

AI-Driven Rubrics for Academic Grading & Feedback Using Fuzzy Clustering & Attention Networks

Rudragouda G. Patil*

Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India

Department of Computer Engineering, Marathwada Mitramandal's College of Engineering, Pune 511052, MH, India

Email: patilrudrag@gmail.com

ORCID iD: <https://orcid.org/0000-0002-1970-859X>

*Corresponding author

Mahantesh N. Birje

Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India

ORCID iD: <https://orcid.org/0000-0003-0669-1829>

Manisha Tapale

Department of Computer Science and Engineering, KLE Technological University Dr. MSSCET, Belagavi-590018, Karnataka, India

ORCID iD: <https://orcid.org/0000-0003-1291-9818>

Nagaraj V. Dharwadkar

Department of Computer Science, Central University of Karnataka (CUK), Kadaganchi, Kalaburgi-585367, Karnataka, India

ORCID iD: <https://orcid.org/0000-0003-3017-0011>

Received: September 25, 2025; Revised: December 10, 2025; Accepted: February 12, 2026; Published: August 8, 2026

Abstract: Educational assessment has shifted from a “one size fits all” model to learner-centered, responsive, and adaptable dynamic rubrics and feedback procedures. Dynamic rubrics boost engagement and assessment confidence by improving instructor-student communication and transparency. Dynamic rubrics could improve feedback and assessment. Pre-processed standard dataset texts are used. Pre-processed texts are delivered to Word2Vec to extract key features and vectorize them. A fuzzy clustering model with dynamically weighted rubrics evaluates the assignment. The dynamic rubric clearly outlines the assignment's evaluation criteria, and weights assist in deciding how much each criterion contributes to the overall grade using the Foraging Phase Updated Addax Optimization (FPUAO) algorithm. Subsequently, the assessment score is obtained, and based on this score, feedback is generated using a Generative Attention Long Short-Term Memory (GenA-LSTM) model. Finally, the developed model provides the optimal responses to the students by using the dynamic rubric with a deep learning model and an enhanced optimization algorithm. Experimental results on two datasets show an average accuracy of 99.36%, representing a 2.7% improvement over K-Means Clustering. The average Mean Absolute Error (MAE) is 31.85%, an 81% reduction compared to K-Means. Average efficiency increases by 21.30% compared to other models. Thus, the developed model is more effective and robust than the existing approaches.

Index Terms: Grade Prediction, Dynamic Rubrics, Fuzzy Clustering, Foraging Phase Updated Addax Optimization, Generative Attention Long Short Term Memory.

1. Introduction

Assessments and feedback are crucial to learning and teaching. Rubrics have traditionally provided objective, transparent, and systematic criteria for assessing student performance. A rubric usually lists criteria and defines quality levels for an assignment [1, 2, 3]. However, static rubrics cannot respond to changing student learning needs and the dynamic nature of advanced pedagogy. As a flexible and responsive feedback and assessment tool, dynamic rubrics have garnered interest in education. Dynamic rubrics adapt throughout the instructional process, unlike static ones [4, 5, 6]. Dynamic rubrics can be adjusted in real time or occasionally during a course to show updated learning objectives, curriculum standards, student progress, or contextual learning difficulties [7]. The shift from static to dynamic rubric models reflects a pedagogical shift toward student-centered, formative, and individualized learning [8].

According to Fernandez-Sanchez et al. (2025) [9], the growing popularity of dynamic rubrics is due to the importance of formative assessment, which helps improve teaching techniques and learning solutions by assessing students. They allow instructors to alter rubric criteria based on classroom assessments, student responses, and formative data to improve feedback-rich criteria [10]. With educational technologies and digital Learning Management Systems (LMS), dynamic rubrics can be effectively designed and managed, even in large classroom settings [11]. These models provide real-time updates, amended rubrics, and learner customization that paper-based methods cannot [12].

Despite their potential, the adoption of dynamic rubrics is complicated. Poor standardized models for application and development are a major issue [13]. Many instructors develop dynamic rubrics ad hoc, which can lead to inconsistencies in validating descriptive responses, confusing expectations, and compromising objectivity [14]. To ensure uniformity and fairness across student responses, descriptive or open-ended answer estimation requires rigorous criteria. In the absence of a systematic rubric to evaluate student descriptive entries for relevance, critical thinking, depth, and coherence [15], instructors may face significant challenges. Educators must continuously adjust criteria to align with new learning objectives, and students may struggle to understand the multiple evaluation standards [16]. Technology issues hinder dynamic rubric construction for descriptive answer assessments, especially in low-resource contexts. Thus, while dynamic rubrics strengthen descriptive answers, their effective use requires sufficient training, planning, and infrastructure. This study proposes a novel deep learning approach to evaluate descriptive student responses adaptively, fairly, and effectively.

The significant contributions of the proposed approach are listed below.

- To introduce a new model combining dynamic rubrics and deep learning to automate grading and feedback for descriptive student answers.
- To develop the Foraging Phase Updated Addax Optimization (FPUAO) algorithm to optimize weights for differentiated grading criteria in a dynamic rubric system. This approach updates the foraging phase of the Addax Optimization Algorithm (AOA) using a fitness-aided strategy. The developed FPUAO learns from response patterns to dynamically alter grading focus, improving fairness and accuracy. This approach ensures improved grading and reduces random weight assignments.
- A fuzzy clustering approach with weight-optimized dynamic rubrics is used to aggregate student responses based on contextual and semantic similarities. Fuzzy logic and adjusted rubric weights provide an intelligent and versatile model that adapts to different writing styles. Thus, this model efficiently predicts scores for fair assessment.
- The GenA-LSTM model is employed to generate context-aware, meaningful feedback tailored to each student's answer. This model's generative attention strategy focuses on the most important elements of the response, providing accurate, individualized feedback that gives students comprehensive advice on how to improve their responses. Thus, the GenA-LSTM helps transform educational grading.

2. Existing Work

2.1. Related Works

Kim et al. (2024) [14] used an analytic rubric to validate student assessments in Model-Eliciting Activities. This work implemented and estimated the rubric used to assess engineering students' creativity and cyber security problem-solving. Results showed that the rubric validated transparently and consistently.

Campbell and Padayachee (2023) [17] created a rubric framework for engineering mathematics assessment. Teachers trying to create rubrics that match engineering mathematics competencies and graduation factors found this work useful. A standard rubric showed how to validate operations.

Asli et al. (2024) [18] created and evaluated student self-assessment Primary Trait Writing (PTW) rubrics. Students and teachers used the Many-Facet Rasch technique to validate the rubrics' construct validity. Results showed TW rubrics effectively differentiated student writing abilities.

Chavira et al. (2022) [19] developed an Objective Structured Clinical Examination (OSCE) framework using socio-formative rubrics to assess medical students' clinical skills. The investigation showed significant performance

evaluations across exam times and a strong positive association between written exam outcomes and OSCE. The results scored the rubric-integrated model as dependable.

Cheung (2023) [20] presented process and product-based classroom assessment rubrics for online English peer and self-evaluation. The four-week intervention used these rubrics. The results showed that rubrics aid formative assessments.

A model for automatic natural language evaluation was proposed by Hashemi et al. (2024) [21]. This methodology helped answers across text quality parameters with rubric-aided prompts. The proposed methodology improved user satisfaction score prediction scores, demonstrating rubric-assisted automated evaluation.

Reference [6] presents a radiology AI measurement model. A fresh focus on local efficacy and a 7-level hierarchy validated AI value across cost, clinical, and technological phases in this approach. The approach organized AI component validation to meet stakeholder needs and improve value-aided radiology.

Bommanaboina and Bommanaboina (2024) [22] suggest using reflective analytic rubrics to improve formative assessment. A hybrid approach showed the importance of individualized training and rubric model and development reflection. Results showed how reflective rubrics improved language validation literacy.

2.2. Research Challenges and Gaps

Poorly designed dynamic rubrics might unintentionally bias learning outcomes or create inequitable conditions. Non-traditional learners, people with cognitive problems, and those inexperienced with flexible learning approaches may be disproportionately affected by frequent evaluation criterion changes. Since rubric evolution is subjective, students and instructors may interpret grading criteria differently. Thus, research is needed to determine how dynamic rubrics affect fair learning by evaluating how rubric changes affect performance and perceptions among diverse learners. This work uses deep learning to address existing challenges and constructs a novel model. The following section outlines the research gaps in existing models.

- Most conventional models rely heavily on qualitative heuristics for personalized assessment, which may compromise model robustness and generalizability. Thus, a new deep learning approach is needed. These models automatically learn hierarchical properties, allowing complex patterns in student performance to be captured.
- Current approaches are time-consuming for rubric creation and complicate the provision of precise feedback. Pre-processing addresses these issues. Pre-processing helps instructors identify key concepts, strengthening the rubric and reducing grading bias by ensuring consistency.
- Conventional systems may overlook subtle aspects of student work; therefore, designing models with an attention mechanism is necessary. By focusing on relevant features and allocating appropriate weights within the network, an attention mechanism can improve feedback.
- The lack of established research models leads to inconsistent findings regarding criteria, design, and feedback, often failing to capture student sentiment and performance status. A clustering model with weighted adaptation could increase consistency, tailor feedback, and better analyze student sentiment.

3. An Intelligent Grade Prediction and Feedback Generation System using Dynamic Rubrics and Generative Models

3.1. Proposed Framework Overview

In advanced educational contexts, notably remote and digital learning, grading open-ended or descriptive student answers is subjective and complex. Descriptive responses require coherence, clarity, topic competence, and context, unlike objective or multiple-choice assessments. Instructors traditionally use rubrics to organize and transparently grade. Standardized rubrics are effective but do not always address student diversity. They cannot adjust to changing answer complexity, learning goals, or writing styles. Disparities in grading bias and student responses may occur. Conventional automated grading models decrease manual labour, but they often rely on rule-based or shallow machine learning techniques that fail to capture the context and meaning of student writing. These models prioritize syntax, phrase length, and keyword matching over context. Consequently, feedback is often generic and uninformative, hindering student learning. The rising number of digital educational domains and the requirement for fair, timely, and constructive evaluation require flexible, intelligent, and context-based grading models. Dynamic rubrics offer a solution by updating evaluation weights and criteria based on the answer, curriculum objectives, or student performance trends. Dynamic rubrics combine adaptive methods with grading to improve feedback relevancy and score fairness, delivering an effective and personalized education. Dynamic rubrics create a new assessment and feedback structure. Figure 1 illustrates the grade prediction and feedback generation process.

Grading descriptive student answers remains challenging due to linguistic diversity and subjectivity. The goal is to develop a powerful, automated system that uses dynamic rubrics and an advanced deep learning model to grade and provide feedback. This smart grading method reduces human bias, scales well, and accurately predicts digital

descriptive answers from many students. The process begins with student responses from typical educational data sources. These responses are pre-processed to remove punctuation, special characters, redundant words, stop words, and noise. This stage standardizes and prepares the data for feature extraction. Reducing linguistic noise and focusing on answer content enhances system reliability. After pre-processing, Word2Vec creates high-dimensional vectors from text. Vectors capture semantic meanings and contextual commonalities across words, letting the technique discover student responses' important themes. Word2Vec enables the model to distinguish between correct and incomplete answers based on context, ensuring reliable evaluations. A dynamically weighted rubric approach is integrated into the fuzzy clustering process. The rubric defines evaluation criteria such as content accuracy, relevance, grammar, and coherence, assigning dynamic weights to reflect their importance. The proposed FPUAO adjusts these weights to evaluate diverse answer patterns and subjects. For fair and flexible evaluation, dynamic rubrics adjust grading criteria to question type and context. FPUAO optimises parameters for accurate and consistent grading. Grades are predicted by weighting and clustering related answers. This grade is used by the novel GenA-LSTM model to generate automatic feedback. This model's attention method highlights key student responses to provide targeted and meaningful feedback. Feedback generated by GenA-LSTM aids student learning and improvement by providing context-specific comments. Finally, insights from dynamic rubrics and deep learning deliver optimal evaluations and feedback. The system is tested and validated using established methodologies to assess performance and robustness. Comparative validation demonstrates that the proposed system improves grading fairness and feedback utility, thereby enhancing learning.

