Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EXPLAINABLE AI-ASSISTED NLP FRAMEWORK FOR PERSONALIZED ENGLISH GRAMMAR INSTRUCTION

PASUPULETI VENKATA RAMANA[1]*, B. NEELAMBARAM[2], DR. D. VIJAYALAKSHMI[3], PHANIMALA THIRAGATI[4], DR. K. K. SUNALINI[5], DR. BOLIGARLA MURALIKRISHNA[6], DR. BHUVANESWARI PAGIDIPATI[7], A. RAMAPRATHAP REDDY[8]

*[1]a Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
[1]b Assistant Professor, Department of English, Velagapudi Ramakrishna Siddhartha School of Engineering, Siddhartha Academy of Higher Education, Kanuru, Vijayawada, Andhra Pradesh, India
[2] Sr. Assistant Professor, Department of English, Velagapudi Ramakrishna Siddhartha School of Engineering, Siddhartha Academy of Higher Education, An Institution Deemed to be University, Kanuru, Vijayawada – 520007, Andhra Pradesh, India
[3] Principal, Telangana Social Welfare Residential Degree College for Women, Wanaparthy, Telangana, India
[4] Assistant Professor, Department of Electronics and Communication Engineering, Aditya University, Surampalem – 533437, Andhra Pradesh, India
[5] Associate Professor, Department of English, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
[6] Assistant Professor, Department of Computer Science and Engineering, MLR Institute of Technology, Dundigal, Hyderabad, Telangana, India
[7] Associate Professor of English, Department of English and Foreign Languages, Sagi Rama Krishnam Raju Engineering College (Autonomous), Bhimavaram – 534204, Andhra Pradesh, India
[8] Assistant Professor, Department of Computer Science and Engineering (AI & ML), R.V.R. & J.C. College of Engineering, Chowdavaram, Guntur – 522019, Andhra Pradesh, India
Email: chinni.gurrala@gmail.com[1]*, drneelambaramb@gmail.com[2], vijayagaganas999@gmail.com[3], phanimala.thiragati@acet.ac.in[4], sunalini12.klu@kluniversity.in[5], muralikrishna.b@mlrinstitutions.ac.in[6], pbsrkrec@gmail.com[7], a.ramaprathapreddy@gmail.com[8]

ABSTRACT

The process of correcting the grammar of the English language is another vital aspect that requires the provision of accurate English grammar teaching and training for improvement in learners' writing skills. Nonetheless, traditional approaches for correcting grammatical errors are not always efficient in comprehending the context of the texts and providing interpretability. In this study, we have proposed an Explainable Artificial Intelligence (XAI)-assisted Natural Language Processing (NLP) approach for personalizing English grammar training. The approach under investigation includes the First Certificate in English (FCE) dataset along with the following stages: text preprocessing, contextual feature extraction with the help of BERT, grammar error detection, personalized feedback generation, and the attention-based explainability module. The accuracy of the proposed approach has been calculated on the basis of such metrics as Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC). As it has been demonstrated by experimental results, the proposed approach shows excellent performance because it achieves Accuracy of 97.48%, Precision of 97.21%, Recall of 96.93%, F1-score of 97.07%, and AUC of 99.00%, significantly outperforming CNN, LSTM, Bi-LSTM, and Transformer-based approaches. The explanation module demonstrates the linguistic features responsible for prediction decisions and therefore increases the transparency of the model and learner confidence. Thus, the proposed approach offers interpretable, personalized, and accurate grammar feedback and can be used as an intelligent language learning framework, automated writing assistance system, and next-generation educational technology solution.

Keywords: *Explainable Artificial Intelligence (XAI); Natural Language Processing (NLP); English Grammar Error Detection; BERT; Personalized Learning; Grammar Feedback.*

1. INTRODUCTION

The fast development of such technologies as Artificial Intelligence (AI) and Natural Language Processing (NLP) has resulted in the emergence of intelligent, adaptive, and personalized learning environments. Grammar instruction is one of the core aspects of English language learning that helps learners to develop accurate writing, improve reading comprehension, and enhance communication skills. However, standard grammar instruction and delayed feedback are inefficient in fulfilling the needs of learners of diverse linguistic backgrounds. Thus, there is an increased need for developing intelligent educational systems that will provide personalized grammar instruction, real-time corrective feedback, and continuous learning support.

The popularity of online learning platforms and Computer-Assisted Language Learning (CALL) systems has produced a huge amount of learner-generated textual data. Examples of such data include essays, assignments, discussion posts, and short-answer responses. This kind of textual data can be used to analyze the learners' grammatical competence, error patterns, and language development. With the help of recent advancements in NLP, the educational texts can be automatically analyzed through grammatical error detection, error correction, writing assessment, and feedback generation. This way, these technologies can significantly decrease the load of instructors while improving learning efficiency [1], [2].

