Design and Implementation of Machine Learning Algorithms for Automated Student Assignment Grading

Duo Chen[1], Fang Xu[2]
[1]Business English Department, School of Foreign Languages, Harbin University, Harbin 150086, Heilongjiang, China
[2]Business English Department, Harbin Finance University, Harbin 150030, Heilongjiang, China

Corresponding author: Duo Chen (e-mail: franki007@163.com)

ABSTRACT With the continuous advancement of educational digital transformation, automated student assignment grading systems have become indispensable for improving teaching efficiency and ensuring objective assessment in large-scale learning environments. As intelligent educational platforms increasingly rely on wireless communication infrastructures and electromagnetic information transmission for real-time data interaction and distributed computing, reliable automated grading algorithms play a critical role in supporting scalable digital education. To address the low efficiency and subjectivity of traditional manual grading, this study proposes a deep learning-based grading framework that integrates a grading-standard awareness mechanism with multi-dimensional feature extraction. By dynamically learning the semantic characteristics of grading criteria, the proposed algorithm accurately evaluates assignments across different subjects and educational stages while adapting to diverse instructional objectives. Experimental results obtained from 15,000 Chinese essays and 20,000 mathematical solution assignments demonstrate grading accuracies of 87.6% and 83.2%, respectively, representing improvements of 12–15 percentage points over existing approaches. A one-year pilot deployment in three middle schools reduced teachers' grading time by an average of 62% and increased grading consistency to 91.5%, while scoring consistency under varying teaching objectives reached 89.3%. The results indicate that the dynamic grading-standard adaptation mechanism is the primary contributor to performance improvement and provides an effective technical framework for intelligent educational assessment systems operating over smart communication networks and electromagnetic information infrastructures.

INDEX TERMS automated grading, deep learning, educational assessment, smart communication networks, electromagnetic information transmission

I. INTRODUCTION

EDUCATIONAL digital transformation is progressing at an unprecedented pace, driving the adoption of automated student assignment grading systems as a vital component of education informatization to improve teaching efficiency [1]. According to the China Educational Informatization Development Report (2023), investment in China's educational informatization has exceeded 500 billion yuan, signaling the rapid advancement of digital transformation in education. However, traditional assignment grading methods remain predominantly manual. Teachers spend an average of 30–40 minutes grading a complete set of student work (comprising multiple complex tasks), which not only imposes a heavy workload but also suffers from inconsistent standards and strong subjectivity.

In the context of complex engineering problems or multi-step mathematical proofs, evaluating the logical progression of a single multi-part question typically requires several minutes of focused manual review. In practical teaching scenarios, the author participated in an educational digitization project at a provincial key middle school and witnessed firsthand the fatigue and pressure teachers experience while grading assignments. A junior high school Chinese teacher once confided: "Grading essays feels like manual labor every time. I have to focus on content quality while also checking for grammatical errors. Often, I can only grade 20 essays in one class period." This situation not only affects teachers' motivation but also limits the timeliness of feedback, making it difficult for students to receive prompt and accurate responses to their work.

It is noteworthy that existing automated grading systems are mostly based on rule-matching or simple machine learning methods, possessing limited capabilities in understanding complex semantics, leading to significant deviations from manual grading results. According to a survey by the Editorial Department of China Educational Technology & Equipment, over 65% of teachers consider existing automated grading systems "too rigid" and incapable of adjusting scoring standards based on teaching objectives [2]. Furthermore, these systems often lack targeted design for the characteristics of assignments in different subjects and educational stages, resulting in uneven grading effectiveness.

This research focuses on core issues in automated student assignment grading. By deeply analyzing the semantic relationship between grading standards and assignment content, a grading algorithm integrating a grading standard awareness mechanism with multi-dimensional feature extraction is designed to accurately evaluate technical reports and fabric surface morphology data. Compared to existing research, the innovations of this paper are: (1) Constructing a grading standard semantic repository, enabling the system to adapt to different teaching objectives by dynamically learning the deep meanings of standards; (2) Designing a multi-dimensional feature extraction method combining surface features and semantic features to improve grading accuracy; (3) Validating the system's general applicability across different educational scenarios in practical applications.

