Beyond Marks: NLP-Based Learning Gap Detection and Remediation

Tanuja Gorte[1], Vrushali Labade[2], Vikrant Gawai[3], Prof. Rashmi Pathak[4]
[1,2,3,4]*Department of Computer Application, MIT Arts, Commerce and Science College, Pune, India*
doi.org/10.64643/ijirt.208433-459

***Abstract*—Conventional grading in Computer Science education evaluates students mainly through numeric marks. The issue with this approach is that a single score rarely reflects what a student truly understands, partially grasps, misinterprets, or misses entirely. To solve this, this research outlines a domain-specific Artificial Intelligence method leveraging Natural Language Processing (NLP) and semantic analysis to pinpoint exact knowledge gaps within written answers. The framework processes theoretical responses by comparing them against expected concept-level knowledge for a given question, then categorizes a student's comprehension into four distinct states: correctly understood, partially understood, misunderstood, or missing. Using these mapped gaps, the system generates customized learning aids, offering simplified breakdowns, clear examples, curated study resources, and targeted drill questions. To test the system's performance, the NLP-based gap detection will be evaluated against standard score-based grading. Expert review alongside standard statistical metrics specifically accuracy, precision, recall, and F1-score will be applied to measure how reliably the system identifies these gaps. Additionally, a follow-up post-assessment will be run to check if tailored remediation actually enhances student conceptual retention. Centered on Computer Science coursework, this study seeks to show how specialized NLP tools can enable deeper, adaptive, and diagnostic learning beyond basic mark-based testing.**

***Index Terms*—NLP, Natural Language Processing, AI in Education, Computer Science Education, Learning Gap Detection, Semantic Analysis, Personalized Learning.**

I. INTRODUCTION

Standard testing in education focuses primarily on final marks, but a total score rarely shows what a student actually understands or where their learning difficulties lie. Two students can earn the exact same grade while holding completely different levels of comprehension across individual topics. This limitation is particularly critical in Computer Science, where core subjects like programming, data structures, databases, and computer networks rely on deeply interconnected concepts. A student might grasp one core idea perfectly, yet have a partial or completely incorrect understanding of another. Identifying these specific learning gaps at the concept level provides far more meaningful insight than relying on an overall examination mark. Consequently, recent research in conceptual gap detection and personalized learning explores automated analysis to pinpoint student struggles and offer tailored educational support.
To address this gap, this paper outlines a domain-specific Natural Language Processing (NLP) framework designed specifically for Computer Science education. The framework takes a student's unstructured theoretical answer and runs it through foundational NLP preprocessing steps, such as tokenization, text normalization, and stop-word removal.

Key concepts are then extracted from the response and analyzed semantically using sentence representations, comparing them against expected concepts stored in a reference knowledge base. Through this process, each concept is categorized as correct, partially understood, or missing—allowing the framework to systematically map a student's weak points. Using these identified gaps, the system generates targeted remediation through simplified explanations, clear examples, and concept-specific practice questions. Rather than distributing identical study materials to every student, this approach directly connects assessment with personalized learning. Finally, as this study presents a conceptual framework, its practical effectiveness and overall learning impact are designed to be evaluated in future implementations using empirical performance metrics and expert reviews.

II. LITERATURE REVIEW

Identifying learning gaps has become central to modern educational assessment because total examination scores fail to show which individual concepts a student has mastered and where specific misunderstandings persist. This diagnostic limitation is especially noticeable in Computer Science. Theoretical answers in CS typically contain multiple interconnected knowledge elements; a student might demonstrate a firm grasp of one concept while missing another, earning a decent overall score despite underlying technical weaknesses. Moving past aggregate grades, the CORE-GAP framework emphasizes identifying these precise conceptual deficiencies to enable targeted learning support [3]. Foundational NLP research supports this direction, particularly through semantic text representations capable of evaluating student answers against model responses [4].