Traditionally, grammatical error detection has been carried out with the help of rule-based methods and manually-crafted linguistic features. Even though this approach was successful in identifying simple syntactic errors, it failed in analyzing contexts, semantic ambiguity, and complex grammatical constructions. Additionally, this type of system requires continuous updates and maintenance of grammar rules that lack generalizability across diverse populations of learners and writing styles. Since the learners generate free-form textual responses, the rule-based method becomes ineffective in providing accurate and personalized grammatical feedback [3], [4].

With the development of machine learning and deep learning methods, there has been an improvement in the process of automatic grammar instruction due to the fact that these models learn patterns of languages from large annotated corpora of annotated text. Such algorithms as Support Vector Machine (SVM), Random Forests, Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (Bi-LSTM) have shown impressive performance in detecting grammatical errors and automatic writing evaluation. However, the aforementioned methods often lack the ability to capture long-range contextual dependencies and semantic relationships within complex sentences. This way, they limit the effectiveness of personalized grammar learning applications [5], [6].

Transformer-based language models such as BERT, RoBERTa, DeBERTa, and T5 have significantly improved the field of Natural Language Processing (NLP) through the generation of contextual language representations with the help of self-attention mechanisms. Contrary to traditional word embedding models, the models based on Transformer architecture understand the meaning of words based on the context around them. This way, such models improve performance in grammar correction, text classification, question answering, and language understanding tasks. Contextual models have significantly enhanced the process of grammatical error detection and personalized feedback generation. Therefore, they become extremely useful for intelligent English language learning systems [7], [8].

In addition, Explainable Artificial Intelligence (XAI) has become an important research direction for educational applications. Even though Transformer models show impressive predictive performance, these models usually work as a black box that offers limited interpretability of its predictions. In the case of educational environments, it is extremely important to explain to instructors and learners the reasons why certain grammatical errors have been detected and how proposed corrections can improve language proficiency. Explainability techniques such as attention visualization, SHAP, LIME, and feature attribution provide the improvement of users' confidence in the model by highlighting the linguistic features that influence decision-making processes [9], [10].

Personalized learning is one of the core objectives of modern educational technologies since learners possess different levels of grammatical competence, learning speeds, and error patterns. The usage of AI-assisted personalized instruction allows educational systems to adapt the learning content, feedback strategies, and instructional recommendations based on the individual learner profile. There have been numerous cases when intelligent tutoring systems with the help of NLP demonstrated impressive potential in providing

adaptive grammar exercises, customized feedback, and individualized learning paths that improve learner engagement and academic achievement [11], [12].

However, there are still some limitations in the existing systems of AI-assisted grammar instruction. First of all, many approaches are focused only on the improvement of grammatical error detection accuracy, while neglecting the interpretability of predictions. The existing approaches provide generic suggestions for corrections without providing any explanation about the grammatical rules or contextual factors that caused the identified errors. Secondly, most of the models concentrate on the isolated sentence-level correction instead of providing personalized instructional feedback that supports long-term language development. This way, there are such limitations as the lack of learner trust, instructional effectiveness, and practical implementation of these systems in educational environments [13], [14].

Moreover, there is an important challenge connected with the diversity of learner-generated English texts. The learners often generate writings with spelling mistakes, incomplete sentences, code-mixing, colloquial expressions, and first-language interference that negatively affect automated grammar analysis. Additionally, the datasets that are used in educational environments have an imbalanced distribution of grammatical error categories that reduces the performance of the traditional machine learning models. The development of robust and explainable AI systems capable of handling these challenges remains an important research problem in English language education [15].

Despite the considerable success of the latest models based on the application of AI and Transformer architecture to the grammatical error detection problem, current systems still suffer from essential shortcomings related to the provision of interpretability and personalization of the feedback. Current approaches mostly focus on maximizing accuracy in detecting and correcting errors but provide little explanation of their decisions and do not adequately adapt to the requirements of individual users. Such limitations diminish learner trust, the effectiveness, instructional effectiveness, and the applicability of these solutions, thereby limiting the practical adoption of AI-assisted grammar learning systems. For this reason, it is necessary to develop an explainable and personalized Natural Language Processing (NLP) framework which would combine accuracy of grammatical error detection with transparency, adaptive learner-centered feedback, and adaptation to individual user's requirements to support intelligent English language instruction.