II. RELATED WORK

In recent years, scholars domestically and internationally have conducted extensive research in the field of automated assignment grading. In the Chinese context, Xia Linzhong et al. [3] built an automated essay grading system based on an LSTM model, achieving an accuracy rate of 82.3%. Qu Tiehua and Gao Haibing [4] explored the theoretical foundations of digital transformation in educational assessment from an ontological perspective. Rong Juan [5] designed an automated grading system for university assignments, focusing on improving grading efficiency. Duo Chen F X et al. [6] proposed a multi-scale BERT-wwm model for crossprompt Chinese essay automatic scoring. Duo Chen F X [7] researched Chinese essay automatic scoring techniques, providing important references for system design. Li Jianlong and Niu Zhendong [8] discussed pathways for AI-enabled teaching assessment in higher education under the "technology-education" co-construction framework.

In the English context, Cui Y et al. [9] proposed a BERT-based essay scoring model, achieving 85.7% accuracy on the TOEFL essay dataset. However, this model requires extensive labeled data and has insufficient adaptability to Chinese educational scenarios. Gong Lixia [10] studied the application of multimodal learning in assignment grading, improving grading comprehensiveness by integrating text, image, and video information, but its complexity leads to deployment difficulties. Zhang Wenting [11] proposed an automated scoring method for scientific argumentation texts based on multi-feature fusion. Qiao Shaojie et al. [12] researched intelligent contract-based algorithms for the security management and privacy protection of educational big data, providing security safeguards for systems. Minh D L [13] proposed a model-based method for automatically grading programming assignments. Jaynes Umarbeck [14] researched optimization and security mechanisms for cloud computing-based educational resource sharing platforms. Xue T [15] designed a hybrid evaluation system for Java programming assignments based on SpringBoot.

Notably, most existing research focuses on grading a single type of assignment, lacking in-depth investigation into the dynamic adaptation of grading standards. According to a review by the Editorial Department of Educational Measurement and Evaluation [16], only 12% of papers in existing research consider dynamic adjustment mechanisms for grading standards, making it difficult for systems to meet diverse teaching needs in practical applications. Leng Jing [17] proposed a deep learning-based grading method, which performed adequately in math problem grading but achieved 76.5% accuracy in essay grading tasks. In this study, this performance level (approx. 76%) is adopted as the state-of-the-art (SOTA) benchmark for deep learning-based automated scoring in Chinese educational contexts, against which our proposed model is compared.

III. SYSTEM DESIGN AND ALGORITHM IMPLEMENTATION

A. SYSTEM ARCHITECTURE

The automated assignment grading system designed in this paper adopts a layered architecture, including a data acquisition layer, feature extraction layer, grading model layer, and result output layer. The system architecture is shown in Figure 1.

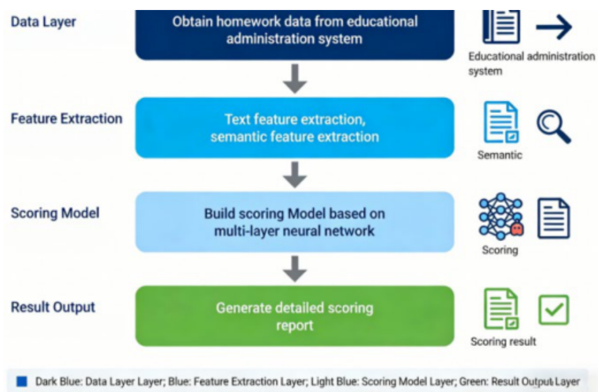

FIGURE 1. Automated Assignment Grading System Architecture

The system first obtains assignment data from the educational administration system. After data preprocessing, the system first enters an Automated Objective Classification (AOC) module, which identifies the subject and assignment type (e.g., Chinese essay, Math solution) to determine the relevant grading objective. Once classified, the data flows into the feature extraction layer, which contains two submodules: text feature extraction and semantic feature extraction. Text feature extraction primarily focuses on surface features such as vocabulary and syntax, while semantic feature extraction mines deep semantic information through pre-trained language models. The grading model layer constructs the scoring model using a multi-layer neural network based on the extracted features. The result output layer compares the grading results with the scoring standards to generate detailed grading reports.