Several studies have explored semantic analysis for evaluating open-ended technical answers. Ndukwe et al. examined automated short-answer grading within a Computer Science networking course using semantic language models [6]. Their findings confirmed that NLP techniques can effectively grade unstructured student text, though performance varies depending on question design and answer format. While their work establishes a valuable baseline for applying semantic analysis to CS theory questions, its primary objective remains automated scoring rather than mapping specific conceptual gaps. Similarly, Botelho et al. applied sentence-level semantic representations alongside machine learning algorithms to grade and deliver feedback on open-ended mathematics responses [1]. Although their model effectively predicted student performance and automated feedback recommendations, it operated primarily at the full-response level rather than explicitly isolating which sub-concepts were missing or partially understood.

Beyond grading, research has increasingly shifted toward generating adaptive instructional support. Kochmar et al. designed an NLP and machine learning pipeline to deliver personalized feedback within an intelligent tutoring platform, proving that tailored guidance significantly enhances student learning outcomes [2]. The CORE-GAP model aligns closely with this objective by detecting specific concepts students understand or miss and linking those gaps directly to customized learning interventions [3]. However, existing literature largely addresses these challenges in isolation focusing separately on automated grading, semantic answer processing, gap identification, or feedback generation without integrating them into a unified domain-specific framework for CS theory answers.

Taken together, existing research reflects a clear shift from basic automated scoring toward semantic analysis and adaptive educational support [4]. Nevertheless, a distinct gap remains in unifying free-form CS answer parsing, automated concept extraction, semantic alignment against expected knowledge bases, explicit classification of concepts (as correct, partial, or missing), and targeted concept-specific remediation within a single pipeline. The research presented here directly addresses this gap by creating an integrated NLP framework that bridges student answer evaluation, weak-concept detection, and tailored instructional remediation.

III. RESEARCH GAP

Existing studies have addressed several important components of automated analysis and personalized learning. Ndukwe et al. demonstrated the application of Sentence-BERT (SBERT) to automated assessment of short answers in a Computer Science networking context [6]. Botelho et al. examined NLP-based assessment and feedback for open-ended student responses, showing the value of semantic analysis in interpreting free-form answers [1]. Personalized feedback has also been investigated as a means of responding to individual learning needs [2], while the CORE-GAP framework focuses more directly on identifying concept-level gaps and connecting them with personalized learning support [3]. These studies therefore establish useful foundations for answer analysis, semantic assessment, feedback, and concept-gap identification.

However, an insufficiently addressed aspect is the structured integration of these capabilities specifically for free-form Computer Science theory answers. Automated grading can indicate the quality or score of a response, but a total score does not necessarily identify which individual concepts a student understands, partially understands, or has not demonstrated. Similarly, identifying a conceptual gap becomes more useful when it can directly guide appropriate remediation. This study therefore proposes a framework that connects NLP preprocessing, concept extraction, SBERT-based semantic comparison, correct/partial/missing classification, learning-gap detection, and concept-specific remediation. The proposed framework is intended to provide a more detailed diagnostic pathway from a student's written response to targeted learning support, without claiming that the individual underlying techniques are themselves new.

IV. OBJECTIVES

The primary objectives of this proposed study are as follows:

1. Process free-form Computer Science theoretical answers using standard NLP techniques.
2. Extract core concepts from student text and evaluate their semantic similarity against expected responses.
3. Isolate specific conceptual misunderstandings and omitted topics to map individual learning gaps.
4. Deliver automated, targeted remediation through concept-specific explanations, practical examples, and practice questions.

V. METHODOLOGY

The proposed framework evaluates student learning gaps in Computer Science by examining unstructured, free-form theoretical responses. Rather than relying on aggregate marks, the system centers its evaluation on the specific concepts articulated within a student's text. The framework operates across eight sequential stages, tracing the pipeline directly from initial answer submission to tailored educational support. The complete end-to-end workflow is illustrated in Fig. 1