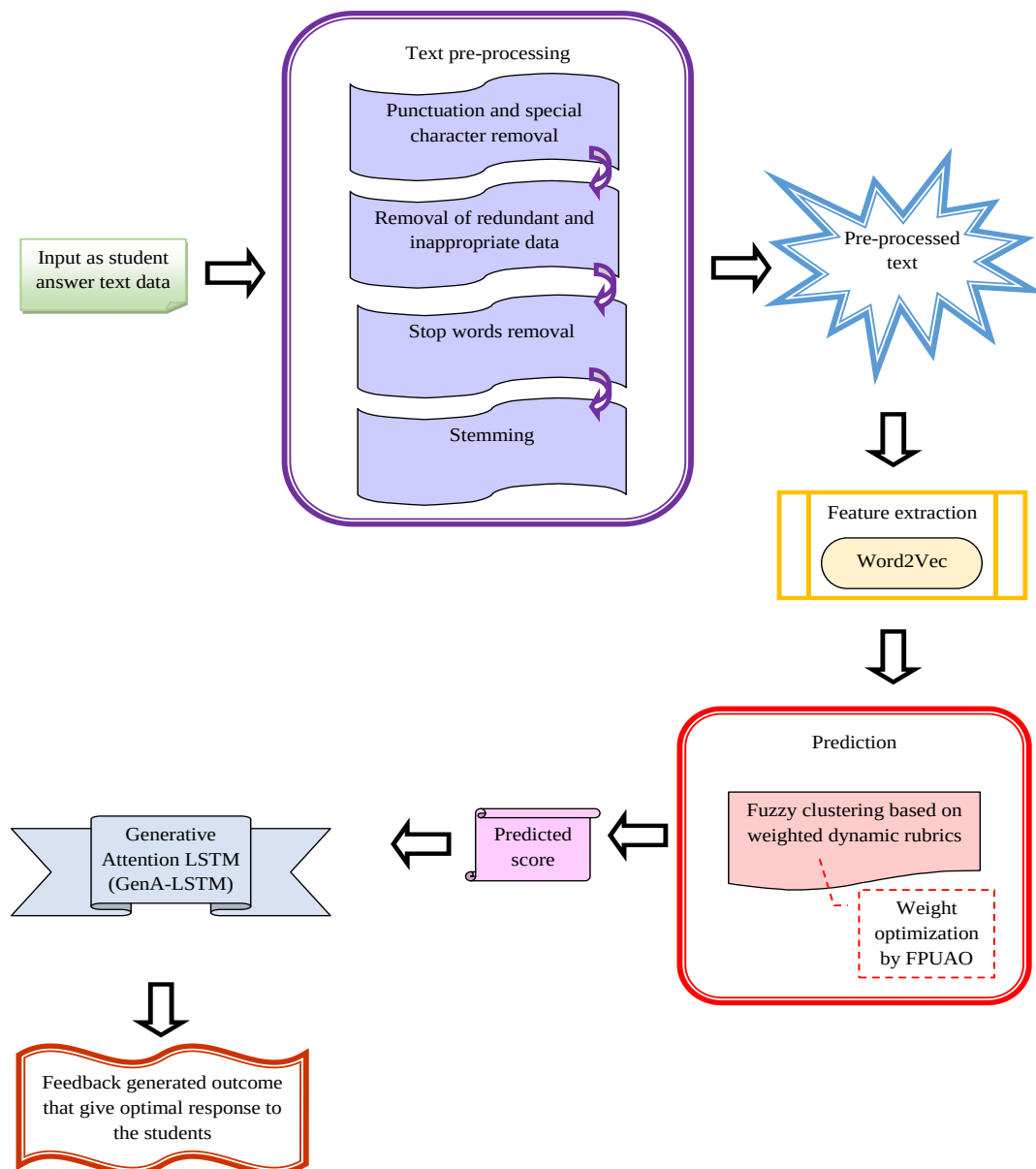


Fig. 1. A diagrammatic view of the introduced grade prediction and feedback generation system using dynamic rubrics.

3.2. Student Answer Details Dataset

The developed system utilizes the following student answer datasets.

The data source 1 is available at “https://github.com/imardini98/ASAG-Aphorisms_Spanish”. (accessed September 13, 2024). Moreover, the database for the automated brief answer ranking assignment consists of the questions related to aphorisms in the Spanish language and 3772 higher education student answers. The major dataset is named as Spanish dataset, which consists of different subsets to do the process. In such a way, the English language dataset is taken to generate the feedback against written answers.

Dataset 2 is “<https://www.kaggle.com/competitions/learning-agency-lab-automated-essay-scoring-2/data>”. Access Date: 2024-12-04. The given dataset is described as the Hewlett Foundation: Automated Essay Scoring II. It is given in three different configurations: a Microsoft Excel 2003 spreadsheet, a Tab-Separated Value (TSV) file, and a Microsoft Excel 2010 spreadsheet. Every file of this dataset contains a total of 6 columns, dataset is split into training-testing-test data (70-15-15) split. Further, the written text documents are gathered, and it is indicated $W_t, t = 1, 2, \dots, T$ respectively.

The text data garnered from these data sources is indicated as T_f , here $f = 1, 2, 3, \dots, F$. The overall count of the text data is pointed as F. Table 1 shows the sample texts gathered from the above data sources for the suggested work. Table 2 shows breakdown of 12,976 essays in “Automated Essay Scoring 2.0”, dataset including grade level and Score classes.

Table 1. Sample Texts for the Suggested Grade Prediction and Feedback Generation System.

Samples	Input data	Structured data	Feedback
Dataset 1			
1	The pond refers to the magnitude of the word, or what it is express and to generate an effect on the author should depend on what you are expressing.	Word Count - 23 Sentence Count - 1 Word Lengths - 7 Avg Word Length- 6 Punctuation Count- 1 Unique Words - 17 Unique Word Count- 17 Word Frequency- 2 Avg Sentence Length- 23 LSTM Score-2 (Grade)	The response shows an attempt to explain the concept but lacks clarity and grammatical correctness. The idea should be expressed more precisely with correct sentence structure. Improve punctuation and focus on clearly defining the intended meaning.
2	The writers are responsible for telling fragments of stories to their readers to work and feed the mind and soul, however, said messages will arrive if the reader opens his mind and this is ready to receive it.	Word Count- 36 Sentence Count- 1 Word Lengths- 9 Avg Word Length- 6 Punctuation Count- 2 Unique Words- 28 Unique Word Count- 28 Word Frequency- 2 Avg Sentence Length- 36 LSTM Score- 3 (Grade)	The response presents a meaningful idea and demonstrates conceptual understanding. However, sentence length is excessive and affects readability. Consider splitting the sentence and refining grammar to enhance clarity and flow.
Dataset 2			
1	Dear people of the @LOCATION1, I would like to inform you on the dangers of computers. Picture an empty world where people stay indoors, absorbed by screens, ignoring society. Children abandon outdoor activities, and adults neglect responsibilities, leading to chaos, obesity, antisocial behavior, and societal decline. Computers are dangerous if overused, but beneficial when used in moderation.	Word Count- 118 Sentence Count- 5 Word Lengths - 28 Avg Word Length - 6 Punctuation Count - 9 Unique Words- 82 Unique Word Count- 82 Word Frequency- 6 Avg Sentence Length - 23 LSTM Score- 4 (Grade)	The response demonstrates strong creativity and persuasive intent. The argument is meaningful and supported with examples. However, the content is overly lengthy, contains grammatical inconsistencies, and should be more concise. Improving sentence structure and reducing repetition would significantly enhance clarity and impact.
2	I agree that computers have a good effect on people because they help us learn more, explore the web, and find useful information. Computers help children learn, search topics of interest, and even calm situations through entertainment when used properly.	Word Count - 92 Sentence Count- 4 Word Lengths- 24 Avg Word Length- 5 Punctuation Count- 6 Unique Words - 61 Unique Word Count- 61 Word Frequency- 5 Avg Sentence Length - 23 LSTM Score - 3 (Grade)	The answer presents relevant ideas and a clear positive stance on computer usage. However, grammar, spelling, and sentence organization need improvement. The argument would be stronger if ideas were structured clearly and supported with concise examples.

Table 2. Summary of Automated Essay Scoring.

Prompt (Essay Set)	Approx. Essays	Target Grade Level	Essay Type	Score Range	Number of Score Classes	Class Nature
1	1,783	Grade 7–8	Narrative	2–12	11 classes	Ordinal
2	1,800	Grade 10	Persuasive	1–6	6 classes	Ordinal
3	1,726	Grade 10	Source-based	0–3	4 classes	Ordinal
4	1,772	Grade 10	Source-based	0–3	4 classes	Ordinal
5	1,805	Grade 7–8	Narrative	0–4	5 classes	Ordinal
6	1,800	Grade 7–8	Narrative	0–4	5 classes	Ordinal
7	1,569	Grade 7–8	Narrative	0–30	31 classes	Ordinal
8	723	Grade 10	Source-based	0–60	61 classes	Ordinal

4. A Novel Score Prediction using Fuzzy Clustering and Dynamic Rubrics and Feedback using Generative Attention LSTM Framework

4.1. Feature Extraction

The feature extraction operation is a significant step, as the models can't process the original text. This operation minimizes the dimensionality and also improves the model performance. This work leverages the Word2Vec word embedding technique for extracting the significant features, where the pre-processed text data is used as input. Word2Vec [23] is a powerful approach that produces word embedding's, which are dense numerical word representations in a continuous vector space. These embeddings efficiently capture the semantic relations among the words. This model leverages two-layer neural networks to understand the word embedding's by validating the large text. It is done utilizing two models. The Continuous Bag-of-Words (CBOW) forecasts the target word on the basis of context offered by its nearby words within a specific window size. The Skip-gram utilizes the target word as input and forecasts the context words that are likely to occur around it. The suggested Word2Vec model converts the high-dimensional, sparse representations into lower-dimensional, dense vectors, which helps minimize the computational complexity and makes it simple for the techniques to learn. Moreover, this model can result in strong generalization capabilities, most importantly when dealing with huge data sources. Thus, from the pre-processed text, the significant features are retrieved and the resultant features are denoted as .

4.2. Criteria's to Consider in Dynamic Rubrics

Conventionally, the rubrics have been static, fixed rubric templates developed by the educators to estimate the performance of the students on the basis of predetermined criteria. While these have supported bringing fairness and structure to grading, often they fail to accommodate the large range of student responses, context-specific nuances, or evolving curricular objectives. In contrast, the dynamic rubrics represent a transformative shift toward highly intelligent, adaptive, and student-centered assessment models. These rubrics are developed to adjust in real time or across distinct scenarios, altering the valuation parameters such as weights (criteria and the performance indicators) on the basis of distinct factors such as response complexity, historical performance data, domain-specific objectives, or student learning levels. By incorporating computational intelligence such as optimization algorithms, natural language processing, or machine learning, the dynamic rubrics make the assessment operation more scalable, personalized, and responsive. They not only enhance the consistency and fairness of the grading but also allow the delivery of contextualised feedback that can assist the students very efficiently in their learning journey. The dynamic rubrics revolution is also highly tied to the rise of digital education platforms, where the automated yet smart assessment models are in high demand because of the enhanced amount of learning and the requirement for timely validations. In addition, the dynamic rubrics empower educators with data-driven insights into the student performance trends, allowing highly informed instructional decisions. As education starts to shift toward competency-based and personalized approaches, the dynamic rubrics are developing as a next-generation assessment model's cornerstone, blending the pedagogical rigour with the technological flexibility to help deep learning and highly meaningful educational outcomes.

The dynamic rubrics estimate the descriptive student answers employing adaptable criteria rather than the constant rules. The primary criteria include relevance of the content, conceptual understanding, coherence and clarity, completeness, usage of appropriate terminology, supporting examples, critical thinking, language originality and quality. Each criterion is allocated a dynamic weight from 0% to 100% that varies based on student performance features, subject area, difficulty level, and question type. The optimization methods refine these weights to guarantee context-aware grading. The rubric is aligned with feedback creation, allowing personalized feedback by concentrating on the student's weaknesses and strengths. Figure 2 represents the criteria for dynamic rubrics in educational assessments in pictorial form.

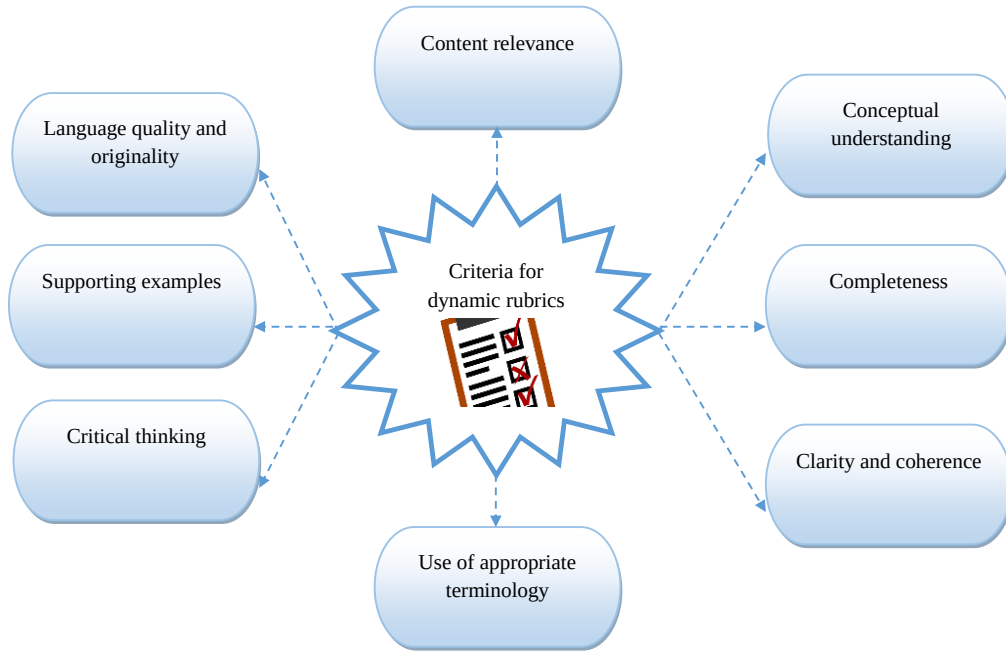


Fig. 2. Criteria for dynamic rubrics in educational assessment settings.