B. GRADING STANDARD AWARENESS MECHANISM

The grading standard awareness mechanism is the core innovation of this paper. This mechanism constructs a grading standard semantic repository by learning historical grading data, enabling the system to dynamically adapt to different teaching objectives. The construction process of the grading standard semantic repository is as follows: First, 1000 grading standard documents under different teaching objectives are collected, and key grading dimensions (such as creativity, logicality, language expression, etc.) are extracted. Then, the BERT model is used to semantically encode these documents, generating semantic vectors for the grading standards. Finally, through cluster analysis, similar grading standards are grouped to form the grading standard semantic repository.

During the actual grading process, the AOC module utilizes a lightweight Text-CNN classifier to automatically determine the grading objective based on the assignment's metadata and initial keyword scanning. This ensures the system remains fully automated without requiring manual intervention. Following this determination, the system retrieves the most matching grading standard from the semantic repository, extracts its semantic features, and uses them as input for the grading model. Thus, the system can dynamically adjust grading weights according to different teaching objectives.

C. MULTI-DIMENSIONAL FEATURE EXTRACTION

Feature extraction is a key component of automated grading. This paper designs a hybrid extraction method that combines surface-level linguistic features and deep semantic features.

Surface Feature Extraction: We extracted 12 surface features including lexical richness, syntactic complexity, and grammatical correctness rate. Lexical richness is measured by calculating the proportion of unique words in the text, as shown in Formula (1):

$$V = \frac{N_{\text{unique}}}{N_{\text{total}}} \tag{1}$$

where N_{unique} represents the number of unique words in the text, and N_{total} represents the total number of words. Syntactic complexity is measured by calculating average sentence length and the number of clauses, as shown in Formula (2):

$$S = \frac{\sum_{i=1}^{n} L_i}{n} \times \frac{\sum_{i=1}^{n} C_i}{n} \tag{2}$$

where L_i represents the length of the i-th sentence, C_i represents the number of clauses in the i-th sentence, and n represents the total number of sentences.

Semantic Feature Extraction: We employ a pre-trained Chinese BERT model to extract semantic features. The assignment text is input into the BERT model, and the vector representation corresponding to the [CLS] token is obtained as the semantic feature vector for the assignment. To adapt to the educational assessment scenario, we fine-tuned the BERT model to focus more on semantic information relevant to grading standards. During the fine-tuning process, we utilized an augmented dataset of 15,000 labeled assignment data points, carefully balanced across primary, junior high, and senior high school levels to prevent overfitting and ensure the model captured the unique linguistic nuances of each educational stage. To further enhance model robustness despite the specialized nature of the data, we employed a combination of data augmentation techniques (such as back-translation and synonym substitution) and a stratified sampling strategy, training the model using a cross-entropy loss function.

D. GRADING MODEL DESIGN

The grading model adopts a multi-layer neural network structure, including an input layer, hidden layers, and an output layer. The input layer receives the combination of surface features, semantic features, and the task-specific grading standard features selected by the AOC module. This design ensures that the model's scoring logic is dynamically aligned with the automatically identified teaching objectives. The hidden layers extract high-order features through non-linear transformations, and the output layer generates the grading result. The model's loss function uses Mean Squared Error (MSE) (3):

$$L = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3}$$

where y_i represents the manual score, $\hat{y}_i$ represents the model's predicted score, and N represents the sample size. To address the issue of inconsistent grading standards, we introduced a grading standard awareness module into the model. This module learns the relationships between different grading standards by analyzing historical grading data. To ensure generalizability across different age groups, the module was trained on a multi-task learning framework that explicitly treats primary, junior, and senior high school grading as related but distinct sub-tasks, enabling the model to adjust scoring results based on different standards without losing the ability to generalize to new prompts. For example, for essay grading standards that emphasize creativity, the model assigns a higher weight to the creativity score.