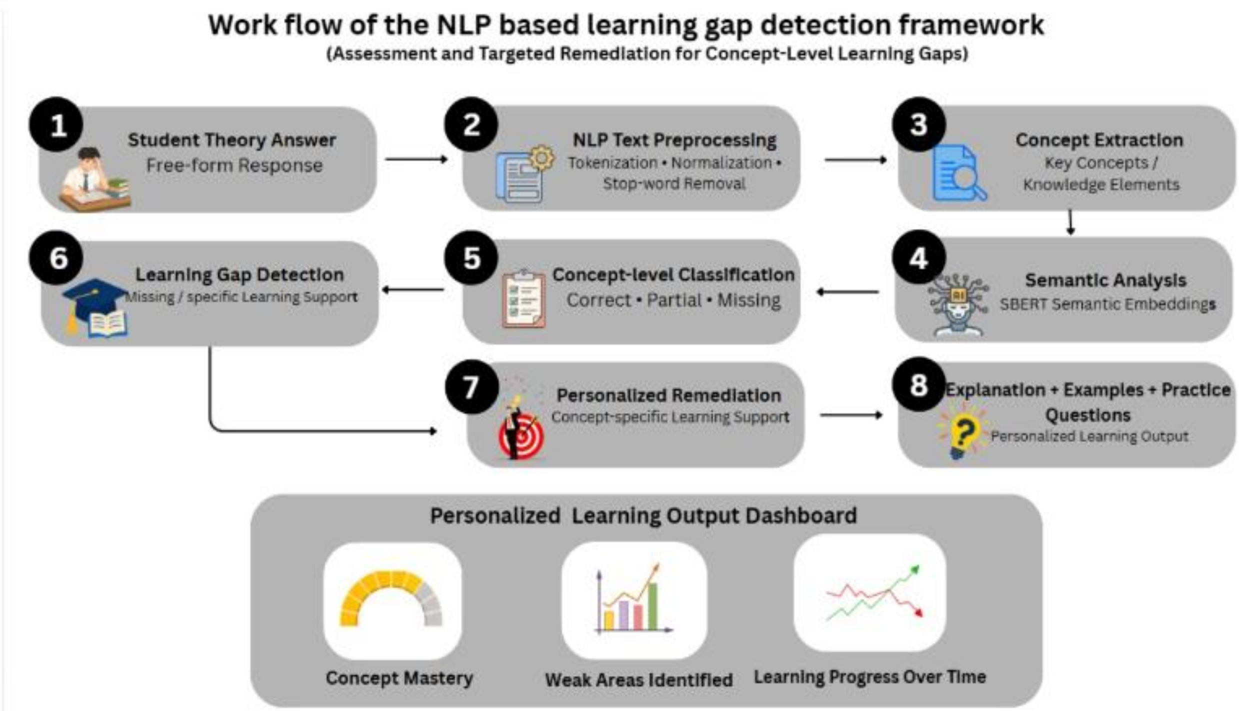

Fig. 1. Workflow of the proposed NLP-based learning gap detection framework.

A. Student Theory Answer

The pipeline initiates when a student submits a written, free-form answer to a Computer Science theoretical question. This submitted text typically contains a mix of correct statements, partial explanations, active misconceptions, or completely omitted concepts. Unlike standardized multiple-choice testing, an open-ended theoretical response yields rich textual data suitable for extracting and evaluating a student's actual conceptual grasp.

B. NLP Text Preprocessing

To prepare the submitted text for detailed analysis, the raw response passes through standard NLP preprocessing routines. The system tokenizes and normalizes the string to establish a uniform textual format, while stripping out non-informative stop words. Removing these lexical variations cleans the response string, enabling reliable downstream concept extraction and semantic comparison.

C. Concept Extraction

Following preprocessing, the system isolates key theoretical concepts and knowledge components embedded within the student's text. For instance, in a networking question addressing TCP mechanics, the extraction routine isolates distinct operational concepts such as reliability, acknowledgments, retransmission, and flow control. Segmenting these ideas individually prevents the system from evaluating the response as a monolithic block, allowing granular concept-level assessment.

D. Semantic Analysis

Once extracted, student concepts are evaluated against benchmark concepts linked to the specific question within a curated reference knowledge base. To measure conceptual alignment, the proposed framework employs Sentence-BERT (SBERT) to map textual statements into dense vector embeddings. SBERT converts text into mathematical vectors that capture contextual meaning, placing semantically equivalent phrases in close proximity within the vector space to calculate cosine similarity [5].

For example, if a reference answer states, *"TCP retransmits packets when acknowledgment is not received,"* and a student writes, *"TCP sends the data again if the receiver does not acknowledge it,"* SBERT detects high semantic alignment despite differences in vocabulary. This semantic matching ensures students are not penalized for using alternative phrasing, and the resulting similarity score serves as the input for conceptual classification.