4.3. Recommended Foraging Phase Updated Addax Optimization (FPUAO)

The FPUAO is a newly introduced algorithm, especially for the suggested work.

Purpose: The FPUAO is an advanced optimization approach concentrated on enhancing the weight optimization operation in dynamic rubrics for estimating the student responses. In a typical grading scenario, the rubrics include distinct criteria such as creativity, correctness, and clarity, with distinct levels of importance in estimating the total score. The designed FPUAO optimizes the weights related to each criterion, guaranteeing that the grading operation is accurate and fair. Dynamically, it adjusts these weights based on the nature of the assignment and the features of the student's answer, offering a highly adaptive and personalized grading model. The optimization operation supports assigning a significant weight to each criterion, improving the grading accuracy and the feedback relevance. This enables better handling of intricate rubrics, where several criteria are subjective, such as creativity, and others are objective, such as correctness, guaranteeing that the overall evaluation aligns very closely to the quality of the student's task.

Reasons for choosing AOA: The existing AOA is selected as the foundation for FPUAO because of its demonstrated capacity to balance the exploitation and exploration in the optimization issues. The existing AOA [24] copies the addax's foraging and digging behavior; this biological inspiration makes it highly applicable for solving optimization problems with dynamic and complex search spaces. AOA is especially advantageous in the operations where robust exploration and quick convergence are required, which is significant when estimating the optimal weights for the rubrics criteria. Contrasted to other algorithms, the existing AOA is very effective in converging to global optima, making it a better choice for optimization tasks.

Novelty: Though conventional AOA is a strong algorithm, it has a limitation. In the foraging phase, it employs a random number, which can often lead to suboptimal issues, most importantly when fine-tuning is needed. This problem can cause inaccuracies in the operations, such as dynamic rubric weight assignment, where nuanced, small adjustments are significant. The FPUAO enhances this by creating a new formulation by replacing the existing random number on the basis of fitness values of the solutions rather than depending on a simple (0, 1) range. This fitness-dependent mechanism enables highly refined, informed exploration of the solution space. This operation minimizes the unnecessary exploration and concentrates more on fine-tuning the existing solution. Also, this results in better weight assignments that are very closely aligned with the rubric's true validation requirements. Eq. (1) gives the mathematical formulation of the updated foraging phase.

$$V_{b,w}^{New} = V_{b,w} + \frac{C_{fit} + B_{fit}}{C_{fit} + W_{fit}} \times (S_{b,w} - L_{b,w} \times V_{mean,w}) \quad (1)$$

Here, the b^{th} addax's newly validated position for the W^{th} dimension is $V_{b,w}^{New}$. The previous position of the b^{th} addax with a W^{th} dimension is given as $V_{b,w}$. The current, worst, and best fitness values are signified as,

C_{fit} , W_{fit} and B_{fit} . The term $S_{b,w}$ indicates that its w^{th} dimension is that particular region. The random attribute from $[0, 1]$ is taken as $L_{b,w}$ and the average estimated region of the addax in the w^{th} dimension is taken as $V_{mean,w}$.

Thus, the FPUAO is designed effectively. The novel aspect of FPUAO lies in its fitness-aided formulation, which offers highly accurate exploration of the search space. By employing the integers derived from the fitness values of current solutions, the FPUAO can fine-tune the weight optimization operation more efficiently than the existing AOA. This approach enables highly tailored optimization that captures subtle changes in rubric criteria, resulting in highly dynamic and precise weight assignments. This improvement improves the overall grading operation, making it highly adaptable to distinct types of assignments and distinct grading criteria. Hence, the FPUAO provides a highly accurate and effective solution for dynamically adjusting the rubric weights in educational settings. The pseudo-code of the recommended FPUAO is given in Algorithm 1.

Algorithm 1. Implemented FPUAO

Input: Criteria based Weight for dynamic rubrics

Output: Optimized weight for dynamic rubrics

Initialize the input details

Set maximum iterations = 50 and population size = 10

Randomly construct the initial population

Fitness function calculation

For $m=1$ to M_{it}

 For $b=1$ to F

Updated Foraging phase (Exploration)

 Calculation of the candidate's foraging region

 Randomly select the target's foraging area

 Derive the addax's new position using Eq. (1)

 Update the position of the addax

Exploitation: Digging phase

 Calculation of addax's new position

 Update the position of the addax

 End for

 Save best solutions

End for

Output the optimal weights for dynamic rubrics

4.4. Score Prediction by Fuzzy Clustering with Weight Optimized Dynamic Rubrics

The score prediction task is very important for achieving consistency when grading detailed answers written by various students. This is highly significant in educational settings, where consistency in grading is significant for guaranteeing fairness. The conventional approaches, though efficient to some extent, often encounter complexities in handling the subjective aspects of student answers such as argumentation, clarity or creativity. They tend to have poor adaptability, that can result in inconsistencies in scoring across distinct kinds of answers. Therefore, this work introduces fuzzy clustering to rectify this issue by grouping similar student responses, allowing for highly accurate grading on the basis of patterns within the data. Here, the extracted features T_f^{w2v} are given as input.

The fuzzy clustering [25] is an unsupervised learning approach that groups data points into clusters, allowing each data point to belong to distinct clusters with distinct membership degrees. The objective function of the fuzzy clustering is denoted as and derived in Eq. (2)

$$K_n(Q, W) = \sum_{j=1}^l \sum_{k=1}^m \mu_{j,k}^n \|y_k - v_j\|^2, n > 1 \quad (2)$$

Here, the term Q indicates the dataset's Y fuzzy partition constructed by the clusters $D_1, D_2, \dots, D_L$. The attribute n is indicated as weight that estimates the degree to which each datum owns to the cluster partially. This attribute also impacts the clustering outcome. The Fuzzy C-Means (FCM) approach concentrates on discovering a better partition by discovering the cluster centers w_j that reduce the objective function K_n . The FCM approach discovers the values of membership $\mu_{j,k}$ that reduce K_n . To attain both objectives of discovering a better membership and partition values, a condition is required to ensure the local minima of K_n that can be formulated in Eq. (3) by estimating the partial

derivatives and discovering where they have the value of zero. The formulation will generate the following expressions to discover the memberships as well as the cluster centres.

$$\mu_{j,k} = \frac{1}{\sum_{i=1}^l \left(\frac{\|y_k - w_j\|^2}{\|y_k - w_l\|^2} \right)^{\frac{2}{n-1}}}, 1 \leq j \leq l, 1 \leq k \leq m \quad (3)$$

According to Eq. (3), the FCM can be leveraged for updating the cluster centres as well as the membership degrees iteratively until convergence criteria are satisfied. Thus, this fuzzy clustering process efficiently manages the overlapping data and provides robustness to outliers and noise. It gives enhanced interpretability by showing the membership degree for each data point across distinct clusters. However, fuzzy clustering alone is insufficient, as it fails to account for the varying significance of distinct evaluation criteria across different types of assignments. This necessitates the dynamic rubrics, where the grading criteria can vary based on the particular assignment nature. Dynamic rubrics represent a novel assessment mechanism in education that is more adaptable, flexible, and responsive to student learning than conventional, static rubrics. They are designed to evolve based on distinct factors such as assessment context, student performance, and the type of evidence generated. The dynamic rubric is a flexible assessment tool that can be adjusted according to the particular requirements of an assessment, the learning context, and the student's progress. Unlike conventional rubrics with fixed criteria and performance levels, dynamic rubrics can evolve and adapt, often incorporating AI-driven validation of student interactions and work. While dynamic rubrics enhance system flexibility, they may fail to assign appropriate importance to each criterion without weight optimization. To address this gap, this work proposes the FPUAO. The FPUAO plays a crucial role in this work by optimizing the weight for each rubric criterion, ensuring that the grading system accurately reflects the varying significance of each aspect of the answers. Through a fitness-based formulation, the designed FPUAO refines the weights to be context-specific and highly accurate. The objective function of the dynamic rubrics weight optimization is expressed in Eq. (4).

$$O_b = \arg_{\{w^*\}} \min \left[\frac{1}{Ar} + MAE \right] \quad (4)$$

Here, the optimized weights for the dynamic rubrics are constrained within the range [0.01, 0.99]. The optimization process increases the accuracy and reduces the Mean Absolute Error (MAE).

Accuracy Ar : It validates the proportion of correctly predicted labels from the overall predictions made by the model. It is expressed in Eq. (5).

$$Ar = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (5)$$

Here, the “true positive, true negative, false positive, and false negative” values are indicated as T_p, T_n, F_p , and F_n respectively.

MAE: It quantifies the error's average magnitude in a prediction set. It is expressed in Eq. (6)

$$MAE = \frac{1}{n} \sum_{s=1}^s |a_s - \hat{a}_s| \quad (6)$$

Here, the term n is the total number of observations, and the predicted value is $\hat{a}_s$. Then, the actual value is taken as a_s .

Thus, the fuzzy clustering with weight-optimized dynamic rubrics is constructed for score prediction. This model allows for highly reliable and precise score prediction by dynamically adapting the evaluation model to the nature of the student's response. It enhances grading consistency by ensuring that similar answers receive similar grades. When employed in educational settings, this approach promotes fair and transparent grading, offering an automated yet highly accurate method for score prediction and personalized, insightful feedback. Figure 3 displays the flowchart of the implemented fuzzy clustering with weight-optimized dynamic rubrics-based score prediction process.

4.5. Generative Attention LSTM (GenA-LSTM)

The GenA-LSTM model generates contextually relevant and coherent text sequences, such as summaries or feedback, from input data. By integrating LSTM and generative attention, the model effectively handles complex sequential inputs such as natural language. The LSTM network [26] optimizes information flow using gating mechanisms and memory cells to capture temporal patterns and long-range dependencies in sequential data. It is efficient for sequential context comprehension tasks, such as evaluating student descriptive answers. Standard LSTMs may struggle with long sequences due to vanishing gradient issues. The generative attention mechanism is crucial. Generative Attention [27] enables deep learning models to focus on the most salient elements of input data while downweighting irrelevant information. The encoder layer processes predicted scores and answers into a sequence of context vectors. The attention layer weights input sections differently to highlight essential phrases, concepts, and keywords. The decoder layer generates a sequence of output (feedback), as shown in Figure 4(a). When integrated with generative attention, the LSTM model comprehends the input sequence and generates new sequences, such as tailored feedback. In education, the GenA-LSTM can evaluate students' descriptive responses, focus on key content areas, and provide constructive, tailored comments on their strengths and weaknesses. Beyond scoring and categorization, this generative feature enables the system to provide human-like feedback to enhance learning. The GenA-LSTM provides adaptive, context-aware, and meaningful feedback on student-generated content, advancing smart educational technologies.