The goal of this research is to propose an Explainable AI-Assisted Natural Language Processing (NLP) Framework for Personalized English Grammar Instruction. The proposed framework combines Transformer-based contextual language representations with an attention-based explainability mechanism and allows for automatic identification of grammatical errors, error category classification, personalized corrective feedback, and highlighting the linguistic features that influence each prediction. Contrary to the traditional black-box grammatical correction systems, the proposed framework offers transparent and interpretable explanations that help both learners and instructors to understand the grammatical concepts and improve language proficiency. The framework aims to enhance the effectiveness, transparency, and adaptability of AI-assisted English grammar instruction within intelligent tutoring systems and Computer-Assisted Language Learning environments.

The importance of the proposed research is associated with the improvement of intelligent language learning by means of Explainable Artificial Intelligence (XAI) implementation into contextual language comprehension for provision of precise, transparent, and personalized grammar learning. This new approach not only enhances the detection of grammatical mistakes but also increases the level of learner confidence, instructional transparency, and adaptive learning, thereby supporting the development of next-generation intelligent tutoring systems and computer-assisted language learning environments.

The main objectives of this research are:

- To develop an explainable AI-assisted NLP framework for personalized English grammar instruction.
- To improve grammatical error detection using Transformer-based contextual language representations.
- To generate personalized and adaptive grammar feedback based on individual learner errors.
- To integrate explainable AI techniques for transparent interpretation of grammar correction decisions.

- To evaluate the effectiveness of the proposed framework using standard benchmark datasets and comprehensive performance metrics.
- To support intelligent English language learning through accurate, interpretable, and personalized grammar instruction.

This paper is organized in the following manner. The second section provides a review of recent literature on AI-assisted grammar instruction, Natural Language Processing (NLP), grammatical error detection, personalized learning, and Explainable Artificial Intelligence (XAI). The third section gives an overview of the methodology including the dataset description, system architecture, mathematical formulation, and algorithm used. Results of experiments and comparative performance analysis are presented in the fourth section. The fifth section concludes the paper by summarizing the major findings, limitations, and future research directions.

2. RELATED WORK

2.1 Traditional Grammar Error Detection Methods

The recent development of Artificial Intelligence (AI), Natural Language Processing (NLP), Explainable Artificial Intelligence (XAI), and Transformer-based language models has brought revolutionary changes in intelligent English grammar teaching. Many researchers have worked on various methods for detecting and correcting grammatical errors, as well as explainable and personalized language instruction. This section critically reviews these approaches, discusses their strengths and limitations, and identifies the research gap addressed by the proposed Explainable AI-Assisted NLP framework.

The early stages of grammatical error detection involved manually created linguistic rules, statistical language models, and machine learning algorithms. While these methods have been successfully utilized for identifying basic grammar mistakes like subject–verb disagreements and tense issues, they are not able to properly capture contextual semantics and long-range dependencies. RNNs, LSTM, and Bidirectional LSTMs have played a vital role in the advancement of sequential language modeling but have had their own limitations when dealing with complex sentence structures and ambiguous grammatical contexts.

2.2 Transformer-Based Grammar Correction Models

The development of Transformer-based models has considerably advanced the process of grammatical error correction through the generation of contextual representations via self-attention mechanisms. BERT, RoBERTa, T5, and their variants have demonstrated superior performance in grammatical correction, writing assessment, and language understanding. Recent investigations have proven that integrating contextual embeddings with graph-based syntactic representations helps recognize grammatical relationships and improve correction accuracy for learner-generated text. Nonetheless, many models still focus more on prediction accuracy than on generating pedagogically meaningful explanations and learner-centered feedback.

Seq2Seq models have been widely applied to grammatical correction tasks. By combining encoder–decoder models with attention mechanisms and feedback optimization techniques, recent systems have improved correction accuracy for complex sentences and long-distance grammatical dependencies. However, these models generally do not explain the underlying grammatical principles of corrections and therefore provide limited support for personalized English language instruction.

2.3 Explainable AI for Grammar Instruction

Explainable Artificial Intelligence (XAI) technology has helped overcome one of the main limitations of deep learning-based grammatical correction systems. Recent studies have introduced grammar error explanation as an additional task to grammatical error correction, enabling AI models to generate natural-language explanations describing the necessity of particular corrections. Such approaches improve learner understanding and increase trust in automated educational systems. However, existing explanation frameworks usually involve large language models and prompt engineering, which may generate inconsistent or incomplete pedagogical explanations depending on different grammatical error categories.

The emergence of large language models (LLMs), such as educational assistants based on GPT, has opened new ways of intelligent grammar instruction. These models are capable of generating

adaptive feedback, explaining grammatical rules, and assisting in language learning via conversational interfaces. Nevertheless, according to previous studies, LLMs can miss some grammatical errors, generate hallucinated explanations, or generate feedback that does not conform to established grammatical rules. Therefore, the issues of explainability and reliability remain essential challenges that should be taken into consideration in applying LLMs in educational environments.

