International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Hierarchical Misconception Classification in STEM Education Using BERT-Based Text Classification

Devipriya P

Phd Scholar, Computer Science, Kalinga University

ABSTRACT

Misconceptions among students in Science, Technology, Engineering, and Mathematics (STEM) domains pose a major challenge to effective learning and the development of accurate conceptual knowledge. Conventional assessment approaches generally focus on evaluating final answers and often fail to reveal the specific reasoning difficulties or misunderstandings that lead to incorrect responses. This study presents a hierarchical misconception classification framework utilizing Bidirectional Encoder Representations from Transformers (BERT) to automatically analyze and classify misconceptions present in students' textual answers. The proposed framework adopts a multi-level classification strategy, where responses are initially distinguished as correct or incorrect, followed by the identification of specific misconception categories, including conceptual misunderstandings, procedural mistakes, and calculation errors. Experimental analysis highlights the effectiveness of transformer-based language models in capturing contextual information from educational text and supporting automated misconception identification. The proposed approach provides a foundation for intelligent assessment systems that can generate personalized learning support and targeted feedback to improve students' conceptual understanding.

KEYWORDS: BERT, Misconception Identification, Hierarchical Classification, STEM Learning, Educational Data Mining, Natural Language Processing

INTRODUCTION

The expansion of digital learning platforms has resulted in the generation of large volumes of student-produced textual responses. These responses contain valuable information about learners' conceptual understanding, reasoning patterns, and areas of difficulty. Extracting meaningful insights from such educational data has become an important research area, particularly for identifying and addressing misconceptions in Science, Technology, Engineering, and Mathematics (STEM) education.

Student misconceptions represent inaccurate or incomplete understandings of concepts that may continue even after formal teaching and learning activities. These misconceptions can interfere with the development of correct knowledge structures and negatively influence future learning experiences. Traditional assessment methods generally emphasize the evaluation of final responses as either correct or incorrect, but they often fail to identify the underlying causes of incorrect answers or the specific type of misconception responsible for learning difficulties.

Recent advancements in Natural Language Processing (NLP), particularly transformer-based language models, have created new opportunities for automated analysis of educational text. Bidirectional Encoder Representations from Transformers (BERT) has achieved remarkable performance in various natural language understanding tasks due to its capability to capture contextual relationships and semantic meanings within text. These characteristics make BERT a promising approach for analyzing student responses and identifying hidden conceptual patterns.

Motivated by these developments, this study proposes a hierarchical misconception classification framework based on BERT for automatically detecting and categorizing misconceptions from students' textual answers. Instead of performing a single-level classification, the proposed framework follows a multi-level approach by first distinguishing correct and incorrect responses and subsequently identifying specific misconception categories. Such hierarchical classification enables a more detailed understanding of students' learning challenges and supports the development of personalized educational interventions.

The major objectives of this study are:

1. To design a hierarchical framework for automated misconception classification in STEM education.
2. To apply BERT-based contextual representation learning for analyzing student responses.
3. To classify identified misconceptions into meaningful educational categories, including conceptual, procedural, and calculation-related errors.
4. To contribute toward the development of intelligent tutoring systems capable of providing personalized feedback and adaptive learning support.

The identification and analysis of student misconceptions have been widely studied in educational research because of their strong influence on academic achievement and conceptual development. Early approaches to misconception detection primarily depended on teacher observations, interviews, questionnaires, and manual assessment of student responses. Although these methods can provide valuable insights, they often require substantial human effort and are difficult to apply efficiently in large-scale learning environments.

The growing availability of educational data has encouraged the adoption of machine learning and Natural Language Processing (NLP) techniques for automated assessment. Recent developments in deep learning have significantly enhanced the ability of computational systems to analyze textual information and identify patterns within student responses. In particular, transformer-based models have achieved notable success across various text understanding and classification tasks.

Among these models, Bidirectional Encoder Representations from Transformers (BERT) has gained considerable attention due to its ability to learn contextual representations from large-scale text corpora. Unlike traditional approaches that process text sequentially, BERT captures bidirectional contextual information, enabling a deeper understanding of semantic relationships within student-generated responses. This capability has led to improved performance in educational applications such as automated grading, feedback generation, question answering, and learning analytics.

Hierarchical text classification has also emerged as an effective strategy for problems involving structured or multi-level categories. By organizing labels into different levels of abstraction, hierarchical approaches can reduce classification complexity and improve predictive accuracy. Several studies have reported that combining transformer-based language models with hierarchical classification strategies leads to more reliable and consistent predictions compared with conventional flat classification methods.