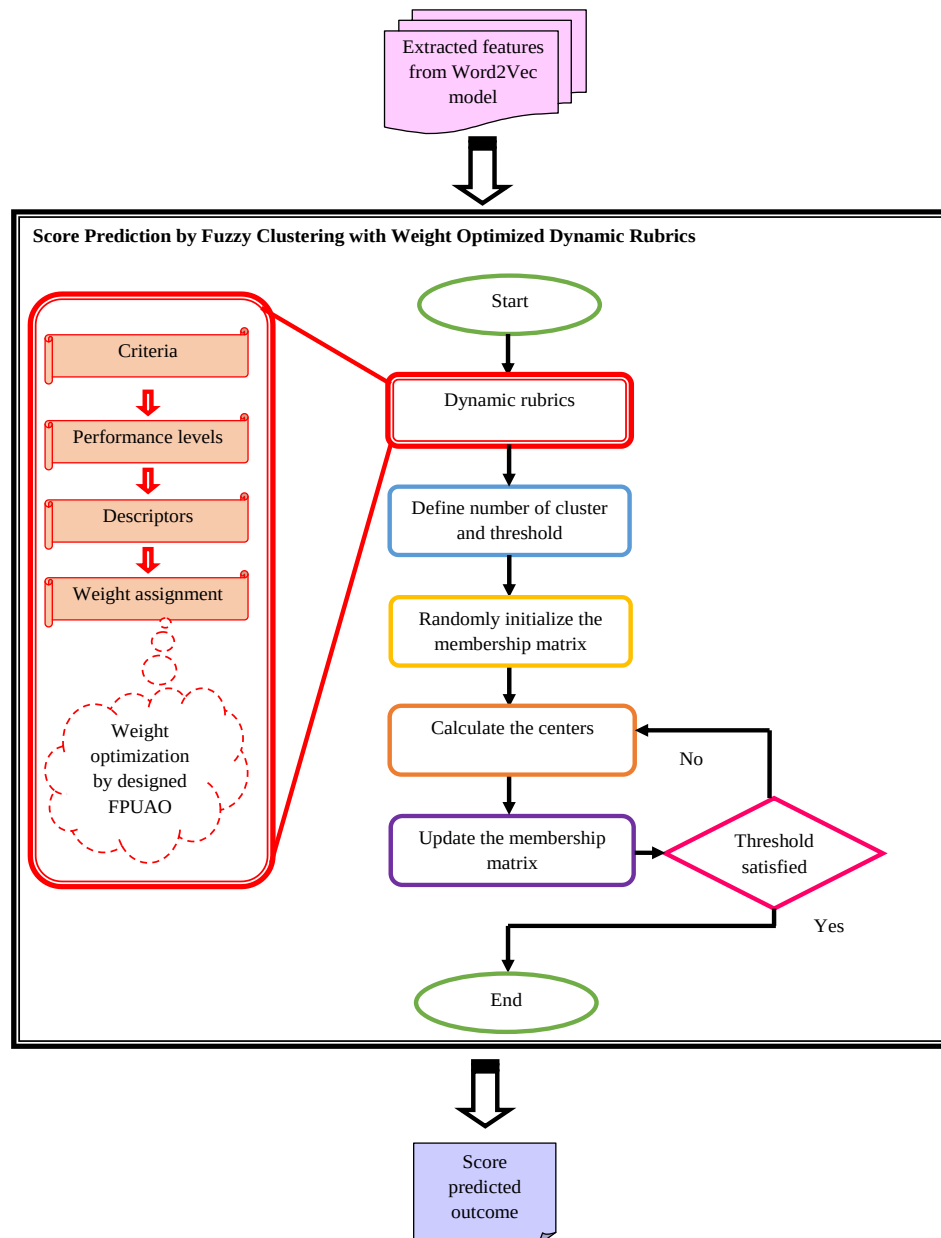


Fig. 3. Flowchart of score prediction using Fuzzy Clustering with weight-optimized dynamic rubrics.

4.6. Generative Attention LSTM for Feedback Generation

Feedback generation constitutes a significant portion of evaluation tasks, as it provides students with detailed insights into their performance. Therefore, this work employs a GenA-LSTM approach for the feedback generation. The GenA-LSTM employs a 2-layer architecture, with each layer containing 512 hidden units. The first layer receives input from 300-dimensional Word2Vec embedding units. The hidden states are passed sequentially to the next layer, with a dropout rate of 0.32 to mitigate overfitting issues. The attention layer, with a dimensionality of 256, computes alignment scores between encoder and decoder states. The context vector is concatenated with decoder hidden representations to enhance feedback quality. This model utilizes both the original student responses and the predicted scores as input. The GenA-LSTM is an advanced deep learning model designed to generate feedback by focusing on key features of student responses. The model identifies quality-determining text features, such as clarity, relevance, and question-specific argument structure. Thus, it provides useful feedback tailored to individual responses by focusing on these crucial areas. It scales to handle a large number of students without compromising relevance or quality, thereby supporting dynamic and interactive learning processes. Consequently, utilizing GenA-LSTM for feedback generation creates a holistic grading and learning experience, helping students not only understand their grades but also gain actionable insights for improvement. Figure 4(b) illustrates the step-by-step process of GenA-LSTM-based feedback generation.

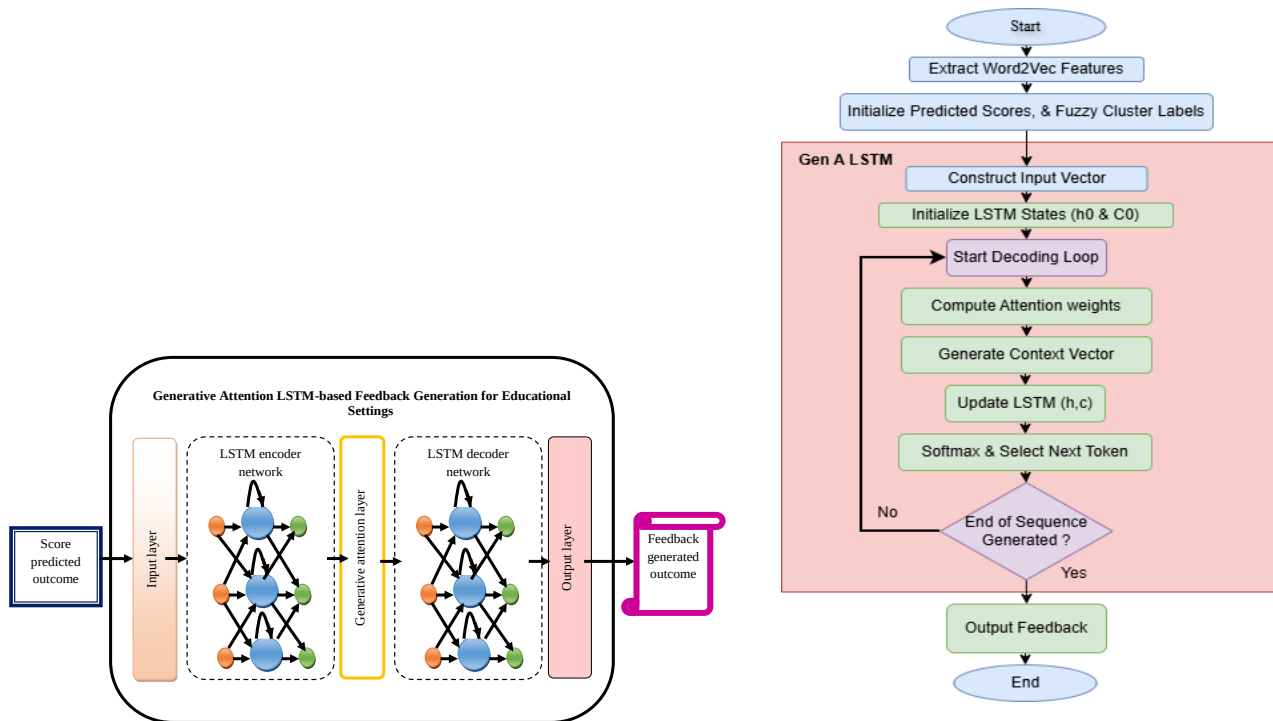


Fig. 4. (a) GenA-LSTM Encoder and Decoder Model; (b) Flowchart of GenA-LSTM-based Feedback Generation.

5. Results and Discussion

5.1. Evaluation Metrics

The following measures are taken for evaluation.

Accuracy and MAE are described in Eq. (5) and Eq. (6).

Correlation Coefficient:

$$Coeff = \frac{\sum (v_i - \bar{v})(z_i - \bar{z})}{\sqrt{\sum (v_i - \bar{v})^2 (z_i - \bar{z})^2}} \quad (7)$$

Here, the terms v_i and z_i are the values of within the v and z samples. The variables $\bar{v}$ and $\bar{z}$ denote the averages of the values of within v and z samples.

F1 score:

$$F1score = 2 \times \frac{Pre \times Rec}{Pre + Rec} \quad (8)$$

Precision:

$$Pre = \frac{T_p}{T_p + F_p} \quad (9)$$

Quadratic Weighted Kappa (QWK):

$$QWK = 1 - \frac{\sum x_{j,k} \cdot g_{ojk}}{\sum x_{j,k} \cdot g_{ejk}} \quad (10)$$

Here, the term $x_{j,k}$ indicates the weighting factor and the observed frequency is given as g_{ojk} . Also, the expected frequency is indicated as g_{ejk} .

Recall:

$$Rec = \frac{T_p}{T_p + F_n} \quad (11)$$

BERTScore: It is utilized to estimate the quality of the generated feedback.

Bilingual Evaluation Understudy (BLEU): It estimates the quality of the machine-translated text by contrasting it to more than one reference translation.

Metric for Evaluation of Translation with Explicit Ordering (METEOR): It gives a nuanced informed estimation of the created text.

Perplexity: It validates how effectively the model generates the feedback.

Sentence Error Rate (SER): It estimates the percentage of sentences that include at least one word error.

Word Error Rate (WER): It estimates the number of errors a system makes at the word level.

ROUGE-1: It estimates the overlap of individual words between the reference and generated texts.

ROUGE-2: It estimates the overlap of bigrams (pairs of adjacent words).

ROUGE-L: It estimates the length of the longest common subsequence of words appearing in the same order in both texts.

5.2. Experimental Setup

The entire framework was executed on an Intel 8th Gen i5 processor (2.5 GHz) with 16 GB RAM and an 8 GB GPU. Python was chosen as the primary implementation tool for the proposed work. Python's flexibility allows for seamless integration of distinct components. Furthermore, Python's dynamic nature aligns well with the concept of dynamic rubrics, which require real-time adjustment of assessment criteria based on student response patterns and context. Thus, Python provides the modularity, scalability, and deep learning capabilities needed to develop an adaptable, dynamic, and intelligent scoring and feedback model.

The designed FPUAO, utilizing the AOA optimizer, was configured with a population size of 10, a maximum of 50 iterations, and 40 epochs. Based on the dataset size and the number of criteria in the dynamic rubrics, the chromosome length of the designed FPUAO was set to 4. The hyperparameters included a learning rate of 0.0012, a dropout rate of 0.32, and an attention dimension of 256. For performance evaluation, several algorithms were considered, including the Crisscross Optimization Algorithm (COA) [28], Orangutan Optimization Algorithm (OOA) [29], Equilibrium Optimizer (EO) [30], and AOA [24]. Additionally, methods such as K-Means Clustering [31], Support Vector Machines (SVM) [32], Deep Neural Networks (DNN) [33], and Fuzzy Clustering [25], as well as feedback generative models like LSTM [34] and Bi-LSTM [35], were included.

5.3. Convergence Analysis

Figure 5 presents the convergence analysis of the FPUAO compared to existing methods for the two datasets. The convergence curves show that the FPUAO-Fuzzy+dynamic weighted rubrics rapidly minimize the cost function in both datasets during the early stages and steadily approach the optimal solution with minimal fluctuation, demonstrating stability and speed. At the 25th iteration of the first data source, the designed FPUAO-Fuzzy+dynamic weighted rubrics achieve a cost reduction of 66.6% compared to COA, 75% compared to OOA, 87.5% compared to EO, and 75.83% compared to AOA. These percentages show that the proposed FPUAO converges faster with fewer iterations. The faster convergence of the foraging phase updating mechanism provides better exploration and exploitation strategies in fewer

iterations. Convergence validation assures that the proposed FPUAO efficiently removes local optima and enhances convergence. Thereby improving robustness in complex optimization tasks such as dynamic rubric weight optimization.

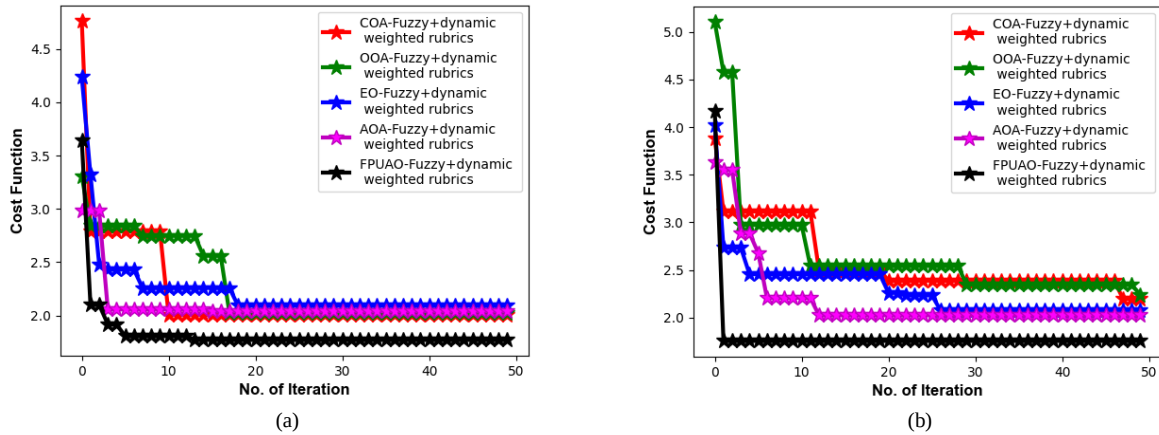


Fig. 5. Convergence examination of the designed FPUAO over traditional algorithms for two data sets a) 1 & b) 2.

5.4. Statistical Analysis

Table 3 compares FPUAO to traditional algorithms for two data sources statistically. This study uses statistical measurements across executions to estimate the FPUAO's consistency and robustness. The designed FPUAO had a smaller standard deviation than other algorithms, indicating minimal variability and excellent stability. The developed FPUAO beat conventional algorithms in both worst-case and best-case scenarios, demonstrating its capacity to find high-quality solutions. The median and mean values of the developed FPUAO are closer to the global optimum, demonstrating reliable central trends in optimization results. When considering the mean factor in the dataset, 43.33% of COA-Fuzzy+dynamic weighted rubrics, 47.22% of OOA-Fuzzy+dynamic rubrics, 27.22% of EO-Fuzzy+dynamic rubrics, and 21.11% of AOA-Fuzzy+dynamic rubrics improve the efficiency of the FPUAO-Fuzzy+dynamic rubrics. These statistical measurements show that the designed FPUAO outperforms previous algorithms by constantly providing optimal parameter choices for dynamic rubrics with minimal mistakes and high accuracy.

