iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78813

NLP Based System for Automated Assessment of Descriptive Responses

Thangeswari P[1], Swathigaa S M[2], Sivika T[3], Sasi Kasthuri S[4], E Shapna Rani[5]
[1, 2, 3, 4]UG students, Department of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, India
[5]Assistant Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, India

Abstract: Wireless Communication technology has made it much easier for students and faculty members to interact effectively in digital classrooms, ultimately leading to better educational outcomes. However, evaluating descriptive answers still remains a significant challenge, as current online testing systems have not yet provided a completely reliable solution. Manual evaluation, in particular, takes a considerable amount of time and often results in inconsistencies, especially during large-scale examinations, where variations in grading and assessment bias can occur. To address these issues, this paper presents an intelligent Descriptive Answer Evaluation System that combines Wireless Communication infrastructure with advanced Natural Language Processing (NLP) techniques to support real-time automated grading. The proposed system makes use of Sentence-BERT (SBERT) to measure semantic similarity between answers and a finite-tuned Text-to-Text Transfer Transformer (T5) model to evaluate contextual correctness, coherence, and completeness of responses. The system follows a modular design approach, consisting of components such as Admin, Staff, Student, Question Management, Answer Submission, NLP Processing, and Result Allocation modules. These modules work together to ensure smooth operation and allow easy scalability of the system. By reducing the dependency on manual evaluation, the system helps in maintaining consistent grading standards, making it highly suitable for educational institutions conducting digital assessments. Furthermore, experimental evaluation and results demonstrate that the proposed grading system provides better reliability compared to traditional systems that rely on keyword matching methods for assessment.
Keywords: Digital Assessment, NLP, SBERT, T5, Automated Evaluation, Semantic Similarity, Intelligent Grading System.

I. INTRODUCTION

The Education system is undergoing a significant transformation due to the rapid advancement and widespread adoption of Wireless Communication technologies in recent years. While human evaluation offers valuable contextual understanding, it also comes with several limitations, such as subjective bias, time consumption, inconsistent grading standards, and difficulty in handling large volumes of student responses. With the increasing number of students enrolling in educational programs and the growing popularity of online learning, there is a strong need for institutions to develop assessment systems that can evaluate written answers with both accuracy and reliability. NLP provides effective techniques for understanding and analyzing textual data. In particular, recent transformer-based models such as SBERT and T5 have shown exceptional performance in semantic analysis and contextual evaluations. The research focuses on developing an advanced system that integrates wireless-based digital answer submission with AI-powered automated assessment.

The main objectives of the proposed research are as follows:

1) To design a scalable framework for online descriptive answer evaluation.
2) To integrate SBERT for computing semantic similarity.
3) To incorporate T5 for evaluating contextual and logical correctness.
4) To develop a weighted scoring mechanism for fair mark allocation.
5) To ensure real-time result generation while minimizing bias.

II. PROBLEM STATEMENT

The use of online examination systems has become increasingly common in recent years. However, evaluating descriptive answers through these systems still presents several major challenges. Manual correction requires a significant amount of time and effort from faculty members. In addition, evaluation criteria can differ from one examiner to another, leading to inconsistencies in grading. Existing automated systems that rely on keyword matching are also limited, as they fail to capture the actual semantic meaning of student responses.

Furthermore, when the number of students increases, the assessment process becomes difficult to manage on a large scale. Therefore, there is a clear need for intelligent systems that can evaluate descriptive answers using NLP, while ensuring fairness and minimizing bias throughout the evaluation process.

III. LITERATURE REVIEW

Researchers working on automated answer evaluation have mainly focused on keyword matching and rule-based systems. While these methods provide a bias level of automation, they are not capable of identifying complex semantic relationships within the text. Early approached included techniques such as Latent Semantic Analysis (LSA), TF-IDF similarity, and N-gram matching.
However, these systems faced several key challenges. They were unable to properly understand paraphrased answers, had poor contextual comprehension, and offered limited scalability when applied to larger datasets.
With recent advancements in transformer-based models. There has been significant improvement in understanding semantic meaning. Sentence-BERT (SBERT) enhances sentence-level similarity computation, while T5 provides advanced contextual reasoning capabilities. Despite these improvements, existing academic evaluation workflows and Wireless Communication infrastructures still lack proper integration with these models. The research aims to bridge that gap.