E. Concept-Level Classification

Based on the semantic similarity score, each isolated concept is categorized into one of three distinct understanding levels:

1. *Correct:* The concept is accurately and sufficiently explained.
2. *Partial:* The concept is present, but the explanation lacks complete technical detail.
3. *Missing:* The required concept is absent or inadequately covered.

Mapping responses at this granular level ensures that strong conceptual understanding in one area does not mask specific technical confusion in another.

F. Learning Gap Detection

The framework uses these classification outputs to map the student's exact learning gaps. Concepts flagged as either partial or missing are isolated as primary weak points. Rather than returning a static grade, the system constructs a detailed concept-level profile highlighting the specific Computer Science topics that require further study. This diagnostic approach aligns with prior research by Grenander et al., who analyzed student responses to categorize correct and incorrect concepts for generating targeted educational feedback [4].

G. Personalized Remediation

Once specific weak concepts are isolated, the system maps them directly to tailored instructional resources. This ensures that academic support targets the individual student's precise needs rather than distributing uniform, generic review materials to all learners. For instance, if a student's submission indicates a breakdown in understanding "TCP retransmission," the framework delivers targeted material exclusively on that mechanism, bypassing fully mastered networking concepts. This diagnostic intervention reflects established research on automated personalized feedback, which highlights the benefit of adapting hints and explanations to individual learning gaps [2].

H. Explanation, Examples, and Practice Questions

In the final stage, the framework outputs a customized remediation module for every detected learning gap. The system generates a clear, simplified concept breakdown, contextualized technical examples, and focused practice problems. Providing this targeted practice allows students to directly address and resolve their specific conceptual weaknesses prior to future assessments.

VI. PROPOSED RESULTS

The proposed NLP framework expects to deliver significantly richer diagnostic feedback regarding student learning compared to traditional mark-based testing. Rather than summarizing performance into a single numerical score, the system aims to isolate which concepts a student grasps, partially understands, or omits within an open-ended theoretical response.

The first anticipated outcome involves granular, concept-level evaluation of written text. Following standard NLP preprocessing and vector analysis, student answers are compared semantically against core reference concepts. This allows the system to recognize conceptually equivalent phrasing even when vocabulary diverges from the answer key. Consequently, evaluation prioritizes conceptual meaning over simple keyword matching.

The second expected outcome centers on pin-pointing individual learning gaps. Categorizing each expected concept into discrete states—Correct, Partial, or Missing—yields an explicit map of the student's knowledge state. A student securing an acceptable overall grade may still struggle with a specific sub-topic; the proposed framework flags that localized weakness rather than obscuring it within an aggregate grade, as illustrated in Table I.

TABLE I. ILLUSTRATIVE CONCEPT-LEVEL DIAGNOSTIC OUTPUT

Expected Concept	Student Response Evidence	Proposed Classification	Recommended Action
Concept A	Correct Explanation	Correct	Continue to next concept
Concept B	Incomplete explanation	Partial	Provide clarification and example
Concept C	No relevant evidence	Missing	Provide concept explanation and practice

The third expected outcome involves generating a dynamic learner-specific gap profile. Collecting topics marked as Partial or Missing into a unified dashboard provides a clear overview of where academic support is necessary.

This diagnostic profile offers far more utility than assigning uniform remedial content to all low-scoring students indiscriminately.

The fourth key outcome addresses targeted educational remediation. Based on detected knowledge gaps, the system presents concept-specific explanations, practical examples, and drill problems. If an answer reflects partial comprehension of one idea and complete absence of another, the framework tailors' distinct levels of intervention for each, establishing a direct link between diagnostic evaluation and learning support.

Furthermore, the proposed evaluation scheme compares this NLP-driven model against conventional mark-based gap identification, as outlined in Table II. This comparison aims to determine whether concept-level processing yields superior diagnostic clarity, utilizing subject-matter experts to benchmark the accuracy of the Correct/Partial/Missing classifications.