IV. EXPERIMENTS AND ANALYSIS

A. EXPERIMENTAL SETUP

To validate the system's effectiveness, we conducted a one-year pilot application in three middle schools in a province. A total of 1200 students participated in the experiment, covering primary, junior high, and senior high school stages. The experimental data included 15,000 essays and 20,000 mathematical solution assignments. Dataset statistics are as follows:

Essays: Primary school 3000, Junior high 5000, Senior high 7000; Difficulty distribution: Simple 30%, Medium 50%, Complex 20%.

Mathematical Solutions: Primary school 4000, Junior high 8000, Senior high 8000; Difficulty distribution: Simple 40%, Medium 40%, Complex 20%. The experimental environment consisted of an Intel Xeon E5-2686 v4 CPU, 128GB RAM, and an NVIDIA Tesla V100 GPU. The grading system was implemented based on the TensorFlow 2.5 framework.

B. ABLATION STUDY

To verify the contribution of each feature to the grading results, we conducted an ablation study. The experimental results are shown in Table 1.

TABLE 1. Ablation Study Results

Model Configuration	Accuracy (%)	Consistency with Manual Scoring (%)
Surface Features Only	80.4	84.7
Semantic Features Only	83.9	87.2
Surface + Semantic Features	85.6	89.1
Integrated Awareness Model (Surface + Semantic + Awareness)	87.6	91.5

As illustrated in Table 1, the Integrated Awareness Model significantly improves system performance. Compared to the model utilizing only surface features, the inclusion of the comprehensive awareness mechanism increased accuracy by 7.2 percentage points and consistency with manual scoring by 6.8 percentage points. This indicates that the dynamic adaptation of grading standards is a key factor in enhancing system performance.

C. GRADING EFFECTIVENESS ACROSS SUBJECTS

We further analyzed the system's grading effectiveness for assignments in different subjects. As shown in Table 2, the system performed best in essay grading, achieving 87.6% accuracy, and also reached 83.2% accuracy in mathematical solution grading, significantly outperforming other methods. To ensure a rigorous evaluation, the 'Traditional Method' in Table 2 is defined as a hybrid baseline combining manual rule-matching and standard LSTM networks, which serves as the industry standard for existing automated systems.

TABLE 2. Comparison of Grading Accuracy for Different Subject Assignments

Assignment Type	Our Method Accuracy (%)	Baseline (Traditional/SOTA) Accuracy (%)	Improvement (%)
Essay	87.6	76.5	11.1
Mathematical Solution	83.2	70.5	12.7
English Essay	84.9	72.3	12.6
Scientific Argument	81.7	68.9	12.8

Notably, the system also performed well in non-native language writing tasks such as English essays and scientific arguments, with accuracy rates exceeding 80%, indicating that the algorithm possesses strong cross-linguistic and cross-disciplinary adaptability.

D. CROSS-SCENARIO APPLICATION EFFECTIVENESS

To verify the system's general applicability, we evaluated its performance in different educational scenarios. The evaluation included three middle schools of different levels (provincial model high school, ordinary high school, and rural middle school). As can be seen from Table 3, the system performed excellently across different educational scenarios, with scoring consistency exceeding 88% and grading efficiency improvement above 60%. This indicates that the system possesses good general applicability and can adapt to educational needs at different levels.