IV. SYSTEM ARCHITECTURE

The proposed system follows a modular architecture designed for scalability and security. This is what our architecture diagram is:

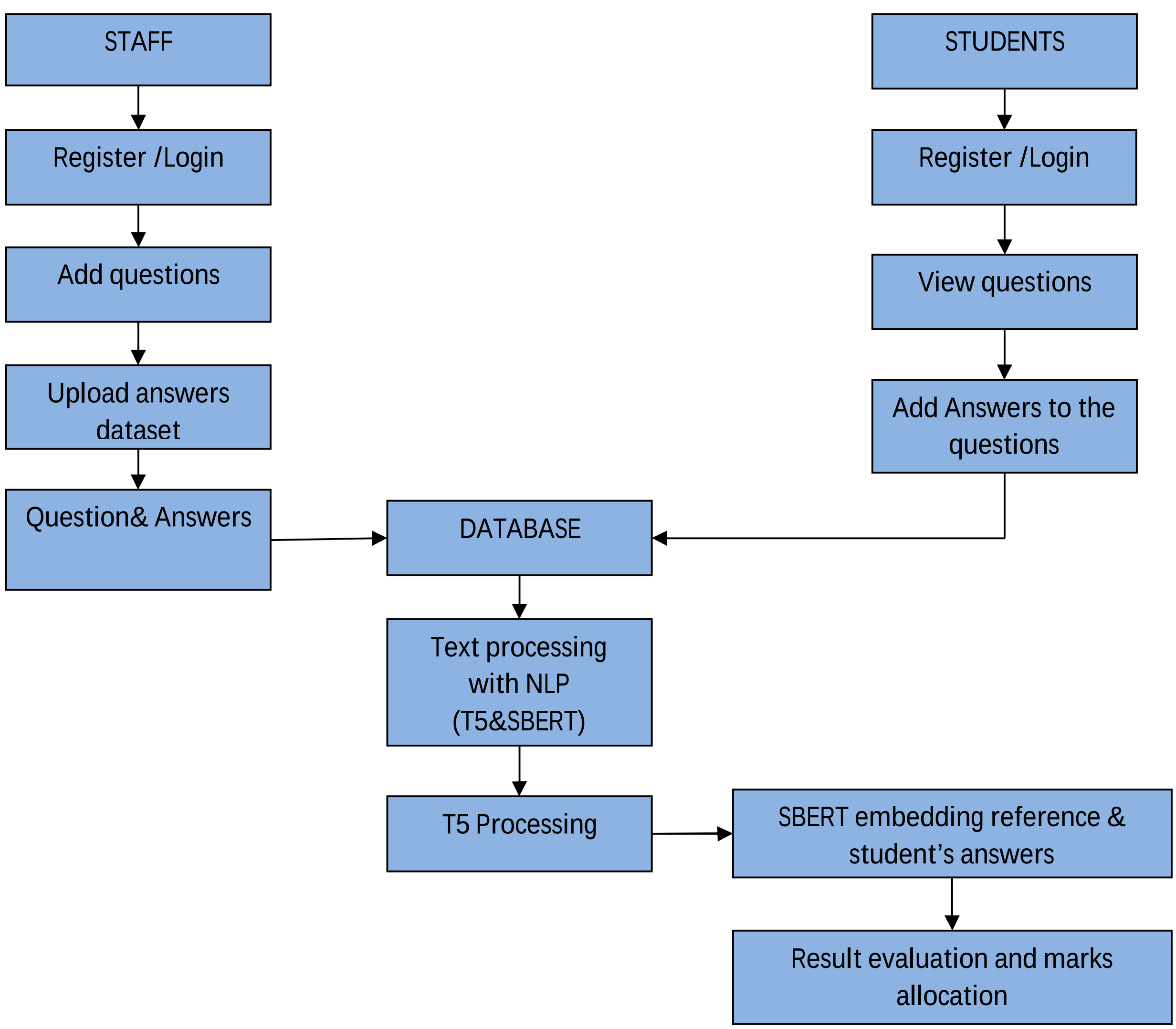


Fig.1. Architecture Diagram

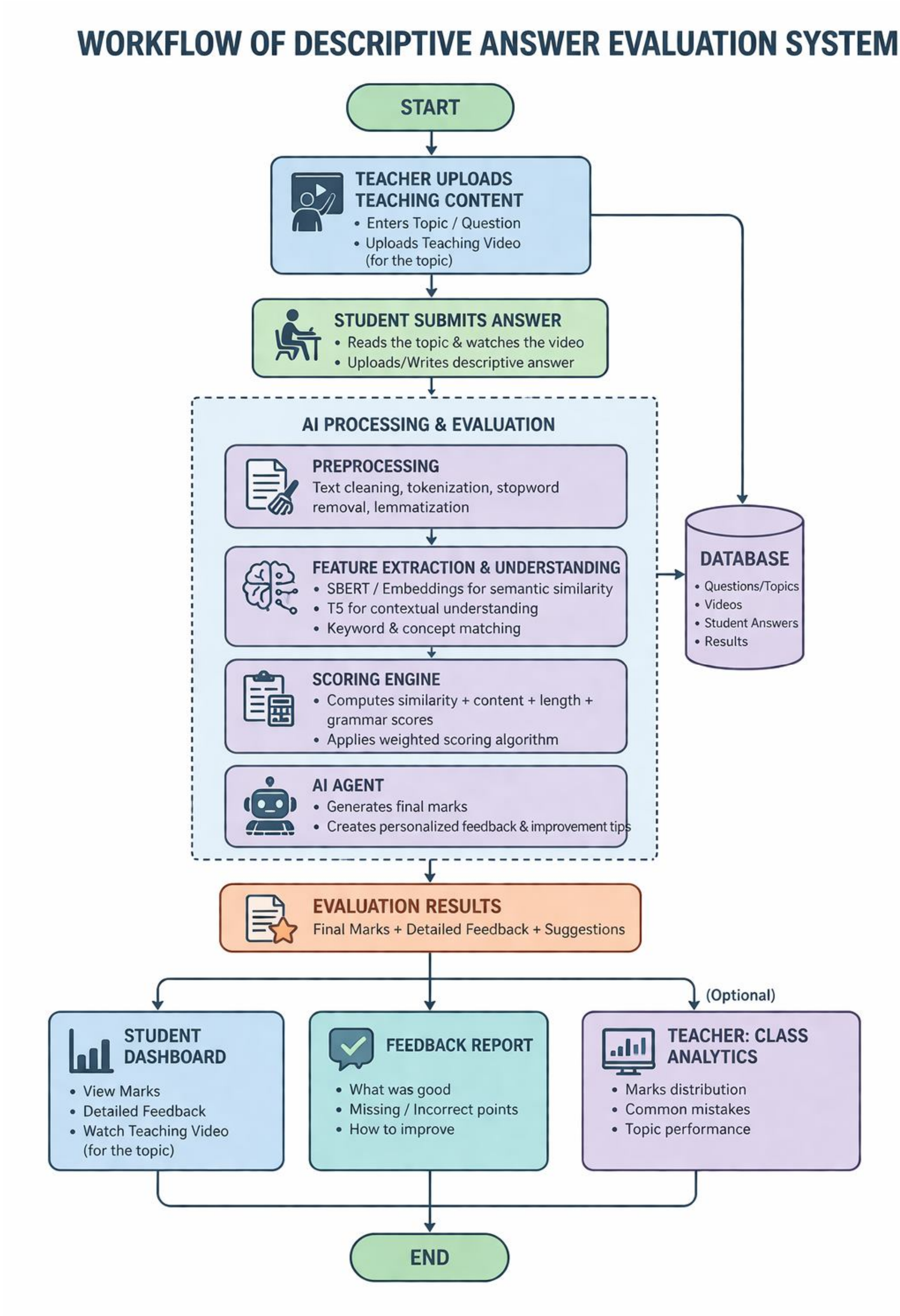


Fig.2. Workflow Diagram

V. MODULE DESCRIPTION

A. Admin Module

The Admin Module is responsible for managing the overall operations of the system. It provides administrators with necessary tools to control and monitor different activities within the platform.