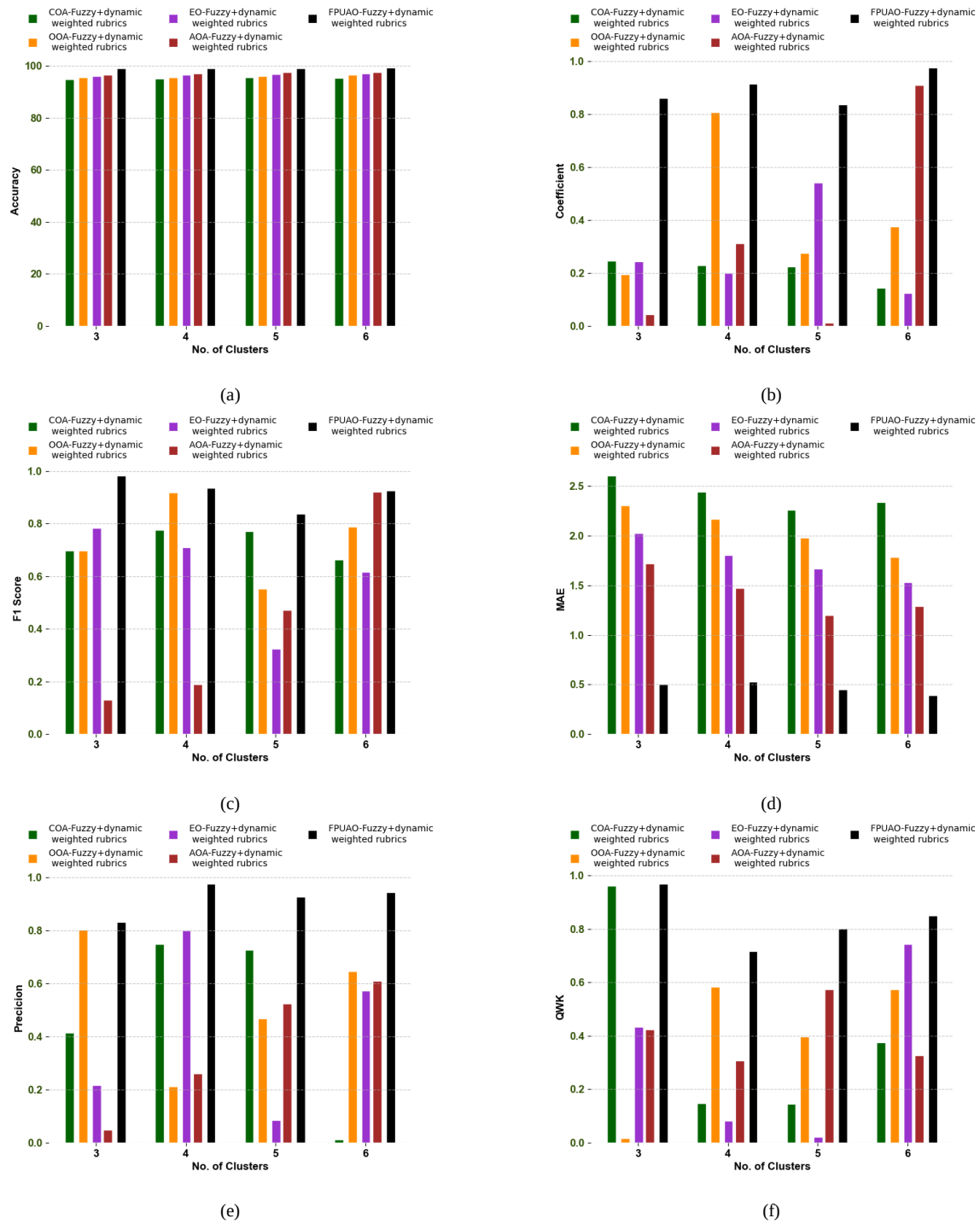
Table 3. Statistical Investigation of the Designed FPUAO over Traditional Algorithms for Two Datasets.

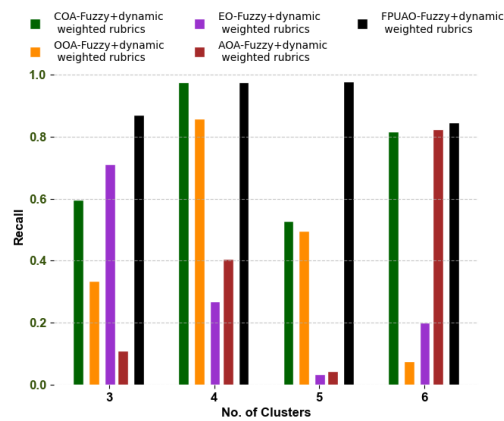
Terms	COA-Fuzzy+dynamic weighted rubrics [28]	OOA-Fuzzy+dynamic weighted rubrics [29]	EO-Fuzzy+dynamic weighted rubrics [30]	AOA-Fuzzy+dynamic weighted rubrics [24]	FPUAO-Fuzzy+dynamic weighted rubrics
Dataset 1					
Best	2.006536824	2.019189277	2.091037025	2.04018173	1.770664961
Worst	4.760706637	3.307119877	4.237499087	2.980565308	3.645495406
Mean	2.203089525	2.27888857	2.230477968	2.100731626	1.832856669
Median	2.006536824	2.024733676	2.091037025	2.04018173	1.770664961
Standard deviation	0.473563649	0.371789728	0.348557525	0.222404923	0.268051037
Dataset 2					
Best	2.197375699	2.240592081	2.075289413	2.023451333	1.754011095
Worst	3.884534007	5.104014505	4.021139528	3.635439681	4.170057977
Mean	2.58385161	2.658588663	2.295427578	2.186575434	1.802332033
Median	2.387744668	2.541884566	2.23310911	2.023451333	1.754011095
Standard deviation	0.356103238	0.575449263	0.32082691	0.401457717	0.338246563

5.5. Performance Analysis for Score Prediction

Figure 6 shows the FPUAO-Fuzzy+dynamic weighted rubrics' performance validation over algorithms and models on two datasets. This assessment uses varied cluster counts across evaluation measures. The accuracy of the designed FPUAO-Fuzzy+dynamic weighted rubrics is enhanced by 3.55% of COA, 3.04% of OOA, 2.53% of EO, and 2.03% of AOA in the 5th cluster of the first dataset. The developed FPUAO-Fuzzy+dynamic weighted rubrics model outperforms existing algorithms and models across all cluster counts, proving its durability and versatility. FPUAO supports dynamic rubric weight optimization, which ensures better performance than previous algorithms. By decreasing subjectivity and altering criteria emphasis, FPUAO-Fuzzy+dynamic weighted rubrics ensure consistent, fair evaluations.

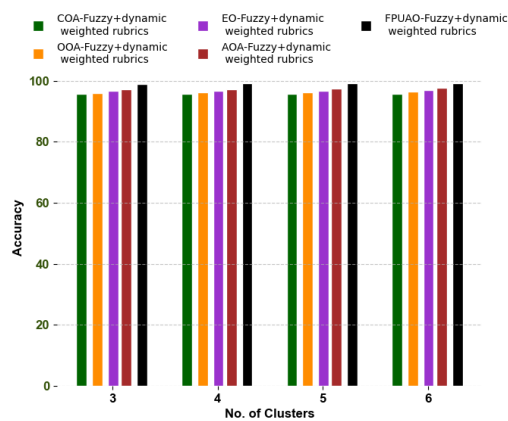
Dataset 1



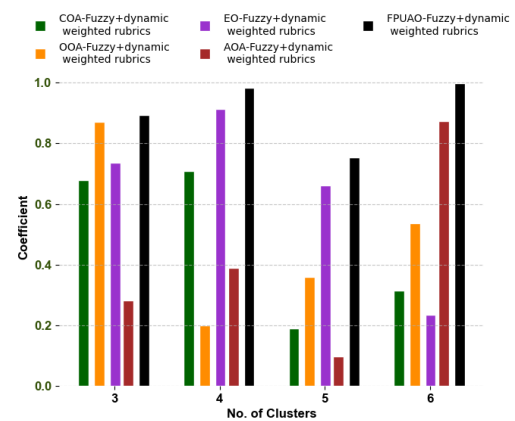


(g)

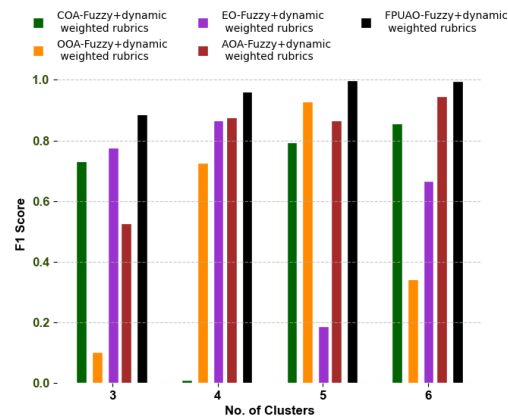
Dataset 2



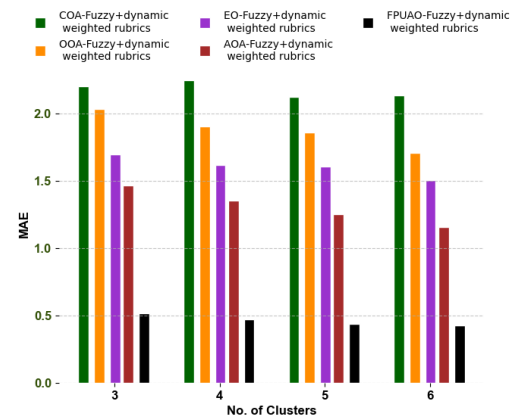
(a)



(b)



(c)



(d)

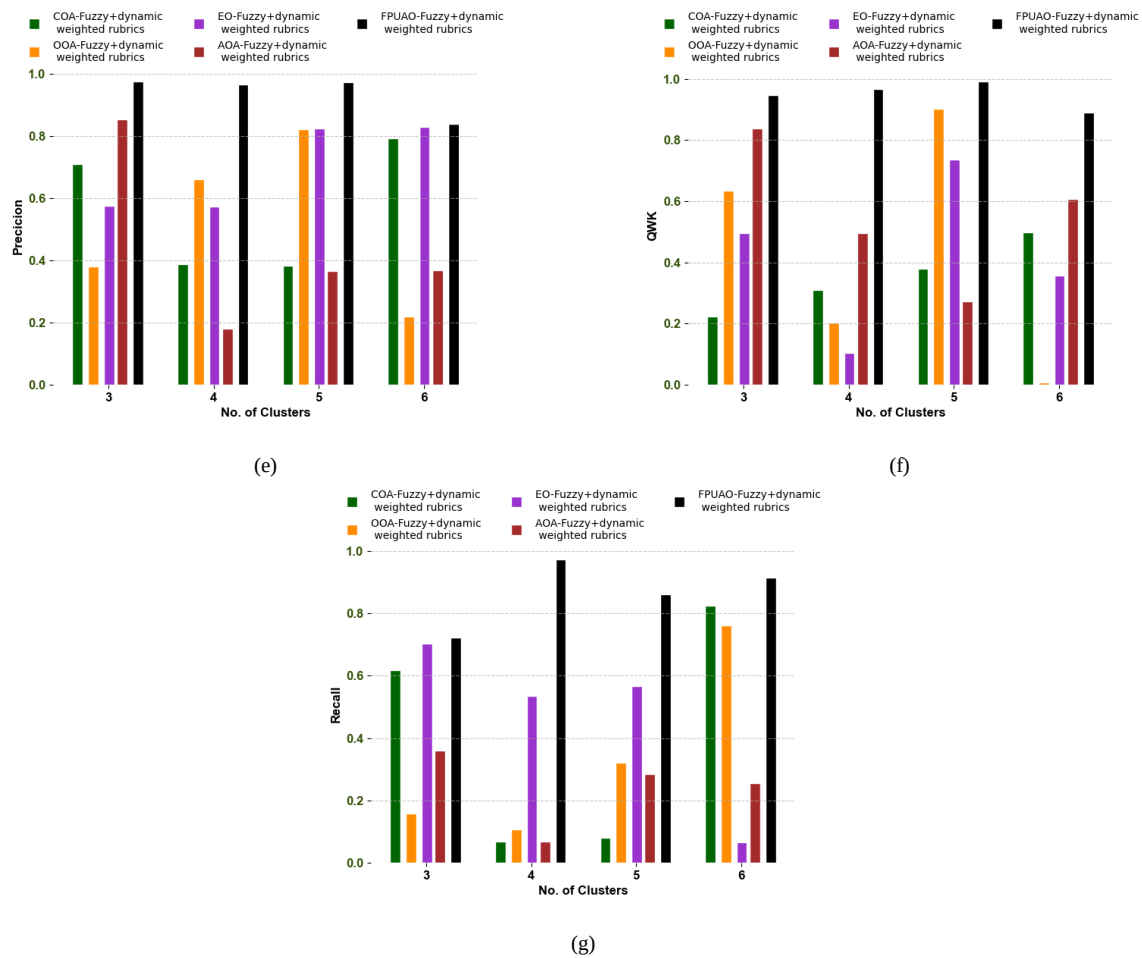


Fig. 6. Performance investigation of recommended FPUAO-Fuzzy+dynamic weighted rubrics technique-based score prediction over existing algorithms for two datasets in terms of (a) Accuracy, (b) Coefficient, (c) F1 score, (d) MAE, (e) Precision, (f) QWK and (g) Recall.