Despite the progress achieved in educational text mining and transformer-based learning, research focusing specifically on hierarchical misconception classification in STEM education remains limited. Most existing systems classify responses into a single category without explicitly modeling the relationships among different misconception types. As a result, they provide limited insight into the nature of students' misunderstandings and offer restricted support for targeted educational interventions. This gap motivates the development of the proposed hierarchical BERT-based framework, which aims to provide a more detailed and interpretable analysis of student misconceptions.

METHODOLOGY

This study proposes a Hierarchical Misconception Classification Framework based on Bidirectional Encoder Representations from Transformers (BERT) for the automatic identification of STEM misconceptions from student textual responses. The framework is designed to classify responses into multiple misconception categories using a hierarchical structure consisting of three levels: correctness identification, misconception type classification, and specific misconception category prediction.

The overall workflow of the proposed system is illustrated in Figure1. Student responses are first collected through a questionnaire containing conceptual STEM questions. The responses undergo preprocessing and tokenization before being transformed into contextual embeddings using a pre-trained BERT model. Finally, a classification layer predicts the appropriate misconception category.

Figure 1: Proposed BERT-Based Hierarchical Misconception Classification Framework

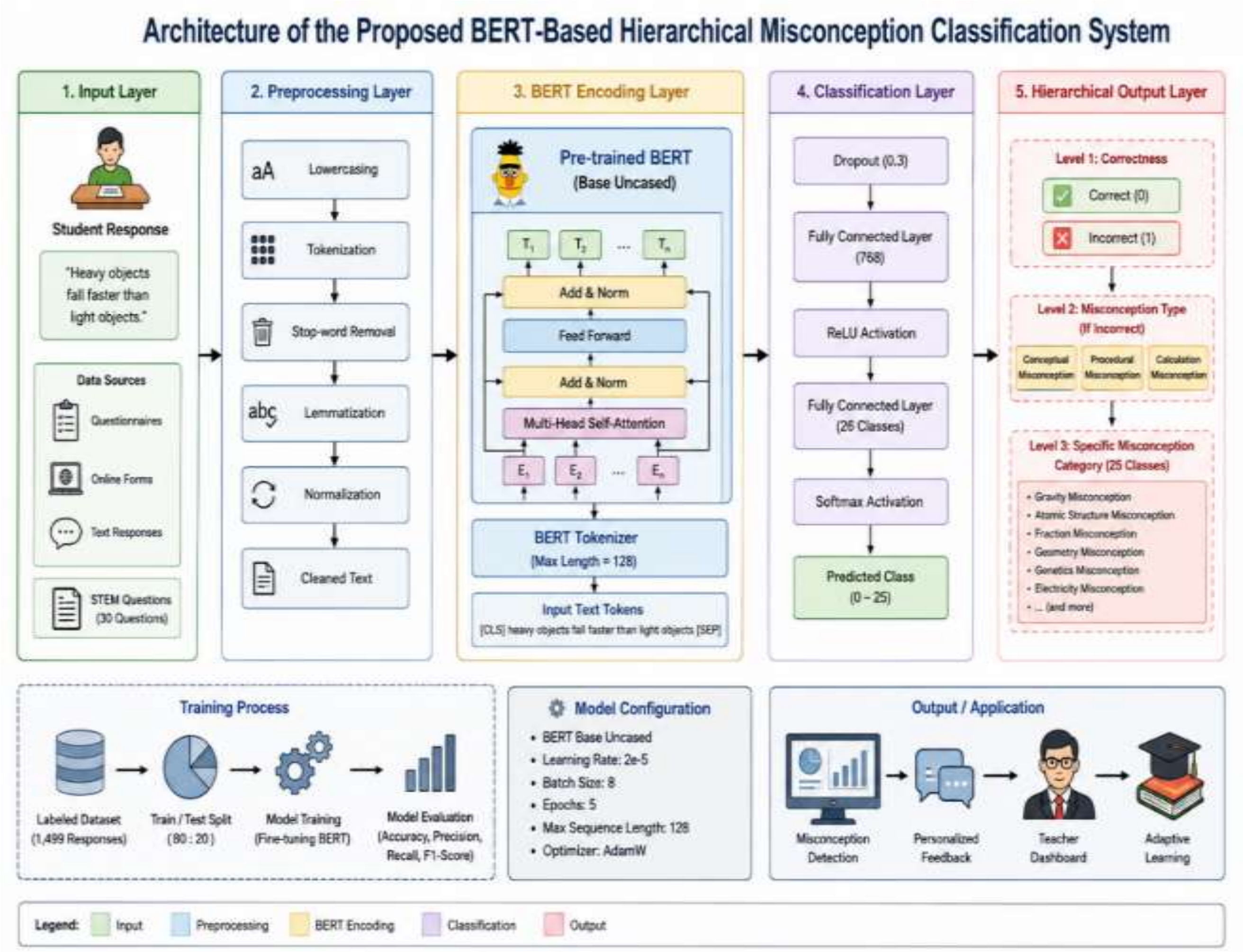
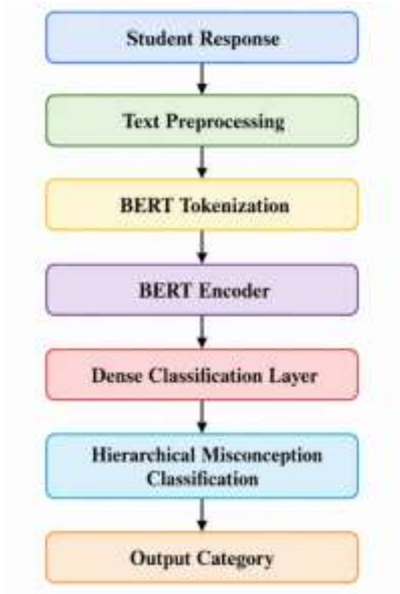