Algorithm:

- bcrypt password hashing: is a secure method used to protect user passwords by converting them into an encrypted hash instead of storing the original password. It adds a random salt to each password and uses a slow hashing process, making it highly resistant to brute-force and hacking attacks

B. Staff Module

The Staff Module is designed to help faculty members to manage examinations in an efficient and organized manner. It provides a secure login interface through which staff can access the system and handle assessment-related tasks with ease. The module acts as a centralized platform where faculty can design, manage, and monitor examinations effectively.

With secure authentication in place, only authorized staff members are allowed to access the assessment tools, ensuring the protection of institutional data. Faculty can create descriptive questions, upload answers, and assign marks based on the academic requirements.

The module also supports real-time activation and deactivation of exams, giving instructors the flexibility to control exam schedules. The feature allows educators to adjust assessments according to course progress and student understanding. Overall, the module simplifies the process of exam management while maintaining accuracy and security.

Algorithm:

- SQL CRUD operations: refer to the basic database functions used to manage data: Create (insert new records), Read (retrieve data), Update (modify existing data), and Delete (remove data). These operations enable efficient storage, retrieval, and management of information within the system.

C. Student Module

The Student Module allows learners to participate in examinations within a safe and well-structured environment. Student can log in, access active tests, submit their answers, and review their results once the evaluation process is completed. It provides a secure and user-friendly platform that makes it easy for learners to take part in assessments.

Student can view available exams, submit their response, and access their results through a clear and organized interface. The system is designed to reduce technical issues and ensure smooth navigation during exams. Features like automated submission confirmation improve reliability and help eliminate uncertainty for students.

Additionally, transparent access to results builds trust in the evaluation process. Overall, the module promotes fairness, accessibility, and efficiency in conducting digital assessments.

Algorithm:

- Role-Based Access Control (RBAC): is a security mechanism that restricts system access based on user roles such as admin, teacher, or student. Each role is assigned specific permissions, ensuring that users can only access the features and data relevant to their role, thereby improving security and control.

D. Question Management Module

The module is responsible for handling all question related data while ensuring that it is stored and retrieved in an organized manner. It also helps prevent duplication by using similarity detection techniques.

The questions Management Module stores, organizes, and maintains all examination questions ina structured format. It ensures that question text, reference answers, and marks are properly indexed, allowing quick and easy retravel when needed. To avoid repetition, the module uses cosine similarity techniques to identify duplicate or highly similar questions.

The functionality supports the creation of balanced question papers and helps maintain academic integrity. In addition, the organized storage enables instructors to analyze the quality and difficulty level of questions. Overall, the module plays an important role in ensuring consistency and reliability across examinations.

Algorithm:

- Cosine Similarity. It measures the similarity between two questions by converting them into numerical vectors and calculating the cosine of the angle between them. A value close to 1 indicates high similarity, while a value close to 0 indicates low similarity. This helps in detecting duplicate or similar questions and avoids repetition in the question bank.

E. Answer Submission Module

The Answer submission Module is responsible for processing student response before evaluation using NLP preprocessing techniques. It prepares the submitted answers to ensure they are suitable for accurate analysis.

Initially, the raw text is tokenized and cleaned to remove unnecessary wordsand symbols. Stop word removal and lemmatization are applied so that the system focuses on meaningful content rather than grammatical variations. Text normalization is also performed to standardize the input format, which helps improve computational accuracy.

In addition, length validation is used to check whether the responses meet the required constraints before evaluation. All these steps work together to improve the overall quality and reliability of the automated grading process.

Algorithm:

- Tokenization: Tokenization is the process of breaking down the student's answer into smaller units such as words or sentences. This helps the system analyze each component individually and makes further processing like parsing and feature extraction easier.
- Stop Word Removal: Stop word removal eliminates commonly used words like "is", "the", and "and" that do not add significant meaning to the text. This reduces unnecessary data and helps the system focus on important keywords, improving processing efficiency.
- Lemmatization: Lemmatization converts words into their base or root form by considering their context. For example, "running" becomes "run". This helps group similar words together and improves the accuracy of comparison and evaluation.
- Text Normalization: Text normalization standardizes the text into a consistent format by converting it to lowercase, removing punctuation and special characters, and correcting variations. This ensures uniformity and improves the reliability of text analysis.