5.6. Overall Performance Analysis for Score Prediction

Table 4 represents the performance investigation of the designed FPUAO-Fuzzy+dynamic weighted rubrics approach over traditional algorithms and models for two data sources. This investigation leverages the maximum iteration values. At the 400th iteration in the second dataset, the max QWK recorded is 0.9534 indicates robust model, the designed FPUAO-Fuzzy+dynamic weighted rubric's MAE is decreased by 83.9% of K-Means Clustering, 80.6% of SVM, 76.6% of DNN, and 68.3% of Fuzzy Clustering, respectively demonstrating greater convergence stability and strong optimization. The model also shows strong agreement at 300th Iteration. This evaluation ensures that the designed FPUAO-Fuzzy+dynamic weighted rubrics benefit from enough training exposure without any overfitting issue.

Table 4. Overall Performance Investigation of the Designed FPUAO-Fuzzy+dynamic Weighted Rubrics-Based Score Prediction Approach Over Existing Algorithms for Two Datasets

Maximum iterations	K-Means Clustering [31]	SVM [32]	DNN [33]	Fuzzy Clustering [25]	FPUAO-Fuzzy+dynamic weighted rubrics
Dataset 1					
MAE					
100	1.844114528	1.664103924	1.313096501	1.087751856	0.376723224
200	1.645015907	1.490455992	1.318133616	0.983828208	0.333775186
300	1.642099682	1.451219512	1.232767762	0.936903499	0.303552492
400	1.590402969	1.371420997	1.155885472	0.904559915	0.271474019
500	1.533669141	1.368239661	1.108960764	0.933457052	0.303022269
Accuracy					
100	96.31061679	96.67380269	97.3938163	97.80536597	99.23151388
200	96.72071367	97.01995706	97.37042522	98.05152869	99.33475998
300	96.70672054	97.08171621	97.56683359	98.08704125	99.39388806
400	96.85544152	97.2690462	97.66332601	98.20705094	99.45285993
500	96.92126747	97.3235076	97.76829301	98.18467725	99.40024557
Precision					
100	0.221936816	0.114683098	0.028209375	0.418138531	0.9436009

200	0.046736671	0.123956001	0.277881811	0.922782226	0.979208866
300	0.484327161	0.345423069	0.404109298	0.135224242	0.905989138
400	0.025886199	0.239576507	0.053769829	0.696429605	0.723371302
500	0.419101874	0.386334299	0.810021663	0.356277471	0.835655694
Recall					
100	0.872764749	0.656969298	0.564734227	0.407044519	0.95529642
200	0.484890189	0.432240207	0.432772782	0.368818101	0.667692427
300	0.775472399	0.199648651	0.624347149	0.422287033	0.969039271
400	0.937289017	0.225716469	0.201607683	0.522059271	0.995190634
500	0.026591262	0.712949373	0.072177553	0.20332122	0.746315839
F1 score					
100	0.405806101	0.52171319	0.017031665	0.309677695	0.991425939
200	0.838958971	0.390817541	0.663166945	0.198400319	0.950383587
300	0.789750405	0.197413312	0.349469688	0.327945993	0.997043727
400	0.270583612	0.533378536	0.093192641	0.531992999	0.971583942
500	0.858799057	0.555168932	0.538160178	0.551878475	0.949960492
QWK					
100	0.475014903	0.748426426	0.031067944	0.172547146	0.928695734
200	0.38552133	0.867375369	0.918653825	0.877317932	0.930903862
300	0.069066675	0.473977783	0.322918641	0.481849046	0.98924642
400	0.419466506	0.844923738	0.644125837	0.374040803	0.955099727
500	0.952589601	0.787178154	0.251698343	0.725379045	0.953307905
Coefficient					
100	0.32276788	0.691261112	0.800438901	0.878783076	0.907128567
200	0.304361818	0.454529693	0.573030837	0.530634791	0.604494728
300	0.796354497	0.426517586	0.248885902	0.294822067	0.900438738
400	0.125212165	0.956082084	0.88931743	0.823932445	0.984474605
500	0.074265157	0.303514278	0.145112115	0.167605896	0.873867776
Dataset 2					
MAE					
100	1.743902439	1.529957582	1.295068929	1.13971368	0.330593849
200	1.772799576	1.518557794	1.278632025	1.028366914	0.370625663
300	1.738600212	1.373806999	1.214209968	0.909862142	0.311240721
400	1.66542948	1.380434783	1.142895016	0.845970308	0.268027572
500	1.642895016	1.311770944	1.133351007	0.88388123	0.315217391
Accuracy					
100	96.49992817	96.9287672	97.38915256	97.69405303	99.32734942
200	96.48627772	96.95990234	97.42967992	97.93273253	99.2733019
300	96.55377178	97.23877018	97.57235238	98.17873063	99.3728867
400	96.66611827	97.24541078	97.72740452	98.28995247	99.4633758
500	96.74351146	97.34988672	97.73526732	98.2356521	99.37587595
Precision					
100	0.340610677	0.842703294	0.097957159	0.754565563	0.905694641
200	0.198856018	0.592898916	0.308886326	0.736022483	0.982930895
300	0.04207001	0.941633923	0.224928368	0.011200722	0.971722926
400	0.53703014	0.712081738	0.576690937	0.422231262	0.788944503
500	0.174421557	0.362695979	0.058497752	0.450216209	0.948636265
Recall					
100	0.104358491	0.072618317	0.245212036	0.401259399	0.994080033
200	0.226947179	0.707979151	0.161027209	0.371208096	0.846214401
300	0.821441629	0.359034244	0.71103141	0.630629445	0.931745242
400	0.155978998	0.073806108	0.615461655	0.293286778	0.915302763
500	0.085938879	0.352625108	0.643471177	0.270739559	0.961908543
F1 score					
100	0.240512783	0.729856723	0.659992086	0.506958279	0.982523698
200	0.084404279	0.110921679	0.231723137	0.856184596	0.864145678
300	0.42183208	0.308361755	0.078176895	0.719954829	0.97828015
400	0.167665158	0.374072168	0.420237665	0.618710683	0.831179393
500	0.757057663	0.581653734	0.587796165	0.137782589	0.933752585
QWK					
100	0.224744055	0.048581883	0.095183409	0.727452369	0.752901217
200	0.380044011	0.428140563	0.087714729	0.118262866	0.893520377
300	0.381174836	0.200985593	0.570231353	0.327552952	0.9727974
400	0.36049928	0.541224328	0.444385427	0.545369832	0.953439158
500	0.005157162	0.47330749	0.375332116	0.115206959	0.723179391
Coefficient					
100	0.667647564	0.033454857	0.158042457	0.066554998	0.912586108
200	0.497420006	0.95805868	0.118477784	0.546033875	0.998273689
300	0.756282213	0.11726651	0.472875297	0.251161656	0.803016796
400	0.476801579	0.260712601	0.824130387	0.777219614	0.919458925
500	0.605491355	0.196439898	0.79366345	0.073228753	0.982786382

5.7. Threshold Analysis for Score Prediction

Figure 7 displays the threshold investigation of the designed FPUAO-Fuzzy+dynamic weighted rubrics for two data sources. The investigation in Fig.7 leverages the existing AOA-Fuzzy+dynamic weighted rubrics for validating the designed FPUAO-Fuzzy+dynamic weighted rubrics. When considering these existing models, the designed model yielded the best balance among accuracy and error minimization. The designed FPUAO's dynamic adjustment of the dynamic rubric weights synergizes efficiently with mid-range learning rates, allowing the technique to effectively adapt during the training while maintaining the stability of the performance. Contrasted to the traditional Fuzzy Clustering, the FPUAO-Fuzzy+dynamic weighted rubrics achieves 18.7% lower MAE rates in the first dataset for a 0.88 threshold rate. Thus, this investigation demonstrated the superior accuracy and adaptability across all tested threshold rates, outperforming the traditional optimization algorithms and models in the grading process.

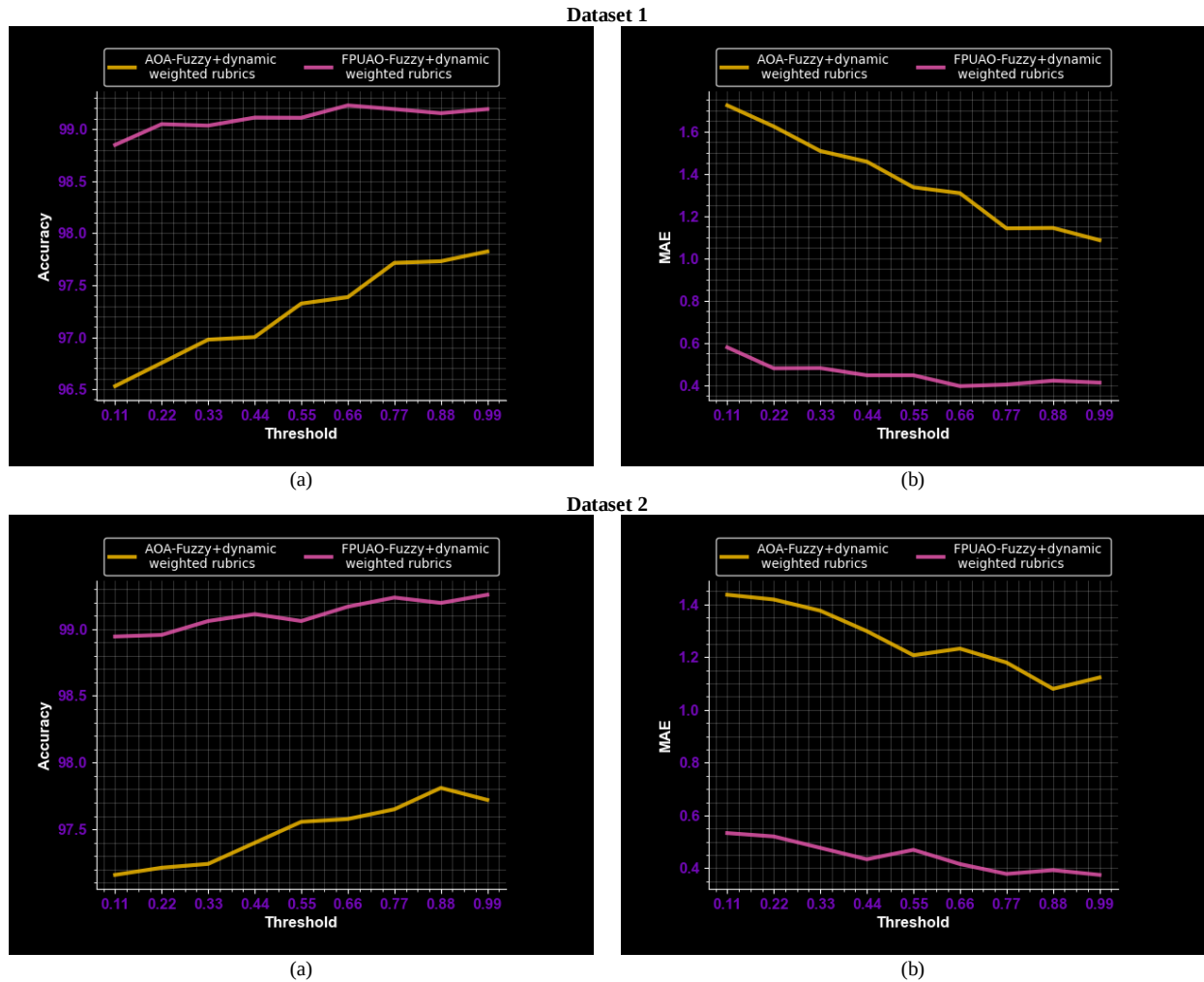


Fig. 7. Threshold-based performance investigation of recommended FPUAO-Fuzzy+dynamic weighted rubrics technique-based score prediction over existing algorithms on two datasets in terms of (a) Accuracy and (b) MAE

5.8. Performance Analysis for Feedback Generation

Figure 8 depicts the performance investigation of the recommended GenA-LSTM technique over traditional models utilizing distinct activation functions. This investigation considers various measures in the feedback generation. The activation functions generated highly favorable solutions for the designed GenA-LSTM than the other techniques with its embedded attention method, particularly in creating grammatically accurate and contextually rich feedback. The average performance of GenA-LSTM improved by 2.05% over LSTM and by 4.5% over Bi-LSTM. The highest average BERTScore of 0.87 for GenA-LSTM surpassed that of LSTM and Bi-LSTM. A BERTScore of 0.89 was achieved by GenA-LSTM with ReLU activation, being the best. The SER of the recommended GenA-LSTM is reduced by 14% compared to LSTM and 4% compared to Bi-LSTM, respectively, for the ReLU activation function in the second dataset, as shown in Figure 8. This superior performance is attributed to the GenA-LSTM's capacity to concentrate on the semantically significant areas of input sequences, guaranteeing precision in both language generation and interpretation, thus improving the quality of the educational feedback given to the students.