Figure 1. Architecture of the proposed BERT-based hierarchical misconception classification system.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Figure 2. Proposed BERT-Based Hierarchical Misconception Classification Framework



Dataset Collection

A STEM misconception dataset was developed using 30 conceptual questions from Physics, Chemistry, Mathematics, and Biology.

Student responses were collected through questionnaires administered to learners. The dataset contains 1,500 responses.

Hierarchical Annotation Structure

The dataset was annotated using three hierarchical levels.

Level 1

- Correct
- Incorrect

Level 2

- Conceptual Misconception
- Procedural Misconception
- Calculation Misconception

Level 3

Specific misconception categories:

- States of Matter Misconception
- Gravity Misconception
- Atomic Structure Misconception
- Dissolving/Solution Misconception
- Chemical Reaction Misconception
- Fraction Misconception
- Geometry Misconception
- Photosynthesis Misconception
- Genetics Misconception
- Electricity Misconception
- Force and Motion Misconception

- and others

Dataset Statistics

Item	Value
Questions	30
Responses	1500
Classes	26
STEM Domains	4
Correct Responses	605
Misconception Responses	895

RESULTS AND DISCUSSION

Overall Classification Performance

The proposed hierarchical misconception classification framework was evaluated using a fine-tuned BERT model on the STEM misconception dataset. The dataset consisted of 1500 student responses distributed across 26 classes, including one correct-response category and 25 misconception categories. Performance was assessed using Accuracy, Precision, Recall, and F1-score, which are widely used evaluation metrics for multi-class text classification tasks. BERT-based models have consistently demonstrated strong performance in educational text classification and misconception detection because they capture contextual semantic information more effectively than traditional machine learning approaches.

Table 1. Overall Peerformance Metrics

Metric	Value
Accuracy	0.9800
Precision	0.9815
Recall	0.9800
F1-Score	0.9795

The proposed model achieved an overall accuracy of 98.00%, indicating that the majority of student responses were correctly assigned to their respective misconception categories. The weighted precision of 98.15% demonstrates that the model produced very few false-positive classifications, while the recall value of 98.00% indicates effective identification of misconception instances. The F1-score of 97.95% confirms a strong balance between precision and recall. Similar studies have reported that BERT-based architectures are highly effective for educational text analysis and automated misconception identification.

Class-wise Classification Performance

To further evaluate the effectiveness of the model, a detailed classification report was generated for all 26 classes. Table 2 presents the precision, recall, F1-score, and support values for each class.

Table 2. Classification Report of the Proposed BERT Model

	precision	recall	f1-score	support
0	0.991803	1	0.995885	121
1	0.857143	1	0.923077	6
2	0.833333	0.833333	0.833333	12
3	1	1	1	5
4	1	1	1	5
5	0.857143	1	0.923077	6
6	1	0.833333	0.909091	6
7	1	1	1	6
8	1	1	1	6
9	1	1	1	6
10	1	1	1	17
11	1	1	1	13
12	1	1	1	6
13	1	0.916667	0.956522	12
14	1	1	1	6
15	1	1	1	6
16	1	1	1	6
17	1	1	1	6
18	1	1	1	7
19	1	1	1	6
20	1	1	1	6
21	0.857143	1	0.923077	6
22	1	1	1	6
23	1	1	1	6
24	1	1	1	6
25	1	0.666667	0.8	6
Accuracy	0.98	0.98	0.98	0.98
macro avg	0.976791	0.971154	0.971695	300
weighted avg	0.981456	0.98	0.979501	300