VI. NLP PROCESSING MODULE

The module serves as the core intelligence engine of the system. It is responsible for analyzing student answers and comparing them with reference responses using advanced NLP models.

The NLP Processing Module acts as the analytical core, Focusing on interpreting and evaluation textual answers. Unlike traditional keyword-based grading systems, it goes beyond simple word matching by understanding semantic meaning and contextual relationships. Advanced language models are used are used to compare student responses with reference answers.

The approach allows the system to recognize conceptual correctness even when the wording is different. By combining multiple NLP techniques, the module enhances the accuracy of evaluation. As a result, it is able to simulate human-like understanding while still maintaining computational efficiency.

A. SBERT Sub-Module

The SBERT model is used to convert textual answers into high-dimensional vector representations, making semantic comparison possible. The sub-module transforms both student responses and references answer into numerical vectors within a high-dimensional semantic space.

Each answer is encoded into embedding that capture its contextual meaning. Cosine similarity is then calculated between these vectors to determine how closely the student's responses matches the reference answer. The method allows the system to identify paraphrased answers as correct, as long as they convey the same concept.

By reducing reliance on exact keyword matching, the SBERT sub-module enables more accurate and meaningful evaluation. As a result, it supports fair and concept-based assessment of student answers.

Process:

Convert reference answer → embedding vector A

Convert student answer → embedding vector B

Calculate cosine similarity

$$\text{Similarity} = (A.B) / (\| A \| \times \| B \|) \quad (1)$$

B. T5 Sub-Module

The T5 model is used to analyze the deeper contextual understanding of student responses. The sub-module evaluates answers based on their context and logical coherence rather than just surface-level similarity.

It examines factors such as sentence structure, flows of ideas, and the completeness of explanations. Unlike similarity-based models, T5 focuses more on the qualitative aspects of writing.

The module also determines whether a response adequately address the requirements of the question. Based on the analysis, a contextual score is generated, reflecting the student's conceptual clarity and reasoning ability. The helps enhance the overall evaluation by taking into account the depth of understanding.

It generates a contextual quality score after analyzing both student and reference answers, ensuring conceptual correctness.

VII. SCORING ALGORITHM

The module is responsible for calculating the final marks using a weighted scoring strategy that balancing semantic and contextual accuracy. It combines the outputs from both SBERT and T5 models to generate the final score.

The Scoring algorithm applies a weighted formula where semantic similarity and contextual quality are considered together for fair evaluation. A higher weight is given to SBERT to prioritize conceptual correctness, while T5 contributes by evaluating explanation quality and logical structure.

Once the final score is computed, it is multiplied by the maximum marks to determine the student's grade. The hybrid approach ensures that the evaluation is reliable, consistent and unbiased.

$$\text{Final Score} = (0.7 * \text{SBERT score}) + (0.3 * \text{T5 score}) \quad (2)$$

$$\text{Marks} = \text{Final Score} * \text{Maximum Marks} \quad (3)$$

Overall, the hybrid scoring method helps maintain fairness and accuracy in the evaluation process.

VIII. DATABASE DESIGN

The System uses PostgreSQL as it structured database platform to securely store examination data. It follows a well-organized approach by maintaining separate tables for Admin, Staff, Student, Question, Answer, and Result entities.

Foreign key relationships are used to maintain referential integrity between these tables, ensuring that all data remains properly connected. The database structure is carefully normalized to reduce redundancy and improve storage efficiency. Each table is designed with appropriate constraints to maintain accurate relationships between different entities.

To enhance performance, indexing mechanisms are implemented to speed up data retrieval during evaluation and result generation. Backup procedures are also integrated to safeguard records against accidental loss, while transaction management ensures that multiple operations are executed without causing data inconsistencies.

The structured database design supports reliability, scalability, and secure maintenance of academic records. It helps prevent duplication, data loss and inconsistencies. Efficient indexing allows quick data retrieval even for large datasets, ensuring smooth system performance.

Separate tables are maintained for each entity:

1) Admin
2) Staff
3) Student
4) Question
5) Answer
6) Result

All tables are interconnected using foreign keys to maintain referential integrity and prevent data inconsistencies.

