RESEARCH ARTICLE

EDUCHECKAI: A PROPOSAL FOR AUTOMATED PAPER EVALUATION USING DEEP LEARNING AND REINFORCEMENT LEARNING

Omkar Sherkhane

Pillai College of Arts, Commerce & Science (Autonomous), New Panvel

Corresponding author E-mail: omkarsherkhane@mes.ac.in

DOI: https://doi.org/10.5281/zenodo.18069116

Abstract:

Grading student examinations has long been recognized as one of the most demanding tasks in education. It requires significant time, careful attention, and consistency, yet human evaluators are often subject to fatigue and bias. This paper introduces *EduCheckAI*, a design framework that aims to automate paper evaluation through artificial intelligence. The system combines three essential technologies: Convolutional Neural Networks (CNNs) for recognizing handwritten responses and parsing document layouts, Transformer-based Natural Language Processing (NLP) models for assessing descriptive answers, and Reinforcement Learning (RL) for refining scoring policies and awarding partial credit fairly. Instead of presenting experimental data, this study draws on established research to outline expected outcomes. Prior work shows CNNs can achieve over 90% accuracy in handwriting recognition, transformer-based essay scoring systems often reach agreement levels of 0.75–0.85 with human graders, and RL methods improve fairness in partial credit allocation. Based on these benchmarks, *EduCheckAI* is expected to deliver reliable, scalable, and equitable grading. The framework provides a roadmap for future implementation and empirical validation, contributing to innovation in educational technology.

Keywords: Automatic Grading, Deep Learning, Convolutional Neural Networks, Reinforcement Learning, Educational Technology, Natural Language Processing.

1. Introduction:

Education systems worldwide continue to rely heavily on manual grading. While this approach ensures human judgment, it is time-consuming and vulnerable to inconsistency. With growing student populations and increasingly diverse assessment formats, the need for scalable and fair evaluation methods has become urgent. Artificial intelligence offers promising solutions by automating repetitive tasks and providing consistent scoring.

Existing systems, however, face limitations. Many struggle to process handwritten responses, evaluate subjective answers, or ensure fairness in scoring. This paper proposes *EduCheckAI*, a conceptual framework designed to address these challenges. By integrating CNNs for handwriting

recognition, transformer-based NLP for semantic evaluation, and RL for adaptive scoring, the system aims to achieve near-human grading accuracy while maintaining fairness and scalability.

2. Literature Review:

Research on automated grading has evolved significantly. Early Automated Essay Scoring (AES) systems relied on statistical models, focusing on surface features such as word count and sentence length. More recent approaches employ transformer-based NLP models, which capture semantic meaning and align more closely with human evaluation.

In parallel, handwriting recognition has advanced through CNN-LSTM hybrids, achieving strong performance in digitizing handwritten text. These models are particularly effective in handling diverse scripts and noisy inputs.

Adaptive scoring has also been explored, with reinforcement learning offering dynamic credit assignment. RL methods allow systems to learn from expert feedback, improving fairness in partial credit allocation.

Despite these advances, gaps remain. Few systems integrate handwriting recognition, semantic evaluation, and RL into a single pipeline. Moreover, issues of fairness, explainability, and deployment in diverse educational contexts are often overlooked.

3. Methodology:

The proposed *EduCheckAI* framework is structured around three core components:

- Dataset (planned): Scanned answer sheets annotated by expert graders.
- Preprocessing: Document segmentation and CNN-based handwriting recognition to digitize responses.
- Grading pipeline:
    - Objective questions → rule-based and semantic similarity checks.
    - Subjective answers → transformer embeddings aligned with rubric criteria.
    - RL module → optimizes partial credit assignment based on expert feedback.
- Evaluation plan: System outputs will be compared with human graders using metrics such as kappa agreement, correlation, and fairness audits.

4. Results:

Although empirical testing has not yet been conducted, expected outcomes can be drawn from literature benchmarks:

- Objective answers: Accuracy around 95% is anticipated, based on prior OCR and rule-based systems.
- Subjective answers: Transformer-based models typically achieve kappa agreement of 0.75–0.85 with human graders.
- Handwriting recognition: CNN-based OCR systems report character accuracy above 90%.
- RL scoring: Studies suggest RL can improve partial credit fairness by 10–12%.
- Feasibility: The system is expected to process hundreds of papers per hour, with robustness against handwriting variation and spelling errors.

5. Discussion:

The proposed framework demonstrates strong potential when viewed against existing benchmarks. CNNs provide reliable handwriting recognition, transformers enable semantic evaluation of descriptive answers, and RL enhances fairness in scoring. Together, these components form a practical pipeline for automated grading.

Implications:

If implemented, *EduCheckAI* could reduce teacher workload, accelerate feedback, and support adaptive learning environments. Limitations: The framework remains conceptual; empirical validation and dataset collection are necessary. Future work: Building a prototype, expanding datasets, enhancing explainability, and integrating with learning management systems are key next steps.

Conclusion:

This paper presents *EduCheckAI*, a design framework for automatic paper evaluation using CNNs, NLP transformers, and reinforcement learning. While empirical testing is pending, literature benchmarks indicate strong potential for reliable, fair, and scalable grading. The framework offers a roadmap for future implementation, contributing to interdisciplinary innovation in education technology.

Acknowledgment:

The author acknowledges the guidance of faculty members and the support of the Department of Computer Science, Pillai College of Arts, Commerce & Science (Autonomous), New Panvel in developing this proposal.

References:

1. Graves, A., & Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in Neural Information Processing Systems*, 21, 545–552.
2. Dong, F., Zhang, Y., & Yang, Y. (2017). Attention-based recurrent neural network for automatic essay scoring. *Proceedings of EMNLP*, 1072–1081.
3. Shermis, M. D., & Hamner, B. (2013). Contrasting state-of-the-art automated scoring of essays. *Journal of Writing Research*, 5(2), 1–23.