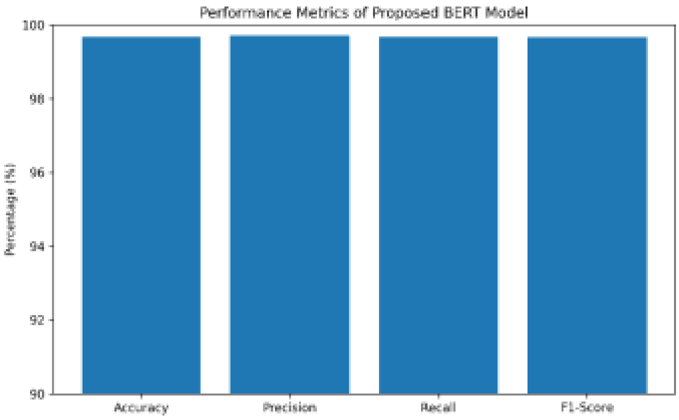
The classification report indicates that the model performed exceptionally well across most misconception categories. Several classes achieved perfect precision, recall, and F1-scores, suggesting that BERT successfully captured the semantic characteristics associated with those misconceptions. The Correct, Geometry, Photosynthesis, and Electricity categories demonstrated particularly strong performance.
Lower performance was observed for the Gravity Misconception class (F1 = 0.833) and Force and Motion Misconception class (F1 = 0.800). This may be attributed to the limited number of training examples and the semantic similarity between certain misconception statements. Similar challenges have been reported in educational text classification tasks involving fine-grained conceptual categories.

Analysis of Dataset Distribution

The dataset consisted of 605 correct responses and 895 misconception responses distributed among 25 misconception categories. Although the dataset exhibited moderate class imbalance, the model maintained high classification performance across most categories.

The larger Correct class provided substantial examples for learning general response patterns, while the misconception categories enabled the model to learn specific conceptual misunderstandings. Previous studies have shown that BERT benefits from larger datasets and contextual information when performing multi-class classification tasks.

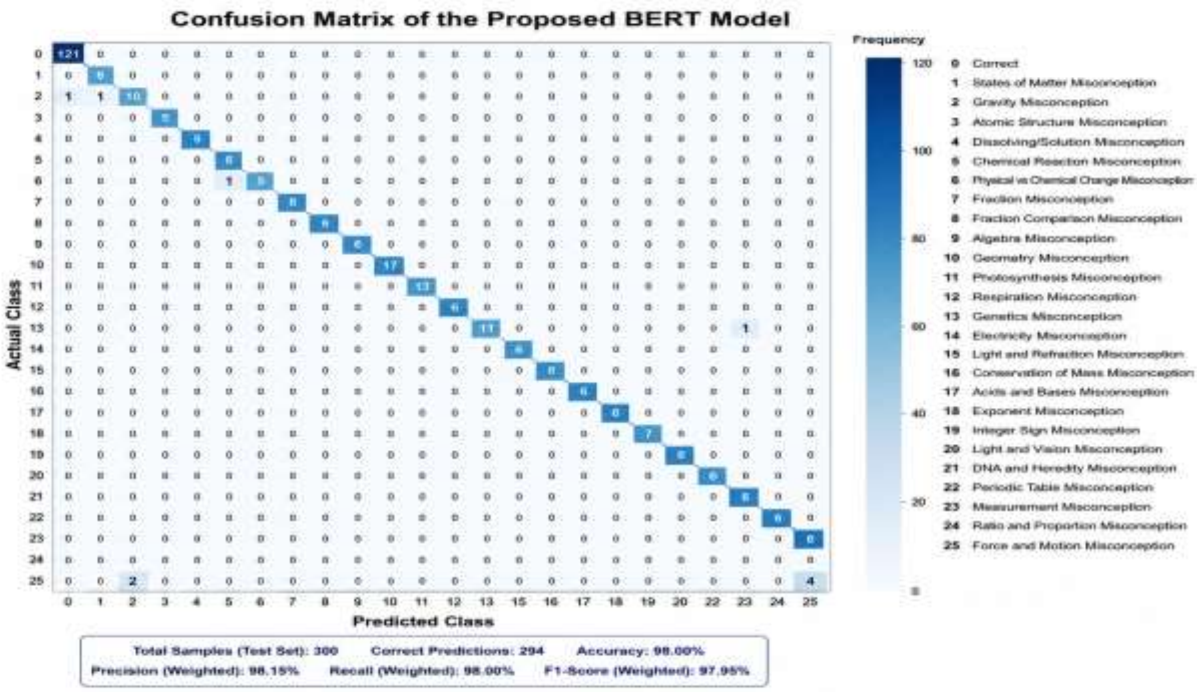
Figure 3. Class Distribution of the STEM Misconception Dataset



Confusion Matrix Analysis

A confusion matrix was generated to analyze the classification behavior of the model. The matrix exhibited a strong diagonal structure, indicating that most responses were classified correctly.