IX. IMPLEMENTATION DETAILS

The system backend is developed using the Flask framework, which provides a lightweight yet powerful environment for handling server-side operations. Database connectivity is managed through SQL Alchemy, ensuring smooth and efficient communication with PostgreSQL.

NLP models are integrated using the sentence-transformers and transformers libraries, running on a torch backend. These tools enable the efficient execution of deep learning algorithms required for answer evaluation.

A. NLP models are integrated using
- Sentence-transformers library
- Transformers library
- Torch backend

Student can submit their answers wirelessly through a secure web interface, making the system easily accessible from remote locations. The implementation follows a modular approach, which simplifies maintenance and supports future upgrades.

Each module is developed and tested independently before being integrated into the system, ensuring overall stability. API endpoints are well-structured to handle requests in an efficient and secure manner. In addition, error-handling mechanisms are implemented to manage unexpected inputs or server interruptions.

Logging systems are also included to monitor system activity, which helps in debugging and performance tracking. Overall, the development approach results in a stable, maintainable, and scalable software solutions.

The backend is also designed to handle multiple user requests simultaneously, ensuring smooth performance even during peak usage. Efficient routing and request handling mechanisms help reduce latency and improve responses time. Security measures such as authentication, and data validation are implemented to protect sensitive user information.

Furthermore, the system supports easy integration with external services and APIs, making it adaptable to different institutional requirements. The modular structure allows developers to update or enhance specific components without affecting the entire system, ensuring long-term sustainability.

B. Dashboard and Feedback Features
In addition to the core functionality, the system includes an interactive dashboard that provides a comprehensive view of student performance and grading details. The dashboard presents information such as marks obtained, comparison with reference answers, and overall progress in a clear and structured format.

It also supports graphical representations of performance, helping students and instructors quickly understand trends and patterns. The visual insight makes it easier to track improvement over time and identify areas that require attention.

A smart feedback mechanism is integrated into the system to provide detailed and meaningful insights. It not only highlights correct and incorrect portions of answers but also explains where improvements are needed. The helps students refine their understanding and improve their response quality.

Additionally, feedback can be personalized based on performance patterns, making the learning experience more effective. The combination of dashboard analytics and intelligent feedback enhances transparency, encourages self-assessment, and supports continuous

X. EXPERIMENTAL SETUP
The system was evaluated using descriptive answers from various Computer Science subjects. Experimental testing was carried out using responses collected from multiple domains within Computer Science. Several performance metrics were considered during evaluation, including semantic accuracy, contextual relevance, and processing speed.

The results showed a strong alignment between automated scores and human grading. The system maintained high consistency across repeated evaluations, while also significantly reducing the manual workload compared to traditional grading methods. These findings confirm the effectiveness of the proposed approach.

A. Evaluation Metrics
1) Semantic similarity accuracy
2) Contextual relevance score
3) Processing time
4) Grading consistency

B. Results
1) Improved grading consistency
2) Reduced manual workload
3) Strong correlation with human evaluation

Further testing was performed using different types of answer patterns, including concise, detailed, and paraphrased responses. The system consistently delivered stable performance across these varied input styles. Statistical analysis indicated minimal deviation between automated scores and expert evaluations.

Additionally, response processing time remained within acceptable limits, even when handling large datasets. Repeated trials confirmed that the results were both reliable and reproducible. Overall, these observations highlight the robustness and validity of the proposed evaluation system.

XI. ADVANTAGES

The proposed system has significant advantages in modern education. It reduces faculty workload by automating the evaluation of descriptive answers, allowing instructors to focus more on teaching rather than manual grading.

It also minimizes bias and ensures consistent scoring, as all responses are evaluated using predefined algorithms. This guarantees fair and uniform assessment for every student.

The system provides instant feedback, helping students quickly understand their performance and improve learning outcomes.

Additionally, it supports large-scale assessments, making it suitable for institutions handling a high volume of students without compromising accuracy.

The system promotes transparency and reliability in grading by maintaining digital records, which can be used for review or auditing purposes.

Finally, its easy integration and modular design allow institutions to adopt and customize the system efficiently, making it a flexible and scalable solution.

XII. FUTURE ENHANCEMENTS

The system can be further improved by introducing several advanced features. Adding multilingual support would allow the evaluation of answers written in different languages, making the system more inclusive. Automated feedback generation could help students understand their mistakes and improve their learning outcomes.