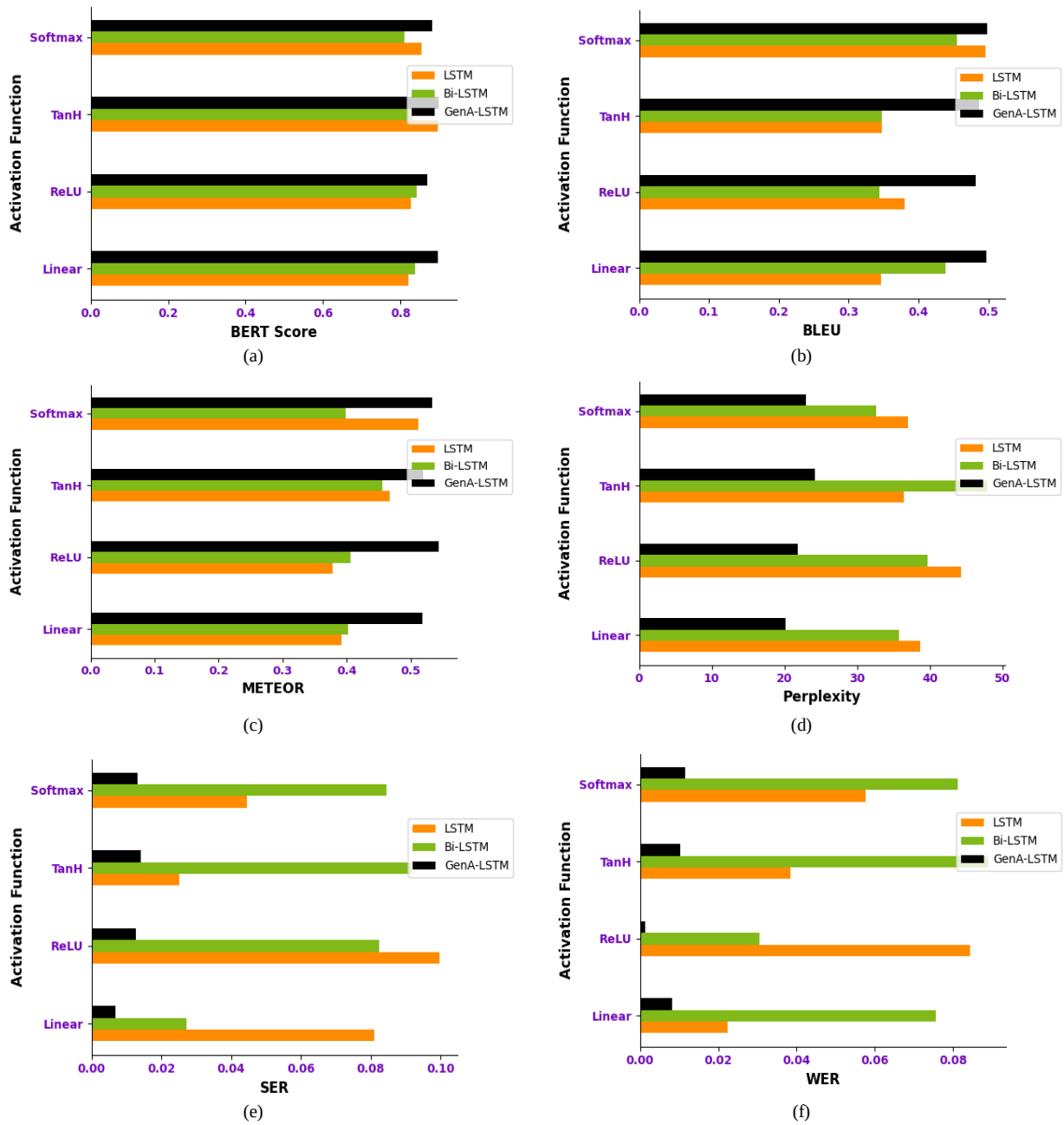


Fig. 8. Performance investigation of recommended GenA-LSTM-based feedback generation over existing models for the 2nd dataset in terms of (a) BERT score, (b) BLEU, (c) METEOR, (d) Perplexity, (e) SER, and (f) WER

5.9. Rouge -based Performance Analysis for Feedback Generation

Figure 9 illustrates the performance evaluation of the designed GenA-LSTM approach by leveraging the rouge scores. The rouge scores are mostly utilized for validating the similarity of the generated feedback and the reference text. These measures are employed for validating the quality of the produced feedback. When considering ROUGE-2 in the second dataset, the designed GenA-LSTM's precision is enhanced by 41.66% compared to LSTM and 35% compared to Bi-LSTM, respectively. Hence, it is ensured that the recommended GenA-LSTM technique attained accurate feedback than the other approaches.

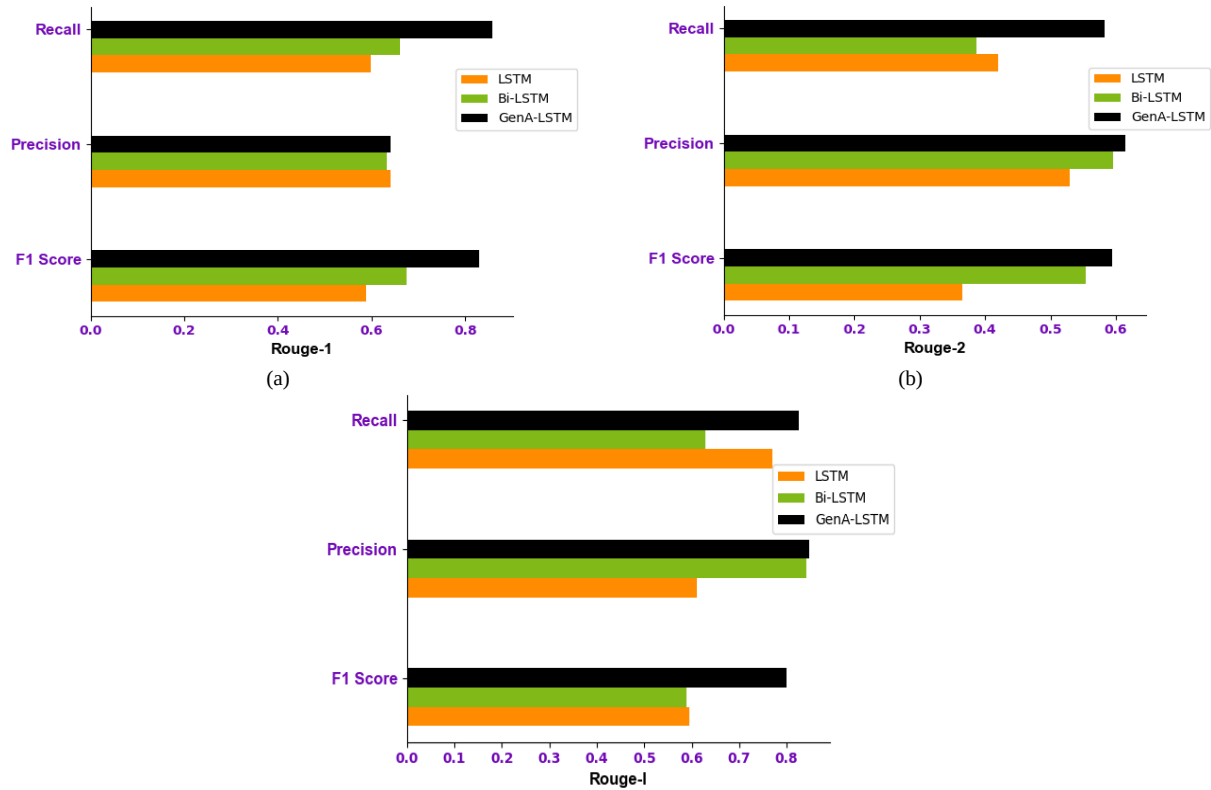


Fig. 9. Rouge-based performance investigation of recommended GenA-LSTM-based feedback generation over existing models for 2nd dataset in terms of (a) Rouge 1, (b) Rouge 2, and (c) ROUGE-L.

5.10. Computational Time & Complexity Analysis

Table 5 and Table 6 give the computational time and complexity validations of the designed framework. This investigation properly analyzes the computational efficiency of the designed FPUAO-Fuzzy+dynamic weighted rubrics. Here, the number of population size is N_{po} , and the chromosome length is given as C_{le} . Also, the maximum iteration is specified as M_{itr} . From the investigations, it is showcased that the implemented FPUAO-Fuzzy+dynamic weighted rubrics require a much smaller amount of time for the computation (13 seconds) than the existing methods. In addition, the designed FPUAO-Fuzzy+dynamic weighted rubrics elucidate that its computational complexity is much lower than the previous algorithms. Hence, it is proved that the designed FPUAO-Fuzzy+dynamic weighted rubrics are computationally more efficient than the existing methods.

Table 5. Computational Time Analysis of the Designed FPUAO-Fuzzy+Dynamic Weighted Rubrics.

Terms	Computational time (Seconds)
Based on Algorithms	
COA-Fuzzy+dynamic weighted rubrics [28]	17.9601
OOA-Fuzzy+dynamic weighted rubrics [29]	16.3654
EO-Fuzzy+dynamic weighted rubrics [30]	18.9854
AOA-Fuzzy+dynamic weighted rubrics [24]	15.8890
FPUAO-Fuzzy+dynamic weighted rubrics	13.002680
Based on Models	
K-Means Clustering [31]	17.8136
SVM [32]	16.3695
DNN [33]	18.1584
Fuzzy Clustering [25]	14.3268
FPUAO-Fuzzy+dynamic weighted rubrics	13.002680

Table 6. Computational Complexity Analysis of the Designed FPUAO-Fuzzy+Dynamic Weighted Rubrics.

Algorithms	Computational complexity
COA-Fuzzy+dynamic weighted rubrics [28]	$O[M_{itr} + 1 + N_{po} + 1 + C_{le} + 3]$
OOA-Fuzzy+dynamic weighted rubrics [29]	$O[M_{itr} + N_{po} + 2 + C_{le}]$
EO-Fuzzy+dynamic weighted rubrics [30]	$O[M_{itr} + 1 + N_{po} + 2 + C_{le} + 3]$
AOA-Fuzzy+dynamic weighted rubrics [24]	$O[M_{itr} + 1 + N_{po} + C_{le} + 2]$
FPUAO-Fuzzy+dynamic weighted rubrics	$O[M_{itr} + N_{po} + C_{le}]$

5.11. Ablation Study

Table 7 presents the ablation study of the designed GenA-LSTM. This study mainly helps to analyze the components of the designed algorithm. This study eliminates the parts of the model to understand the contribution of each model component. When analyzing the GenA-LSTM, it achieves a 95.08% accuracy compared to other LSTM-based methods. Hence, it is ensured that the designed GenA-LSTM provides highly accurate feedback.

Table 7. Ablation Study.

Term	LSTM	Attention LSTM	Generative LSTM	Proposed GenA-LSTM
Accuracy (%)	92.1285	93.0281	94.0698	95.0886

6. Conclusion

This work has designed an efficient framework for providing grades and feedback to the student responses by employing deep learning and a dynamic rubric approach. At first, the necessary student response texts were garnered from the existing and real-time data sources. Further, these texts were provided to the pre-processing stage for eliminating the unwanted contents from the raw text. Further, the obtained pre-processed texts were given to the Word2Vec approach for retrieving the relevant features. After that, the extracted features were subjected to the fuzzy clustering approach, powered by dynamically weighted rubrics, where the dynamic rubrics detected the particular criteria that were utilized to estimate the assignment, and provided weights to estimate how much each criterion contributed to the entire grade. Here, the weight was optimized by FPUAO. Further, the assessment score was achieved, and according to this score, the feedback generation was done by the designed GenA-LSTM. Lastly, the performance of the designed framework was assessed with other traditional approaches. When focusing on the 6th cluster in the first dataset, the designed FPUAO-Fuzzy+dynamic weighted rubrics' accuracy was enhanced by 4.04% over K-Means Clustering, 3.53% over SVM, 3.03% over DNN, and 2.02% over Fuzzy Clustering, respectively. Therefore, the designed framework was effective and powerful in predicting scores and the feedback generation process, ensuring consistent and accurate evaluations in the educational settings.