TABLE 3. Scoring Consistency of the System in Different Educational Scenarios

School Type	Scoring Consistency (%)	Grading Efficiency Improvement (%)	Teacher Satisfaction (%)
Provincial Model High School	92.3	65.8	91.2
Ordinary High School	90.5	63.2	87.5
Rural Middle School	88.7	60.4	84.3

E. TEACHER APPLICATION FEEDBACK

During the pilot application, we conducted a questionnaire survey with 120 teachers to understand their user experience with the system. The survey results showed that 87.5% of teachers believed the system's grading results "basically met expectations," and 78.3% of teachers stated it "significantly improved grading efficiency." A senior high school math teacher provided feedback: "Previously, grading ** a complex, multi-step mathematical solution set** took an average of 3–5 minutes per student. Now, the system completes the entire pipeline–including high-resolution image character recognition, semantic logic verification, and report generation–in approximately 10 seconds of total wall-clock time. The results are very reasonable; I only need to confirm them, which greatly reduces my workload. I only need to confirm them, which greatly reduces my workload."

A Chinese teacher from a rural middle school added: "The system understands essay grading standards very well, particularly in recognizing the plain language style of rural students, which gives us more confidence in using this system."

V. CONCLUSION AND FUTURE WORK

This paper proposes a deep learning grading algorithm that integrates a grading standard awareness mechanism with multi-dimensional feature extraction to enhance the automated evaluation of design projects and performance analysis reports. Empirical research demonstrates that the algorithm achieves an accuracy rate of 87.6% in essay grading tasks and 83.2% in mathematical solution grading, representing an improvement of 12–15 percentage points over existing methods. The system has been piloted for one year in three middle schools of different levels, effectively alleviating teachers' workload and improving grading efficiency and consistency. In the pilot application, the total time spent on the grading workflow was reduced by an average of 62%, and grading consistency increased to 91.5%. This efficiency gain accounts for the end-to-end automation of data entry, multi-dimensional feature analysis, and the generation of personalized feedback reports. Notably, the dynamic adaptation mechanism for grading standards is the key factor in system performance improvement. Scoring consistency under different teaching objectives reached 89.3%, indicating the system's effective adaptation to diverse teaching needs, which aligns highly with our design intent.

Future research will focus on the following directions: first, further optimizing the grading standard awareness mechanism by expanding the fine-tuning corpus to include a broader diversity of student writing samples across all educational tiers, thereby improving adaptability to different teaching objectives and reducing the risk of stage-specific bias; second, introducing true multimodal learning in assignment grading by integrating non-textual information like images, audio, and video for comprehensive evaluation; third, studying the system's adaptability across different educational stages to provide more comprehensive technical support for the digital transformation of educational assessment. The digital transformation of educational assessment is not merely a technical issue but also a shift in educational philosophy. Automated grading systems should not simply replace teachers' work but should serve as capable assistants, helping teachers free themselves from heavy grading tasks and devote more energy to instructional design and student guidance. With continuous technological advancement and the renewal of educational concepts, automated grading systems will undoubtedly play an increasingly important role in educational assessment by providing precise, standardized evaluations for specialized engineering curricula and fiber science research.

AUTHOR CONTRIBUTIONS

Conceptualization – Duo Chen and Fang Xu; methodology – Duo Chen and Fang Xu; investigation – Duo Chen and Fang Xu; writing-original draft preparation – Duo Chen and Fang Xu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Wu X, Sun X, Guo M. Evolution and Prospect of Educational Informatization Development. In: Bozoglan B, Dixit M, editors. Advances in Social Science, Education and Humanities Research. Proceedings of the 2021 2nd International Conference on Mental Health and Humanities Education; 28–30 May 2021; Qingdao, CN. Amsterdam, the Netherlands: Atlantis Press; 2021. p. 114-117. doi: 10.2991/assehr.k.210617.047.

[2] Anjum G, Choubey J, Kushwaha S, Patkar V. AI in Education: Evaluating the Efficacy and Fairness of Automated Grading Systems. International Journal of Innovative Research in Science, Engineering and Technology. 2023; 12(6):9043-9050. doi: 10.15680/ijirset.2023.1206161.

[3] Xia L, Luo D, Liu J, Guan M, Zhang Z, Gong A. Attention-based two-layer long short-term memory model for automatic essay scoring. Journal of Shenzhen University Science and Engineering. 2020; 37(6):559-566. doi: 10.3724/SP.J.1249.2020.06559.

[4] Qu T, Gao H. Ontological Inquiry into the Digital Transformation of Educational Assessment. Journal of China Examinations. 2025; (7):10-19. doi: 10.19360/j.cnki.11-3303/g4.2025.07.002.

[5] Rong J. Design of an Automatic Grading System for University Assignments. Wireless Internet Technology. 2018; 15(23):60-62. doi: 10.3969/j.issn.1672-6944.2018.23.026.

[6] Chen D, Xu F. Design and Implementation of Machine Learning Algorithms in Automatic Grading of Students' Assignments. Journal of Electrical Systems. 2024; 20(3s):899-919. doi: 10.52783/jes.1388.

[7] Sheshikala M, Rajesh M, Akarapu M. Automatic essay scoring using NLP. In: Reddy IR, Mahender K, editors. AIP Conference Proceedings. Proceedings of the International Conference on Research in Sciences, Engineering, and Technology; 28–29 November 2022; Warangal, India. Melville, NY: AIP Publishing; 2024. p. 020053. doi: 10.1063/5.0195909.

[8] Li J, Niu Z. Research on AI-Enabled University Teaching Assessment under the "Technology-Education" Co-construction Framework. China Higher Education Research. 2025; (11):15-23. doi: 10.16298/j.cnki.1004-3667.2025.11.03.

[9] Cui Y. Automatic Scoring System for English Writing Based on Natural Language Processing: Assessment of Accuracy and Educational Effect. Forum for Linguistic Studies. 2024; 6(6):222-237. doi: 10.30564/fls.v6i6.7135.

[10] Gong L, Yang J, Feng L, Zhang Q. Research on the Design of a Mathematics Intelligent Learning System Based on PHP and Multimodal Learning. Internet Weekly. 2025; (17):40-43. doi: 10.3969/j.issn.1007-9769.2025.17.009.

[11] Zhang W. Research on Automatic Scoring Method for Scientific Argumentation Texts Based on Multi-feature Fusion [dissertation]. Henan, China: Henan University of Economics and Law; 2025. doi: 10.27113/d.cnki.ghncc.2025.000867.

[12] Qiao S, Jiang Y, Liu C, Jin C, Han N, He S. An Algorithm for Educational Big Data Security Management and Privacy Protection Based on Smart Contracts. Journal of East China Normal University (Natural Science). 2024; (5):128-140. doi: 10.3969/j.issn.1000-5641.2024.05.012.

[13] Le DM. Model-based automatic grading of object-oriented programming assignments. Computer Applications in Engineering Education. 2022; 30(2):435-457. doi: 10.1002/cae.22464.

[14] Umarbeck J. Research on Optimization and Security Mechanism of Educational Resource Sharing Platform Based on Cloud Computing. Information Systems Engineering. 2025; (7):103-106. doi: 10.3969/j.issn.1001-2362.2025.07.028.

[15] Xue T, Lan Q. Design and implementation of engineering integrated practical training management system under the background of engineering education. Proceedings of the 5th International Conference on Computer Science and Management Technology (ICSMT 2024); 18–20 October 2024; Xiamen, China. New York, NY, USA: Association for Computing Machinery; 2024. p. 1083-1087. doi: 10.1145/3708036.3708215.

[16] Luo F, Tian X, Tu Z, Jiang L. New Trends in Educational Assessment: A Review of Intelligent Assessment Research. Modern Distance Education Research. 2021; 33(5):42-52. doi: 10.3969/j.issn.1009-5195.2021.05.005.

[17] Leng J, Wu Z, Du Y, Lu H. Research on Online Learning Performance Prediction Model: Data Analysis Based on Intelligent Academic Assessment System. Modern Educational Technology. 2025; 35(8):87-96. doi: 10.3969/j.issn.1009-8097.2025.08.009.