In addition, emotion detection techniques could be used to analyze the tone and sentiment of responses. Deploying the system on the cloud would improve scalability and accessibility, allowing it to handle a larger number of users efficiently. Integration with Learning Management Systems would also enable smoother and more streamlined academic workflows.

1) Potential improvements include:
2) Multilingual support
3) Automated feedback generation
4) Emotion detection in responses
5) Cloud-based scalable deployment
6) Integration with Learning Management Systems

XIII. APPLICATIONS

The proposed system can be implemented across a wide range of educational and assessment environments. Universities can use it for internal examinations as well as continuous assessment. Competitive exam authorities may adopt it for evaluating descriptive answers, while certification platforms can integrate it for skill-based testing. Online learning systems can also use it to provide automated grading for assignments. Its flexibility makes it suitable for both academic and professional evaluation contexts.

The proposed system can be applied in multiple educational environments:

1) Universities
2) Competitive examinations
3) Certification platforms
4) Online learning systems

Beyond academic institutions, the system can also be used in corporate training environments to assess employee performance. Recruitment agencies can utilize it to evaluate descriptive aptitude tests more efficiently. Government organizations conducting large-scale examinations can benefit from automated grading as well.

Additionally, distance education platforms can integrate the proposed system for assignment evaluation, and professional certification bodies can adopt it for standardized testing. Overall, its adaptability makes it suitable for a wide variety of educational and professional assessment scenarios.

XIV. CONCLUSION

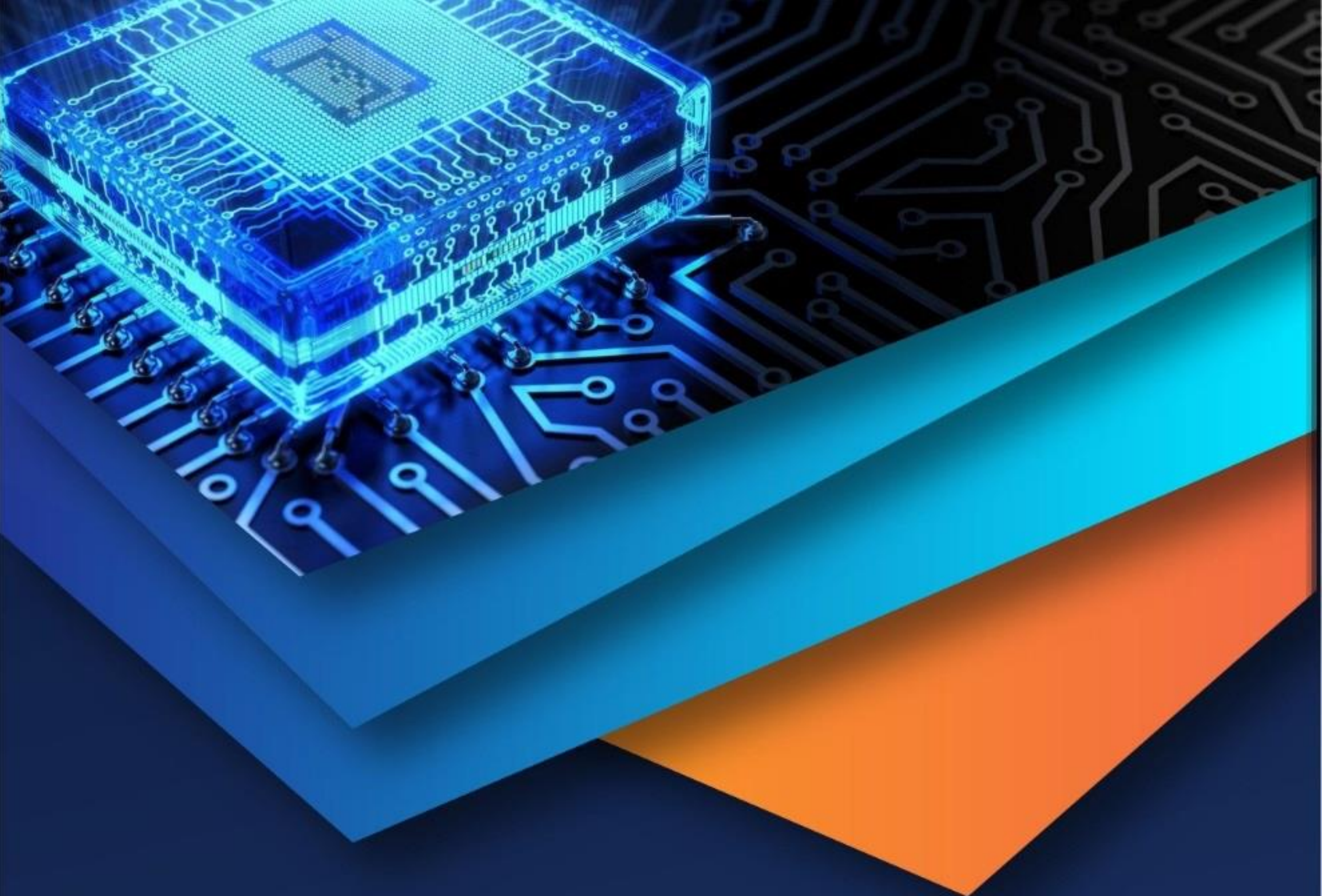
The research introduces an intelligent Descriptive Answer Evaluation System that combines Wireless Communication with advanced NLP models. By integrating SBERT for semantic similarity and T5 for contextual reasoning, the system is able to provide accurate, unbiased, and efficient grading. Its modular architecture supports scalability, and experimental results confirm improved consistency along with reduced manual effort.