TABLE II. PROPOSED COMPARISON

Assessment approach	Main output	Learning-gap information
Traditional score-based	Overall score/grade	General indication of performance
Proposed NLP approach	Concept-level classification	Specific concepts requiring support

Ultimately, the primary goal extends beyond merely raising test scores. The core contribution lies in establishing a diagnostic pipeline connecting a student's written response to specific conceptual gaps and targeted remediation activities. The true utility of this approach will be validated through empirical testing on real student data and expert review.

VII. LIMITATIONS

Several constraints within the proposed framework warrant consideration. Primary among these is that the system currently exists as a conceptual framework without experimental validation; consequently, its real-world classification accuracy and instructional effectiveness remain unverified. Furthermore, the operational scope focuses exclusively on Computer Science theoretical responses, which may not capture practical coding competencies or other forms of student learning. The accuracy of gap detection remains heavily reliant on the quality and thoroughness of the underlying concept knowledge base. While SBERT effectively measures semantic alignment, highly technical jargon, ambiguous phrasing, or poorly structured student text can complicate accurate interpretation. Additionally, establishing boundaries between correct, partial, and missing categories requires fine-tuned similarity thresholds alongside rigorous expert validation. Lastly, the impact of the personalized remediation pipeline depends entirely on having access to high-quality, concept-aligned instructional materials, examples, and practice sets.

VIII. DISCUSSION

The proposed framework addresses a fundamental distinction in Computer Science education: the difference between measuring overall performance and diagnosing genuine conceptual understanding. While an aggregate examination score summarizes how well a student answered a question, it frequently conceals which specific concepts were mastered, partially grasped, or completely omitted. In theory-driven Computer Science coursework, this distinction is critical because a single response often spans multiple interconnected ideas; a student may demonstrate complete fluency in one concept while exhibiting clear misunderstandings in another within the same answer. Analyzing responses at the concept level provides the detailed diagnostic visibility required to guide effective instructional choices.

The primary contribution of this work lies in linking multiple analytical stages into a unified diagnostic pipeline. Standard NLP preprocessing and concept-extraction routines transform unstructured, free-form text into modular knowledge elements, while Sentence-BERT (SBERT) provides the core semantic engine for evaluating student phrasing against reference concepts. SBERT was specifically developed to generate dense sentence embeddings for efficient semantic similarity computation [5], and its application to grading Computer Science short answers confirms its utility in processing open-ended educational text [6]. This capability is essential because students frequently express correct Computer Science principles using varied phrasing or non-standard syntax, rendering rigid keyword matching inadequate. Crucially, SBERT alone does not determine conceptual correctness; within this framework, its similarity scores serve as underlying evidence to drive the subsequent Correct, Partial, or Missing classification.

Furthermore, the framework bridges the gap between diagnostic answer processing and active instructional intervention. Prior studies on open-ended text analysis demonstrate the broad utility of NLP for educational assessment and feedback [1], while research into adaptive tutoring highlights the benefits of tailoring guidance to individual learner profiles [2], particularly through detailed response processing [4]. Concept-gap research further underscores the value of isolating specific knowledge deficits to drive personalized instruction [3]. Building upon these foundations, the framework focuses specifically on CS theoretical answers, integrating semantic parsing, concept classification, gap mapping, and targeted remediation into a single workflow. The contribution is not the introduction of a new SBERT architecture or NLP algorithm, but rather the systematic integration of these established techniques into a practical diagnostic framework.

Several implementation challenges must be managed in future practical deployments. The accuracy of the underlying reference knowledge base and concept-extraction routines directly governs the reliability of subsequent classifications. Highly ambiguous, fragmented, or poorly phrased responses present natural processing hurdles, requiring empirical tuning of semantic similarity thresholds rather than reliance on static assumptions. Rigorous expert review will be essential to validate the boundary conditions of the Correct, Partial, and Missing categories, while the practical value of the remediation pipeline will depend on the clarity and relevance of the instructional materials provided. Ultimately, while this framework offers a foundation for delivering diagnostic insights to instructors and tailored support to students, its empirical reliability and true educational impact remain to be validated through future implementation and testing.

