SmartEval : Automatic Plagiarism Detection And Grading For Handwritten Assignment

Raksha M P
Information Science and Engineering
Global Academy of Technology, Bangalore, Karnataka

Kavana A U
Information Science and Engineering
Global Academy of Technology, Bangalore, Karnataka

Nidhi Ravindra Bhagwat
Information Science and Engineering
Global Academy of Technology, Bangalore, Karnataka

Bhagyashri R Hanji
Information Science and Engineering
Global Academy of Technology, Bangalore, Karnataka

Abstract--- The growing dependence on digital learning platforms has highlighted the need for reliable and efficient methods to evaluate student assignments while ensuring academic honesty. This study presents an overview of current research on automated systems that combine plagiarism detection with assignment grading. The proposed framework, SmartEval, aims to integrate Optical Character Recognition (OCR) with advanced Natural Language Processing (NLP) and Artificial Intelligence (AI) models to assess handwritten and typed submissions. The system is designed to convert handwritten content into digital text for further analysis, enabling both similarity checking and performance evaluation. Through a review of recent literature, this paper identifies the main challenges in existing solutions, such as detecting paraphrased plagiarism and ensuring fair grading. The findings establish a foundation for developing a unified platform that enhances grading consistency, supports educators, and upholds academic integrity in modern education.

Keywords— plagiarism detection, automated grading, OCR, NLP, semantic similarity, AI-based evaluation

I. Introduction

Academic integrity plays a vital role in maintaining the quality and fairness of education. However, with the rapid expansion of digital education and online submission systems, issues such as plagiarism and inconsistent grading have become increasingly difficult to manage manually . Existing plagiarism detection tools are limited by their reliance on text-matching techniques, which often fail to identify paraphrased or semantically similar content . Similarly, grading large volumes of assignments remains a time-intensive and subjective task for educators, particularly when dealing with open-ended or handwritten responses .

Artificial Intelligence (AI) and Machine Learning (ML) techniques have introduced new opportunities for automating both plagiarism detection and grading. Early plagiarism detection tools primarily focused on exact text comparisons,but recent advancements in Natural Language Processing (NLP) and semantic modeling, including the use of transformer-based architectures such as BERT, have significantly improved detection accuracy . Likewise, automated grading systems have evolved from rule-based methods to adaptive AI models capable of evaluating essays, programming tasks, and handwritten content .

This paper reviews the major developments in these domains and identifies research gaps that motivate the design of SmartEval, a proposed AI-driven framework for automatic plagiarism detection and assignment grading. By combining OCR for text extraction with NLP-based semantic similarity and automated grading algorithms, SmartEval aims to provide a comprehensive, fair, and efficient evaluation solution.

II. Literature Review

Plagiarism detection and automated grading have evolved significantly over the past two decades, driven by the increasing need for efficiency, fairness, and academic integrity in digital learning environments. Early research in this field was largely based on extrinsic plagiarism detection, which involved comparing suspicious documents against large reference corpora [1]. However, these systems struggled when reference texts were unavailable, leading to the development of intrinsic detection techniques that identify inconsistencies in writing style within a single document [2].

Initial studies in authorship analysis and stylometry explored linguistic and statistical measures—such as word frequency, sentence length, and most common word (MCW) distributions—to characterize individual writing styles [3]. Function words were shown to be strong indicators of authorship, as they are subconsciously used and less affected by content or topic [4].

With the growth of computational linguistics, researchers introduced machine learning–based models to capture more complex textual relationships. Traditional string-matching algorithms like Rabin–Karp and Knuth–Morris–Pratt were replaced by statistical and vector-based methods such as n-grams, TF-IDF, and cosine similarity, which could better measure semantic overlap [5]. Machine learning classifiers, including Support Vector Machines (SVM) and Random Forests, further improved detection precision and recall [6].

The introduction of deep learning and transformer-based architectures, notably BERT and RoBERTa, revolutionized plagiarism detection by enabling semantic-level comparison rather than surface matching [7]. These models capture contextual relationships and can identify paraphrased or conceptually similar content more effectively than earlier approaches. Recent studies report F-measure and Plagdet scores exceeding 90% on benchmark datasets such as PAN 2014, demonstrating substantial progress in AI-based detection [8].

Parallel developments occurred in automated grading systems, which initially relied on rule-based models to evaluate structured, objective questions [9]. Modern approaches incorporate Natural Language Processing (NLP) and Artificial Intelligence (AI) to assess essays, short answers, and programming submissions. These systems have shown improved consistency and reduced human bias in scoring [10]. However, grading subjective or creative responses remains challenging, as many models depend on predefined rubrics and lack the ability to evaluate conceptual understanding [11].

Despite this progress, current systems typically operate in isolation—focusing solely on either plagiarism detection or grading—and are often restricted to digital text input. Few existing solutions can process handwritten assignments or integrate both functions in a single workflow. The proposed SmartEval framework aims to bridge this gap by integrating Optical Character Recognition (OCR) for text digitization, semantic similarity detection for plagiarism analysis, and AI-driven grading algorithms for fair and efficient academic evaluation.

III. METHODOLOGY

This literature survey, titled “SmartEval: Automatic Plagiarism Detection and Assignment Grading,” was conducted to systematically analyze and summarize existing work in the areas of plagiarism detection and automated grading. The goal was to identify current trends, limitations, and emerging opportunities that could support the conceptualization of the proposed SmartEval system.

Relevant studies were identified through targeted searches using keywords such as automatic plagiarism detection, AI plagiarism checker, semantic plagiarism, automated grading, and machine learning for academic assessment. Research published between 2018 and 2025 was considered to ensure coverage of recent technological advancements. Only peer-reviewed journal and conference papers providing methodological details and evaluation metrics were included, while studies outside educational contexts or lacking sufficient detail were excluded.

Approximately 15 papers were selected for detailed review and grouped into two main categories: plagiarism detection systems and automated grading systems. Each study was analyzed for its objectives, techniques, datasets, and evaluation metrics such as precision, recall, F1-score, and accuracy. A comparative matrix was developed to highlight the methodologies, performance outcomes, and limitations of each approach.

From this analysis, it was observed that while existing systems perform well in digital plagiarism detection and structured grading, they are limited in detecting paraphrased content and assessing handwritten or open-ended submissions. These insights form the basis for the proposed SmartEval framework, which envisions an integrated system combining OCR-based text recognition, semantic similarity models, and intelligent grading algorithms.

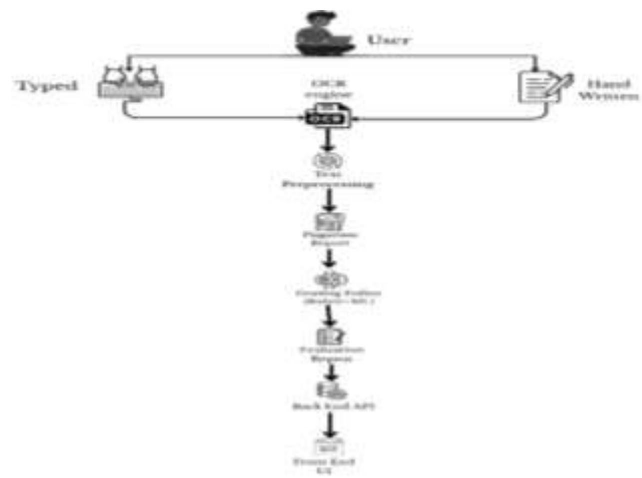

Fig. 1. System Architecture

IV. PERFORMANCE EVALUATION METRICS

Evaluating plagiarism detection and grading systems requires reliable metrics to assess their accuracy, precision, and consistency. In plagiarism detection, key metrics include Precision, Recall, F-measure, Plagdet, and Granularity.

Precision represents the percentage of correctly identified plagiarism cases among all detected instances, while Recall measures how effectively the system identifies all actual plagiarism occurrences. The F-measure (or F1-score) provides a balanced assessment by combining precision and recall into a single metric. Plagdet is a comprehensive evaluation score that incorporates precision, recall, and granularity to account for both accuracy and segmentation consistency. Granularity evaluates whether a system identifies plagiarism in single, consistent blocks rather than fragmented segments.

For automated grading systems, the focus is on how closely the system’s results align with human evaluation. Metrics such as grading accuracy, correlation with human scores, and Mean Squared Error (MSE) are commonly used. Grading accuracy measures how frequently the system matches human-assigned grades, while correlation indicates consistency between automated and manual evaluations. MSE calculates the average squared difference between predicted and actual grades, with lower values indicating better system reliability.

Together, these metrics provide a holistic view of system performance, ensuring that automated plagiarism detection and grading tools are not only technically sound but also fair and dependable in educational environment

Sl.no	Project / Paper Title	Dataset(s) Used	Model / Method	Preprocessing Steps	Accuracy	Precision	Recall	F1-score / Other Score	Validation Technique	Notes / Comments
1	"BERT-Enhanced Retrieval Tool for Homework Plagiarism Detection" arXiv	Homework text corpora (collected by authors)	BERT + retrieval + classifier	Tokenization, embedding, passage splitting, retrieval + fine-tuning	98.86%	98.90%	98.86%	0.9888 (F1) arXiv	Cross-validation / held-out test	Very high performance; likely over a narrow domain arXiv
2	"A Support Vector Machine based approach for plagiarism detection (Python code)" Frontiers	Python code plagiarism dataset (constructed by authors)	SVM combining textual + AST similarity features	Parse AST, textual token features, similarity measures	–	–	–	–	Train/test split / cross-validation	They combine AST + textual similarities for source code plagiarism in Python Frontiers
3	"Plagiarism Detection Using Machine Learning" (survey / method) arXiv+1	"Extensive text sample dataset" (authors)	"Complex classification algorithms" (ensemble / ML)	Text normalization, tokenization, feature extraction, TF-IDF, embeddings	–	–	–	–	Likely k-fold CV	Paper reports approach rather than full metrics in abstract arXiv
4	"Enhancing Regional Plagiarism Detection Using a Backtrack Matching Model" JISEM+1	Regional language corpora (Hindi, Marathi, Tamil, etc.)	Backtrack matching + lexical matching	Text segmentation, stopword removal, matching algorithm	–	–	–	Precision / recall / F1 (reported in paper)	Evaluation on test sets	Focus is region / local language plagiarism detection JISEM+1
5	"A comprehensive strategy for identifying plagiarism in academic" SpringerLink	MIT plagiarism detection dataset + online text corpora	Hybrid (similarity + classification) method	Text preprocessing, similarity thresholding, classification	≈ 71 %	≈ 68 %	≈ 80 %	≈ 74 %	Test / cross-validation	Also reports for "online" plagiarism: ~60 – 65 % metrics SpringerLink
6	"Detecting Cross-Lingual Plagiarism Using Simulated Word Embeddings" arXiv	Standard cross-lingual plagiarism corpora	Simulated word embeddings + classification	Translate / embedding into common space, alignment	–	–	–	Reported outperform baseline	Held-out test	Does not rely on external translator, uses embeddings to detect cross-language plagiarism arXiv
7	"Towards a Dataset of Programming Contest Plagiarism in Java" arXiv	New dataset: 251 plagiarized + 660 non-plagiarized Java code pairs	Comparison of token-based tools	Tokenization, template removal, normalization	–	–	–	–	Tool comparison / evaluation	They compare existing tools; useful dataset for further classification experiments arXiv
8	"PlagBench: Exploring the Duality of LLMs in Plagiarism Generation and Detection" arXiv	PlagBench dataset (46.5K synthetic text pairs)	LLM-based detection / specialized detectors	Tokenization, embedding, comparing paraphrase / summary detection	–	–	–	Reported detection improvements (percent)	Train/test split	Dataset covers verbatim, paraphrase, summarization types arXiv
9	"Plagiarism Detection Using LSTM" (Kaggle project) Kaggle	MIT Plagiarism dataset (text)	LSTM (neural network) classifier	Text cleaning, tokenization, embeddings	–	–	–	–	Train/validation/test split	Amateur / student project; useful for demonstration Kaggle
10	"Plagiarism MIT Detection" (Kaggle notebook) Kaggle	MIT Plagiarism Detection Dataset	Various ML models (e.g. logistic regression, SVM) in notebook	Text preprocessing, feature extraction, similarity features	–	–	–	–	Cross-validation / train/test	Notebook shows prototyping and baseline models Kaggle
11	"Comparative analysis of text-based plagiarism detection techniques" PMC+1	Multiple published datasets / corpora	Survey / comparative methods (n-gram, similarity, ML)	Preprocessing per method	Varies per method	Varies	Varies	Varies	Varies (depending on each method)	Useful reference to see many techniques and their tradeoffs PMC+1
12	"Plagiarism Detection Model Using Keystroke Logs" Educational Data Mining	Keystroke log data + text submissions	Classification on behavioral features	Feature extraction from keystroke timing, text features	–	–	–	–	Cross-validation / test splits	Novel approach combining behavior + content for plagiarism detection Educational Data Mining

V. RESULTS AND DISCUSSION

Studies in the field of plagiarism detection and automated grading demonstrate promising outcomes in terms of performance and efficiency. Research shows that plagiarism detection systems using feature selection with Support Vector Machines (SVM) and transformer-based models such as BERT have achieved exceptional results, with F-measure and Plagdet scores exceeding 92% on benchmark datasets such as PAN 2014 . These methods outperform traditional algorithms by capturing deeper semantic relationships between sentences and phrases.