Overall, the proposed solution represents a significant advancement in automated academic assessment and effectively addresses the growing need for scalable digital education technologies.

The research presents an intelligent descriptive answer evaluation system that integrates wireless technology with advanced NLP models. By combining semantic similarity analysis using SBERT and contextual reasoning using T5, the system achieves accurate grading. Its modular architecture ensures scalability and supports efficient workflow management. Experimental results also demonstrate strong agreement with human evaluators.

REFERENCES

[1] L. Ralitha Manasachanadrapati and C. Koteswara Rao, “Descriptive Answer Evaluation Using NLP Approaches,” IEEE,2024. DOI: 10.1109/ACCESS.2024.3391201.
[2] G. M. Rasiqul Islam, A. Saifat, and M. M. Fahim Hasin, “Mobile Based MCQ Answer Sheet Analysis and Evaluation Application,” IEEE, 2024. DOI: 10.1109/ACCESS.2024.3387456.
[3] P. Sushila Devi, S. Sarkar, S. Sonamani Singh, and H. Hotham, “An Approach to Evaluate Subjective Answers Using BERT Model,” IEEE, 2022. DOI: 10.1109/ACCESS.2022.3164725.
[4] P. Kapprand and S. Koteswara Rao, “Subjective Answer Evaluation Using Keyword Similarity and Regression Techniques,” IEEE, 2024. DOI: 10.1109/ACCESS.2024.3389024.
[5] V. Suresh, R. Agasthiya, J. Ajay, A. Amrith Gold, and D. Chandra, “AI Based Automated Essay Grading System Using NLP,” IEEE, 2023. DOI: 10.1109/ACCESS.2023.3278415.
[6] P. S. Raut, C. Adhesh, D. Choudhari, V. B. Waghole, and P. U. Jadhav, “Automatic Evaluation of Descriptive Answers Using NLP and Machine Learning,” IEEE, 2022. DOI: 10.1109/ACCESS.2022.3149872.
[7] M. Menakshi, “Descriptive Answer Evaluation System Using Cosine Similarity,” IEEE, 2019. DOI: 10.1109/ACCESS.2019.2942156.
[8] A. Chaayni et al., “Automatic Essay Scoring Using Jaccard and Cosine Similarity,” IEEE,2025. DOI: 10.1109/ACCESS.2025.3401267.
[9] V. Sreedhivya and J. Narayan, “Short Descriptive Answer Evaluation Using Word Embedding Techniques,” IEEE, 2021. DOI: 10.1109/ACCESS.2021.3098764.
[10] A. C. Condor, “Exploring Automatic Short Answer Grading as a Tool to Assist in Human Rating,” IEEE, 2020. DOI: 10.1109/ACCESS.2020.2986543.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY