a major advancement in NLP. It replaced sequential processing with a self-attention mechanism, allowing for more efficient contextual learning across the entire sentence. This model laid the groundwork for improved contextual models of language.

Devlin et al. (2019) created the BERT model, a bidirectional transformer. This model learns the relationships between words by considering both the words before and after them. This approach led to better performance on various NLP tasks, such as semantic similarity and text classification.

To address the computational challenges of similarity tasks between sentences, Reimers and Gurevych (2019) created Sentence-BERT. This model produces fixed-length embeddings for sentences that are optimized for similarity analysis. It allows for effective computation of semantic similarity using cosine distance metrics.

Smith (2007) provided a detailed review of the Tesseract OCR system, which demonstrated its ability to scan documents and turn them into digital text. The OCR system is a necessary preprocessing step for automated evaluation systems because it enables the digitization of handwritten text.

Research in automated essay scoring supports AI-based evaluation. Attali and Burstein (2006) showed that machine learning methods can reliably score essays using linguistic features. Similarly, Le and Burstein (2013) provided an analysis of automated essay scoring systems across different fields, emphasizing the move toward semantic-based evaluation models.

Kenter and de Rijke (2015) looked into short-text similarity using word embeddings, highlighting the importance of vector-based semantic comparison. Jurafsky and Martin (2022) offered foundational theories in speech and language processing that back contextual language modeling techniques. Additionally, Wolf et al. (2020) described the Transformers library, which makes it easier to use transformer-based models in real-world applications.

Proposed System

The proposed system will integrate optical character recognition (OCR) technology with a semantic evaluation using a deep learning model, which will be used to automatically grade the descriptive responses of students in a handwritten format.

Working Procedure:

- Uploading of scanned answer sheets.
- Extraction of responses from the answer sheets.
- Comparison of responses with reference answers.
- Calculation of semantic similarity of responses with reference answers.
- Awarding of marks based on the evaluation modality chosen.
- Generation of analytical reports.

Key Features:

- Extraction of responses using Tesseract OCR.
- Transformer-based evaluation of semantic similarity.
- Support for multiple evaluation modes (Liberal, Balanced, Strict).
- Support for batch uploading of directories.
- Accuracy metrics and reports available.
- Bias-free evaluation system.

System Architecture

The proposed Automated Paper Evaluation System is conceptualized as a system with a modular and role-based architecture, comprising user interaction layers and a centralized processing engine. The system is built around a client-server model, where users will be able to interact with the system through a web interface, while computational processing will be done by a server, which will be integrated with an Artificial Intelligence (AI) evaluation engine.

The system architecture will be divided into three major user interaction modules, namely Student, Teacher, and Admin, each of which will be associated with a centralized Database and AI Engine. The proposed system will be modular, ensuring a clear division of responsibilities, as users will be able to interact with the system according to their role, with

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

all transactions happening through a unified backend layer.

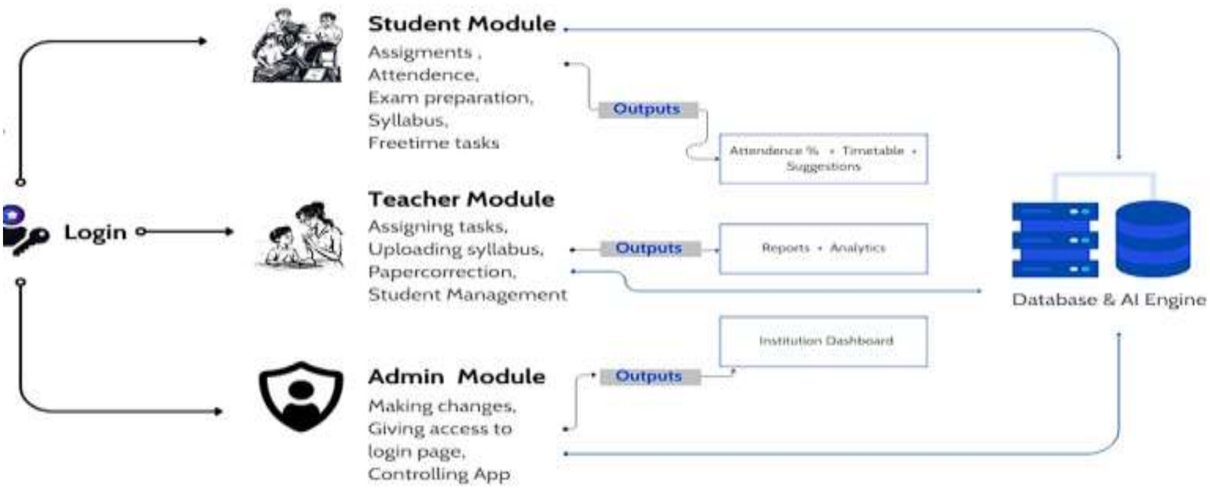

Figure 1

Authentication Layer

The system begins with a secure authentication mechanism. Users log in through a credential validation process that verifies identity and assigns access permissions based on predefined roles. Passwords are securely stored using hashing mechanisms, and session management ensures authorized access throughout user interaction. This layer acts as the gateway to all modules and enforces access control policies.

Student Module

The Student Module provides access to academic resources and evaluation results. Students can view assignments, attendance records, timetables, syllabus content, and examination preparation materials. The module also presents AI-generated outputs, including attendance statistics, personalized suggestions, and evaluation results.

When a student submits an answer script, the document is forwarded to the backend for processing. The extracted and evaluated results are then displayed through the student dashboard. This ensures transparency and real-time feedback while maintaining centralized processing.