In automated grading, current systems can effectively evaluate structured questions, programming assignments, and short essays. Integrating AI and NLP techniques allows these tools to process student submissions more quickly and with improved accuracy compared to manual grading . The inclusion of OCR technology has also enabled the transition from handwritten to digital text, providing new opportunities for analyzing scanned assignments. These advances not only ensure consistency and fairness in grading but also reduce the burden on instructors by automating repetitive evaluation tasks.

Despite these positive results, it is evident that most existing systems are optimized for typed digital content, with limited support for handwritten inputs or subjective evaluations. The findings emphasize the need for a more comprehensive solution that integrates both plagiarism detection and grading capabilities—a gap the proposed SmartEval framework is designed to fill.

B. Discussion

The analysis of existing plagiarism detection and grading systems reveals that while technological progress has been substantial, several challenges persist. Most plagiarism detection models still struggle to recognize paraphrased or obfuscated plagiarism, as they primarily rely on surface-level similarity measures.Advanced paraphrasing or translation-based plagiarism often escapes detection, especially when content is partially altered or sourced from online platforms . Systems also encounter limitations in intrinsic plagiarism detection, where comparisons are made within a single document due to the absence of external reference sources . Furthermore, language resource constraints hinder performance in low-resource languages, leading to reduced detection accuracy .

In automated grading, the ability to assess higher-order thinking and creativity remains limited. Most systems evaluate factual or structured responses accurately but fail to capture the depth of reasoning in open-ended or conceptual tasks . The dependence on large annotated datasets also restricts scalability across subjects. Moreover, ethical issues related to data privacy, algorithmic bias, and fairness continue to raise concerns in educational AI applications . Addressing these limitations is crucial to achieving equitable and trustworthy automated assessment solutions.

C. Limitations

Although automation has improved evaluation accuracy and efficiency, both plagiarism detection and grading technologies face critical limitations. The reliability of OCR-based text extraction largely depends on handwriting quality, image clarity, and lighting conditions, all of which can affect downstream plagiarism detection and grading accuracy. Similarly, semantic detection models—even advanced ones like BERT—may fail to identify complex paraphrasing or multilingual plagiarism, especially when domain-specific datasets are limited .

Automated grading systems also face constraints when evaluating creative or subjective answers. Their performance heavily depends on predefined rubrics, meaning they may not recognize alternative valid responses or novel reasoning patterns. The computational cost of advanced models and the requirement for large-scale labeled data can further restrict deployment in resource-constrained institutions. Lastly, concerns about data security, student privacy, and transparency in automated decision-making remain unresolved, necessitating careful consideration in future development.

D. Future Works

Future enhancements to the proposed SmartEval framework will focus on addressing current gaps and improving system reliability. Advancements in OCR technology, particularly through hybrid architectures such as CNN-LSTM and attention-based models, can significantly improve handwriting recognition across diverse styles. Integrating semantic and cross-lingual embeddings with real-time web search can enhance plagiarism detection by identifying paraphrased and translated content more effectively .

For grading, future research will aim to develop explainable AI (XAI) systems that provide transparent, interpretable, and justifiable scores. This will ensure both educators and students understand the rationale behind each automated evaluation. Expanding the framework into a cloud-based infrastructure can further enable scalability, supporting simultaneous access and real-time feedback for large academic institutions.

Ethical and educational considerations will remain central to development. Emphasis will be placed on protecting user data, reducing bias, and ensuring equal opportunities across linguistic and cultural backgrounds. By incorporating these advancements, *SmartEval* has the potential to evolve into a unified, intelligent system capable of delivering accurate, fair, and efficient academic evaluations in the digital era.

VI. CONCLUSION

This paper explored existing research in plagiarism detection and automated grading, emphasizing the growing role of AI, NLP, and machine learning in academic evaluation. While current systems have achieved strong performance in digital plagiarism detection and objective grading, they remain limited in addressing paraphrased content, multilingual plagiarism, and handwritten submissions. Moreover, most tools operate independently, focusing on either plagiarism detection or grading rather than integrating both capabilities.