Limitations and Future works: Although the innovation and efficacy introduced by the designed FPUAO-based dynamic rubric framework integrated with GenA-LSTM, some problems still remain. One primary problem lies in the reliance on pre-trained word embeddings such as Word2Vec. Though it is efficient in capturing semantic similarities, it may not entirely grasp the evolving linguistic patterns or domain-specific terminologies in the educational content. Therefore, in future work, contextualized learning approaches like BERT will be integrated for better handling of domain-specific variations and richer semantic understanding. In addition, including the reinforcement learning for the adjustment of dynamic rubrics based on teacher feedback will be refined in future work for grading adaptability as well as accuracy.

All the Declarations and Statements

Author Contributions Statement

Rudragouda G. Patil – Ideation and conceptualization of the research, developed the framework and model, performed data processing, training, validation, analysis, and drafted the manuscript.

Mahantesh N. Birje – Guided the methodology, supervised the project, and reviewed and revised the manuscript.

Manisha Tapale – Assisted with validation, formal analysis, and technical review.

Nagaraj V. Dharwadkar – Supported visualization, evaluation, and manuscript refinement.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

No funding received for this research.

Data Availability Statement

This study analyzed publicly available datasets.

Dataset1: https://github.com/imardini98/ASAG-Aphorisms_Spanish

Dataset2: <https://www.kaggle.com/competitions/learning-agency-lab-automated-essay-scoring-2/data>

Ethical Declarations

This article is original and contains unpublished material. The authors confirm that no ethical issues are involved.

Acknowledgments

The authors thank all the experts for their valuable suggestions, which improved the quality of this study.

Declaration of Generative AI in Scholarly Writing

No generative writing tools were used in the preparation of this manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

AI - Artificial Intelligence
AOA - Addax Optimization Algorithm
BERT - Bidirectional Encoder Representations from Transformers
BLEU - Bilingual Evaluation Understudy
COA - Crisscross Optimization Algorithm
DNN - Deep Neural Networks
EO - Equilibrium Optimizer
FCM - Fuzzy C Means
FPUAO - Foraging Phase Updated Addax Optimization
GenA LSTM - Generative Attention Long Short Term Memory
GPU - Graphical Processing Unit
LSTM - Long Short Term Memory
MAE - Mean Absolute Error
METEOR - Metric for Evaluation of Translation with Explicit Ordering
OOA - Orangutan Optimization Algorithm
QWK - Quadratic Weighted Kappa
ReLU - Rectified Linear Unit
RNN - Recurrent Neural Network
SER - Sentence Error Rate
SVM - Support Vector Machines
WER - Word Error Rate

Appendix

None.

References

- [1] M. E. Poehner and L. Yu, "Dynamic assessment of L2 writing: Exploring the potential of rubrics as mediation in diagnosing learners' emerging abilities," *TESOL Quarterly*, vol. 56, no. 4, pp. 1191–1217, 2022.
- [2] C. Lazou, A. Tsinakos, and I. Kazanidis, "A rubric for peer evaluation of multi-user virtual environments for education and training," *Information*, vol. 16, no. 3, p. 174, 2025.
- [3] A. Rezai, A. Naserpour, and S. Rahimi, "Online peer-dynamic assessment: An approach to boosting Iranian high school students' writing skills: A mixed-methods study," *Interactive Learning Environments*, vol. 32, no. 1, pp. 306–324, 2024.
- [4] M. Chomintra, "Critical race digital literacy rubric for assessing mis/disinformation literacy instruction lesson plans," *The Journal of Academic Librarianship*, vol. 50, no. 3, p. 102868, 2024.
- [5] L. Zheng, T. Liu, Y. Feng, X. Gu, and M.-H. Yu, "Dynamic teachers' technology adoption during the COVID-19 pandemic," *SAGE Open*, vol. 14, no. 2, 2024.
- [6] B.-J. Boverhof et al., "Radiology AI deployment and assessment rubric (RADAR) to bring value-based AI into radiological practice," *Insights into Imaging*, vol. 15, no. 1, p. 34, 2024.

- [7] P. Boyd and J. Hill, "Working with rubrics: Codification plus dialogue in developing shared academic standards," in *Academic Standards in Higher Education*, R. Land and J. Gordon, Eds., Routledge, 2024, pp. 169–185.
- [8] J. S. Chavis, J. G. Osborn, D. P. Syed, and R. Terrell, "Architecting an enterprise-of-enterprises with the 10-layer rubric: Transforming EoE decision-making," in *2023 IEEE International Systems Conference (SysCon)*, IEEE, 2023, pp. 1–8.
- [9] A. Fernandez-Sanchez, J. J. Lorenzo-Castineiras, and A. Sanchez-Bello, "Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics," *European Journal of Education*, vol. 60, no. 1, p. e12826, 2025.
- [10] R. Rockwell, "Navigating transition: Developing an instructional coaching rubric and UDL-infused training for seasoned reading and dyslexia teachers in online education," *International Journal on Studies in Education (IJonSE)*, vol. 6, no. 3, 2024.
- [11] B. D. Garber, "A structured rubric for evaluating the many systemic variables that can contribute to parent–child contact problems (PCCP)," *Family Court Review*, vol. 62, no. 2, pp. 343–358, 2024.
- [12] S. Thite, J. Ravishankar, I. Tomeo-Reyes, and A. M. Ortiz, "Design of a simple rubric to peer-evaluate the teamwork skills of engineering students," *European Journal of Engineering Education*, vol. 49, no. 4, pp. 623–646, 2024.
- [13] I. Haight, "A rubric for analyzing legal technology using benefit/risk pairs," *University of St. Thomas Law Journal*, vol. 20, pp. 107–132, 2024.
- [14] Y. R. Kim, J. Yang, Y. Lee, and B. Earwood, "Assessing cybersecurity problem-solving skills and creativity of engineering students through model-eliciting activities using an analytic rubric," *IEEE Access*, vol. 12, pp. 5743–5759, 2024.
- [15] D. P. Drucker et al., "WorldFAIR (D10.3) agricultural biodiversity FAIR data assessment rubrics [Research report]," 2024.
- [16] T. T. Mir, S. A. Khan, and H. Zahra, "Preparing future educators: Assessment literacy and rubric use among prospective teachers," *Journal of Social Sciences Development*, vol. 3, no. 3, pp. 179–191, 2024.
- [17] A. L. Campbell and P. Padayachee, "Design principles for using rubrics in engineering mathematics," *IEEE Transactions on Education*, vol. 67, no. 4, pp. 499–507, 2023.
- [18] N. F. Asli, M. E. E. M. Matore, and M. M. Yunus, "Construct validity of primary trait writing rubrics based on assessment use argument (AUA) validation framework," *Heliyon*, vol. 10, no. 22, p. e40053, 2024.
- [19] J. A. P. Chavira, H. P. Acosta, A. G. Lopez, J. L. Loya, C. R. C. Sanchez, and A. K. N. Acosta, "The socioformative rubrics in the OSCE to assess the level of achievement of the competencies comprising the profile of the physician graduate," *Educacion Medica*, vol. 23, no. 3, p. 100740, 2022.
- [20] A. Cheung, "Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes," *Assessing Writing*, vol. 56, p. 100723, 2023.
- [21] H. Hashemi, J. Eisner, C. Rosset, B. Van Durme, and C. Kedzie, "LLM-rubric: A multidimensional, calibrated approach to automated evaluation of natural language texts. arXiv," 2024.
- [22] L. D. Bommanaboina and R. D. Bommanaboina, "Enhancing ESL preservice teachers' formative assessment skills: Developing reflective analytic rubrics for language assessment tasks," *The International Journal of Humanities Education*, vol. 23, no. 1, pp. 69–102, 2024.
- [23] M. Rahman, M. R. A. Talukder, L. A. Setu, and A. K. Das, "A dynamic strategy for classifying sentiment from Bengali text by utilizing Word2Vec model," *Journal of Information Technology Research*, vol. 15, pp. 1–17, 2022.
- [24] T. Hamadneh et al., "Addax optimization algorithm: A novel nature-inspired optimizer for solving engineering applications," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 3, 2024.
- [25] F. Valdez, O. Castillo, and P. Melin, "Bio-inspired algorithms and their applications for optimization in fuzzy clustering," *Algorithms*, vol. 14, no. 122, 2021.
- [26] Y.-S. Jeong and N.-W. Cho, "Evaluation of e-learners' concentration using recurrent neural networks," *J. Supercomput.*, vol. 79, no. 4, pp. 4146–4163, Mar. 2023, doi: 10.1007/s11227-022-04804-w.
- [27] H. Matsuo et al., "Unsupervised-learning-based method for chest MRI–CT transformation using structure constrained unsupervised generative attention networks," *Scientific Reports*, vol. 12, p. 11090, 2022.
- [28] A. Meng et al., "Solving large-scale combined heat and power economic dispatch problems by using deep reinforcement learning based crisscross optimization algorithm," *Applied Thermal Engineering*, vol. 245, p. 122781, 2024.
- [29] T. Hamadneh et al., "Orangutan optimization algorithm: An innovative bio-inspired metaheuristic approach for solving engineering optimization problems," *International Journal of Intelligent Engineering and Systems*, vol. 18, pp. 45–58, 2025.
- [30] J. Liu, W. Li, and Y. Li, "LWMEQ: An efficient equilibrium optimizer for complex functions and engineering design problems," *Expert Systems with Applications*, vol. 198, p. 116828, 2022.
- [31] A. F. M. Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering analysis for classifying student academic performance in higher education," *Applied Sciences*, vol. 12, no. 18, p. 9467, 2022.
- [32] S. Sarwat et al., "Predicting students' academic performance with conditional generative adversarial network and deep SVM," *Sensors*, vol. 22, no. 13, p. 4834, 2022.
- [33] M. J. S. Baron, J. S. G. Sanabria, and J. E. E. Diaz, "Deep neural network (DNN) applied to the analysis of student dropout in a higher education institution," *Investigacion e Innovacion en Ingenierias*, vol. 10, pp. 202–214, 2022.
- [34] D. Mahajan et al., "Smart Grading System Using Bi LSTM with Attention Mechanism," in *Advances in Communication and Applications*, vol. 1105, N. R. Shetty, N. H. Prasad, and H. C. Nagaraj, Eds., Lecture Notes in Electrical Engineering, Singapore: Springer Nature Singapore, 2024, pp. 247–260. doi: 10.1007/978-981-99-7633-1_18.
- [35] H. Liu, "Research on the construction method of curriculum teaching knowledge graph based on Bi-LSTM and CNN algorithm," *International Journal of High Speed Electronics and Systems*, vol. 34, 2025.

Authors' Profiles



Rudragouda G. Patil, Research Scholar, at Department of Computer Science & Engineering, Visvesvaraya Technological University. He has completed BE and M.Tech in Computer Science & Engineering. He is currently working as Assistant Professor, Department of Computer Engineering, Marathwada Mitramandal's College of Engineering. His research interests include Development of Machine Learning Algorithms for Education Assessments. Published articles in reputed International Journals and Conferences.



Mahantesh N. Birje received B.E., M. Tech. and Ph.D. in Computer Science and Engineering. He serves as Professor in the Department of Computer Science and Engineering and as Director of the IT Cell at Visvesvaraya Technological University (VTU), Belagavi, India. His current research areas include Cloud Computing, Internet of Things and Security. He has authored and co-authored more than 75 research papers published in reputed international refereed journals and conference proceedings. He has served as Chief Editor and Lead Guest Editor for several academic journals and has chaired technical sessions at international conferences. In addition, he is an active reviewer for leading international journals published by IEEE, Elsevier, Springer, and other prominent publishers.



Manisha Tapale holds a Ph.D. degree in Resource Provisioning in Cloud Computing from VTU, Belagavi. She is currently working as Associate Professor in Department of Computer Science and Engineering at KLE Technological University Belagavi. Her areas of interest and research focus on developing efficient mechanisms for resource allocation in cloud and edge computing. She is currently working on use cases in next-generation wireless networks. She has authored and published papers in referred International journals and conferences. She has served as session chair and reviewed articles.



Nagaraj V. Dharwadkar obtained his BE in Computer Science and Engineering in 2000 from the Karnataka University Dharwad, received his M.Tech in Computer Science and Engineering in 2006 from VTU, Belgaum and PhD in Computer Science and Engineering in 2014 from the National Institute of Technology, Warangal. He is Associate Professor in the department of Computer Science at Central University of Karnataka, Kadaganchi. He has 24 years of experience in teaching at professional institutions across India and has published 100 papers in various international journals and conferences. His areas of research interest are multimedia security, image processing, data mining and machine learning.

How to cite this paper

Rudragouda G. Patil, Mahantesh N. Birje, Manisha Tapale, Nagaraj V. Dharwadkar, "AI-Driven Rubrics for Academic Grading & Feedback Using Fuzzy Clustering & Attention Networks", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.4, pp. 1-24, 2026. DOI:10.5815/ijmecs.2026.04.01