Teacher Module

The Teacher Module facilitates academic management and evaluation activities. Teachers can upload assignments, provide reference answers, manage student records, and initiate paper correction processes.

Once answer scripts are uploaded, they are transmitted to the AI Engine for processing. The system performs text extraction using OCR, generates contextual embeddings using

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

transformer-based models, computes semantic similarity, and assigns scores. The results are stored in the database and presented through analytical dashboards.

This module generates performance reports and statistical insights that assist educators in monitoring academic progress and maintaining grading consistency.

Admin Module

The Admin Module controls system-level operations and institutional monitoring. Administrative users manage user roles, control access permissions, configure application settings, and oversee institutional analytics.

Unlike other modules, the Admin Module does not directly interact with the evaluation process. Instead, it ensures system stability, security, and compliance with institutional policies. The centralized dashboard provides a comprehensive overview of user activity and performance metrics.

Database and AI Engine

The Database and AI Engine form the core processing layer of the architecture. The database stores user credentials, attendance records, uploaded scripts, reference answers, and evaluation outcomes. Structured storage enables efficient retrieval and analytics generation.

The AI Engine performs the primary computational tasks. It executes the following sequential operations:

1. Optical Character Recognition (OCR) for handwritten text extraction.
2. Text preprocessing and normalization.
3. Generation of contextual embeddings using transformer-based language models.
4. Semantic similarity computation using cosine similarity measures.
5. Score normalization and result generation.

This centralized processing ensures consistent grading logic across all users and reduces redundancy within individual modules.

Data Flow

The system follows a structured data flow. After authentication, user requests are routed to the appropriate module. Data submitted through the interface is transmitted to the backend server, where it is processed by the AI Engine. The processed results are stored in the database and retrieved dynamically for dashboard visualization.

This layered architecture supports scalability, maintainability, and performance optimization while ensuring secure handling of academic data.

Experimental Results

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Student ID	Student Name	Total Marks	Max Marks	Percentage	Grade
ST001	Alice Johnson	36.05	50	72.1	B+
ST002	Bob Smith	23.94	50	47.88	D
ST003	Carol Davis	28.66	50	57.32	C
ST004	David Wilson	21.19	50	42.38	D
ST005	Emma Martinez	28.69	50	57.38	C

Student ID	Student Name	Question No	Marks Obtained	Max Marks	Percentage	Grade
ST001	Alice Johnson	1	7.45	10	74.47	B+
ST001	Alice Johnson	2	7.61	10	76.08	B+
ST001	Alice Johnson	3	7.01	10	70.13	B+
ST001	Alice Johnson	4	6.67	10	66.7	B
ST001	Alice Johnson	5	7.31	10	73.13	B+
ST002	Bob Smith	1	4.45	10	44.47	D
ST002	Bob Smith	2	4.73	10	47.29	D
ST002	Bob Smith	3	4.95	10	49.48	D
ST002	Bob Smith	4	4.52	10	45.23	D
ST002	Bob Smith	5	5.29	10	52.88	C
ST003	Carol Davis	1	5.92	10	59.24	C
ST003	Carol Davis	2	6.12	10	61.17	B
ST003	Carol Davis	3	5.33	10	53.3	C
ST003	Carol Davis	4	4.8	10	48.02	D
ST003	Carol Davis	5	6.49	10	64.87	B
ST004	David Wilson	1	4.04	10	40.41	D
ST004	David Wilson	2	4.13	10	41.28	D
ST004	David Wilson	3	4.62	10	46.23	D
ST004	David Wilson	4	4.24	10	42.4	D
ST004	David Wilson	5	4.16	10	41.56	D
ST005	Emma Martinez	1	6.02	10	60.22	B
ST005	Emma Martinez	2	5.82	10	58.24	C
ST005	Emma Martinez	3	4.93	10	49.28	D
ST005	Emma Martinez	4	5.46	10	54.64	C
ST005	Emma Martinez	5	6.46	10	64.58	B

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Metric	Value
Total Students	5
Average Marks	27.71
Highest Marks	36.05
Lowest Marks	21.19
Average Percentage	55.41
Pass Rate (>40%)	100.0%
Distinction Rate (>75%)	0.0%

Conclusion

Author(s) Contributions

All authors contributed to system design, implementation, experimentation, and manuscript preparation.

REFERENCES

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova (2019), “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Proceedings of NAACL-HLT, pp. 4171–4186.
- [2] N. Reimers and I. Gurevych (2019), “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” Proceedings of EMNLP, pp. 3982–3992.
- [3] A. Vaswani et al. (2017), “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS).
- [4] R. Smith (2007), “An Overview of the Tesseract OCR Engine,” Proceedings of ICDAR, vol. 2, pp. 629–633.
- [5] C. K. Le and J. Burstein (2013), “Automated Essay Scoring: A Cross-Disciplinary Perspective,” International Journal of Artificial Intelligence in Education, vol. 23, no. 1–4, pp. 1–19.
- [6] Y. Attali and J. Burstein (2006), “Automated Essay Scoring with e-rater® V.2,” Journal of Technology, Learning, and Assessment, vol. 4, no. 3.
- [7] T. Kenter and M. de Rijke (2015), “Short Text Similarity with Word Embeddings,” Proceedings of CIKM, pp. 1411–1420.
- [8] F. Chollet et al. (2015), “Keras,” https://keras.io .
- [9] T. Wolf et al. (2020), “Transformers: State-of-the-Art Natural Language Processing,” Proceedings of EMNLP: System Demonstrations, pp. 38–45.
- [10] D. Jurafsky and J. H. Martin (2022), Speech and Language Processing, 3rd ed., Pearson.