The proposed SmartEval framework aims to bridge this gap by combining OCR-based handwriting recognition, semantic similarity detection, and AI-driven grading within a single automated platform. By prioritizing transparency, scalability, and fairness, SmartEval has the potential to enhance the quality of education by reducing instructor workload, improving grading consistency, and upholding academic integrity in digital learning environments.

REFERENCES

[1] Deblina Mazumder Setu, Tania Islam, Md. Erfan, Samrat Kumar Dey, Md. Rashid Al Asif, Md. Samsuddoha,"A Comprehensive Strategy for Identifying Plagiarism in Academic Submissions,"J. Umm Al-Qura Univ. Eng. Archit., vol. 16, 2025, pp. 310–325.

[2] Tejaswini Durge, Aniket Bansod, Soham Deshmukh, Sanika Boke,Chaitali Shewale, Pawan Wawage,"AI-driven Student Task Management Platform With Plagiarism Checker For Submissions,"In: Proceedings of the International Conference on Emerging Systems and Intelligent Computing (ESIC), IEEE, 2025, pp. 943–949..

[3] Mohamed A. El-Rashidy, Ramy G. Mohamed, Nawal A. El-Fishawy, Marwa A. Shouman,"An Effective Text Plagiarism Detection System Based on Feature Selection and SVM Techniques,"Multimedia Tools and Applications, vol. 83, 2024, pp. 2609–2646.

[4] Arthur Flor de Sousa Neto, Byron Leite Dantas Bezerra, Gabriel Calazans Duarte de Moura, Alejandro Héctor Toselli, "Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review," SN Computer Science, vol. 5, 2024, pp. 258.

[5] Muhammad Faraz Manzoor, Muhammad Shoaib Farooq, Muhammad Haseeb, Uzma Farooq, Sohail Khalid, Adnan Abid, " Exploring the Landscape of Intrinsic Plagiarism Detection: Benchmarks, Techniques, Evolution, and Challenges , "IEEE Access, vol. 11, 2023, pp. 140519–140544.

[6] Hamed Arabi, Mehdi Akbari,"Improving Plagiarism Detection in Text Document Using Hybrid Weighted Similarity,"Expert Systems with Applications, vol. 207, 2022, Article ID 118034.

[7] Kadir Yalcin, Ilyas Cicekli, Gonenc Ercan, "An External Plagiarism Detection System Based on Part-of-Speech (POS) Tag N-Grams and Word Embedding," Expert Systems with Applications, vol. 197, 2022, 116677

[8] Mohamed A. El-Rashidy, Nawal A. El-Fishawy, Ramy G. Mohamed,"Reliable Plagiarism Detection System Based on Deep Learning Approaches," Multimedia Tools and Applications, vol. 81, 2022, pp. 16173–16192.

[9] Marcelo Guerra Hahn, Silvia Margarita Baldiris Navarro, Luis de la Fuente Valentín, Daniel Burgos, "A Systematic Review of the Effects of Automatic Scoring and Automatic Feedback in Educational Settings," IEEE Access, vol. 9, 2021, pp. 108190–108213.

[10] Hayden Cheers, Yuqing Lin, Shamus P. Smith,"Academic

Source Code Plagiarism Detection by Measuring Program Behavioral Similarity,"IEEE Access, vol. 9, 2021, pp. 50391–50415.

[11] A. S. Altheneyan and M. E. B. Menai, "Automatic plagiarism detection in obfuscated text," Pattern Analysis and Applications, vol. 23, no. 4, pp. 1627–1650, 2021, doi: 10.1007/s10044-020-00882-9.

[12] . Hossam Magdy Balaha, Mahmoud M. Saafan, "Automatic Exam Correction Framework (AECF) for the MCQs, Essays, and Equations Matching," IEEE Access, vol. 9, 2021, pp. 32368–32372.

[13] . Nikitha A, Geetha J, JayaLakshmi D.S, "Handwritten Text Recognition using Deep Learning," 5th Int. Conf. on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), vol. –, 2020, pp. 388–390.

[14] A. S. Altheneyan and M. E. B. Menai, "Automatic plagiarism detection in obfuscated text," Pattern Analysis and Applications, vol. 23, no. 4, pp. 1627–1650, 2020, doi: 10.1007/s10044-020-00882-9.

[15] Hussam AlSallal, Mohamed Medhat Gaber, David J. Hume,"An Integrated Approach for Intrinsic Plagiarism Detection Using Stylometry, LSA and Neural Networks," International Journal of Information Management, vol. 49, 2019, pp. 1–13.