IX. FUTURE SCOPE

Future research will prioritize the practical implementation and experimental validation of the proposed framework using empirical Computer Science student responses. Subsequent studies can evaluate alternative sentence-embedding architectures and domain-specific language models alongside SBERT to benchmark classification performance for concept-level gap detection. Furthermore, the concept-extraction pipeline can be enhanced by integrating specialized Computer Science knowledge graphs and expert-curated concept hierarchies. Subject-matter experts and instructors will be engaged to annotate student responses, establishing a gold-standard baseline to validate the accuracy of the Correct, Partial, and Missing classification boundaries.

Beyond core diagnostic functionality, system capabilities can be expanded through adaptive assessments that dynamically recalibrate question difficulty based on detected knowledge gaps. Maintaining longitudinal learner profiles across multiple assessment cycles would further allow the system to track conceptual progress and retention over time. The remediation module can similarly be broadened to incorporate diverse educational modalities, including instructional videos, interactive worked examples, and targeted coding exercises. Once validated within theory-heavy Computer Science modules, the pipeline may be adapted for practical programming tasks and code-based analysis. Finally, future empirical studies should measure whether this targeted remediation yields statistically significant learning gains, while exploring the framework's transferability to other technical and STEM disciplines.

X. CONCLUSION

Traditional testing in Computer Science education typically reduces student performance to a single numerical mark, hiding whether individual concepts were fully mastered, partially grasped, or completely missed. To address this diagnostic limitation, this study outlines a domain-specific NLP framework for analyzing free-form CS theoretical answers at a granular, concept level. Rather than evaluating a student's text as an indivisible unit, the platform isolates core theoretical statements and evaluates them directly against expected knowledge components.

The pipeline integrates standard NLP preprocessing, concept-extraction routines, SBERT-based semantic analysis, and a tri-state classification model (Correct, Partial, Missing) to systematically map learning gaps. These mapped deficiencies directly inform targeted remediation, delivering concept-specific explanations, targeted examples, and drill problems. The primary conceptual contribution of this model lies in bridging diagnostic assessment with active learning support: an answer's evaluation does not terminate with a static grade, but instead drives immediate, tailored educational assistance. This gives learners explicit guidance on specific conceptual weaknesses while offering instructors deeper diagnostic visibility into class-wide learning hurdles.

At present, the framework stands as a conceptual model and awaits practical implementation and empirical validation; consequently, its classification accuracy, operational reliability, and total educational impact remain to be established. Real-world testing using student responses, alongside rigorous expert review and statistical evaluation, will be necessary to benchmark performance. Overall, the proposed architecture establishes a clear blueprint for applying semantic NLP tools to transition beyond basic marks-based testing toward concept-aware, adaptive instruction in Computer Science education.

REFERENCES

[1] Botelho, S. Baral, J. A. Erickson, P. Benachamardi, and N. T. Heffernan, “Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics,” *J. Comput. Assist. Learn.*, vol. 39, no. 3, pp. 823–840, 2023.

[2] H. Chinni, Y. K. Sharma, S. Shevkari, J. L. Vydehi, Saurabh, and M. Kavitha, “Comparative evaluation of custom CNN architectures and transfer learning models for image classification in PyTorch,” in Proc. 2026 13th Int. Conf. Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2026, pp. 1–6, doi: 10.23919/INDIACom70271.2026.11525435.

[3] K. P. Kumar, S. Noureesh, T. Gundu, S. V. Reddy, and S. V. Kumar, “CORE-GAP: Concept gap detection system for personalized learning,” Int. J. Adv. Res. Sci. Technol. (IJARST), vol. 15, no. 5, 2026.

[4] Y. Aruna, M. Kavitha, M. Patil, S. Shevkari, A. Aggarwal, and Y. K. Sharma, “Deep learning based clinical decision-support web framework for heart disease prediction,” in Proc. 2026 13th Int. Conf. Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2026, pp. 1–6, doi: 10.23919/INDIACom70271.2026.11526620.

[5] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019, pp. 3982–3992.

[6] I. G. Ndukwe, C. E. Amadi, L. M. Nkomo, and B. K. Daniel, “Automatic grading system using Sentence-BERT network,” in Artificial Intelligence in Education (AIED 2020), Lecture Notes in Computer Science, vol. 12164, I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Millán, Eds. Cham, Switzerland: Springer, 2020.