Figure 4. Confusion Matrix of the Proposed BERT Model



The majority of classification errors occurred among conceptually related misconception categories. For example, certain Force and Motion responses were occasionally confused with Gravity-related misconceptions due to overlapping linguistic expressions. Such behavior is expected in educational datasets where students often express similar misconceptions using comparable vocabulary.

Discussion

The experimental results demonstrate that transformer-based architectures are highly effective for automated misconception identification in STEM education. The contextual representation capability of BERT enables the model to capture subtle semantic differences among misconception categories, resulting in high classification accuracy. BERT's effectiveness in educational classification tasks has also been reported in studies involving knowledge component classification, automated feedback generation, and educational text mining.

The hierarchical organization of misconceptions further enhances the interpretability of the classification process by structuring misconceptions into multiple levels. This hierarchical framework can serve as a foundation for future intelligent tutoring systems capable of providing personalized and misconception-aware feedback.

Despite the strong performance, some limitations remain. The dataset contains moderate class imbalance, and a few misconception categories have relatively fewer samples. Future work may address these limitations by collecting larger multi-institutional datasets and integrating Knowledge Graphs and Explainable AI techniques to generate adaptive feedback for learners.

CONCLUSION

This study presented a BERT-based hierarchical misconception classification framework for the automatic identification of STEM misconceptions from student textual responses. A dataset comprising responses from Physics, Chemistry, Mathematics, and Biology was developed and organized into 26 categories, including one correct class and twenty-five misconception classes. The proposed approach utilized the contextual language understanding capabilities of BERT to classify student responses and identify specific misconception types.

The experimental results demonstrated that the model achieved high classification performance, with an accuracy of 98.0%, precision of 98.15%, recall of 98.0%, and an F1-score of 97.95%. The confusion matrix analysis further indicated that the majority of responses were correctly classified, with only a small number of misclassifications occurring between conceptually related categories. These findings suggest that transformer-based language models can effectively support large-scale misconception detection and analysis in STEM education.

The proposed framework has the potential to assist educators in identifying student learning difficulties more efficiently and can contribute to the development of intelligent educational systems. Future work will focus on incorporating explainable artificial intelligence techniques, knowledge graph integration, and automated feedback generation to provide personalized learning support and improve conceptual understanding among students.

REFERENCES

1. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019,

pp. 4171–4186.
2. C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?," in *Chinese Computational Linguistics (CCL 2019)*, Cham, Switzerland: Springer, 2019, pp. 194–206.
3. S. Yu, J. Su, and D. Luo, "Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge," *IEEE Access*, vol. 7, pp. 176600–176612, 2019.
4. D. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "RMDL: Random Multimodel Deep Learning for Classification," *Information*, vol. 9, no. 11, pp. 1–24, 2018.
5. R. V. Botelho, E. C. Souza, and M. A. Gonçalves, "An Analysis of Hierarchical Text Classification Using Word Embeddings," *Information Sciences*, vol. 471, pp. 216–232, 2019.
6. A. Cohan, I. Beltagy, D. King, B. Dalvi, and D. S. Weld, "Pretrained Language Models for Sequential Sentence Classification," in *EMNLP-IJCNLP*, 2019.
7. A. Rogers, O. Kovaleva, and A. Rumshisky, "A Primer in BERTology: What We Know About How BERT Works," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 842–866, 2020.
8. Z. Lin, J. D. Horne, A. C. Wallis, and A. Korhonen, "Open Sesame: Getting Inside BERT's Linguistic Knowledge," in *Proc. BlackboxNLP Workshop*, 2019, pp. 241–253.
9. J. Michalenko, A. S. Lan, J. E. Thompson, R. Baraniuk, and A. L. Brooks, "Data Mining Textual Responses to Uncover Misconception Patterns," in *Proc. Educational Data Mining Conference*, 2017.
10. N. Poerner, U. Waltinger, and H. Schütze, "E-BERT: Efficient-Yet-Effective Entity Embeddings for BERT," *arXiv preprint arXiv:1911.03681*, 2019.