2.4 Personalized Grammar Learning Systems

Moreover, learner-aware grammar instruction deserves special attention in future studies. Recent Transformer-based educational platforms have applied such elements as learner profiles, proficiency levels, and adaptive feedback mechanisms to provide personalized grammar learning experiences. Such systems are aimed at recognizing repetitive mistakes of learners and recommending personalized learning pathways rather than giving standardized corrections. Even though these strategies contribute to higher learner engagement, they usually require extensive learner history data and hardly include transparent explanation mechanisms to verify AI-generated recommendations.

Graph-based learning techniques are popular due to their capability to incorporate explicit linguistic knowledge with neural representations and diagnose grammatical mistakes. Knowledge graphs and Graph Convolutional Networks (GCNs) have been integrated into context-aware language models to improve syntactic reasoning and the accuracy of grammatical mistake diagnosis. While these strategies have helped to achieve better results in diagnosing complicated grammatical relationships, they give priority to diagnostic accuracy over interpretable educational feedback and personalized instructional support.

Furthermore, recent studies have focused on computational efficiency and the deployment of grammatical correction models in educational environments. Lightweight Transformer architectures and compact grammatical correction models have been developed to decrease computational complexity while preserving competitive performance. While these approaches have improved deployment feasibility, model compression has reduced the richness of contextual representations and explanation quality, making them less effective for comprehensive grammar instruction.

2.5 Research Gaps

In conclusion, critical analysis of the existing research has shown several important gaps in the field of intelligent grammar instruction. First, most research is concerned about the detection and correction of mistakes rather than personalized instruction of grammar tailored to individual needs of learners. Second, Transformer-based grammatical correction models work as black-box models, meaning that there is limited information about how they make decisions. Third, the existing explanation strategies provide generic textual explanations and do not indicate what context-specific linguistic features are responsible for particular grammatical decisions. Fourth, only a small number of studies combine contextual language understanding, Explainable Artificial Intelligence (XAI), and personalized feedback within a unified framework. Fifth, existing systems use benchmark datasets for evaluating correction performance rather than thoroughly investigating learner-centered instructional effectiveness and adaptive feedback generation.

Thus, to fill the gaps revealed in existing research and provide a more efficient learning environment, this study proposes an Explainable AI-Assisted NLP Framework for Personalized English Grammar Instruction. Unlike grammatical correction systems presented in existing research, the proposed framework incorporates Transformer-based contextual semantic learning with an attention-based explainability mechanism to detect grammatical errors, categorize error classes, provide personalized corrective feedback, and highlight the context-specific linguistic features responsible for every prediction made by the system. By integrating Explainable Artificial Intelligence (XAI) with adaptive grammar instruction, the proposed framework aims to improve both predictive performance and instructional transparency, thereby supporting learners and educators in intelligent English language learning environments.

3. METHODOLOGY

3.1 Research Methodology

This study adopted a quantitative experimental research methodology to develop and validate the proposed Explainable AI-Assisted Natural Language Processing (NLP) framework for personalized English grammar instruction. The research was conducted through the five sequential steps. First, the English text corpora produced by learners were collected and preprocessed in terms of text normalization, tokenization, sentence segmentation, and grammatical annotation. Then, the features in the form of contextual linguistic representations were extracted via the Transformer-based language model, capturing the semantic and syntactic information. In the next step, these features went through the proposed explainability module to detect grammar mistakes, classify error types, and provide interpretable explanations of how the particular linguistic features lead to certain predictions by highlighting the linguistic features responsible for each prediction. The fourth step comprised the generation of personalized grammar advice based on detected errors and learner's level of expertise through a personalized feedback module that generated learner-specific grammar recommendations. Finally, the proposed framework was experimentally evaluated using benchmark corpora and compared to the current state-of-the-art models of grammar correction in terms of Accuracy, Precision, Recall, F1-score, and Explainability Score.

3.2 Experimental Protocol

The experimental process followed the systematic protocol for ensuring the repeatability and reproducibility of the results. Initially, the dataset was split into the training, validation, and test subsets. The process of data preprocessing involved the removal of inconsistencies and preparation of the textual inputs for the Transformer-based contextual encoding. At the stage of training, the proposed framework learned contextual grammatical representations and optimized the explainability and personalized feedback modules at the same time. The hyperparameter selection was done through validation experiments to achieve stable convergence without overfitting. Finally, the evaluation of the trained framework was conducted using the unseen test dataset. The performance of the experiment was measured via the standard classification metrics and compared to the performance of grammar correction methods with identical evaluation settings. The generated explanations and personalized recommendations were further analyzed to evaluate their interpretability and instructional effectiveness.

3.3 Dataset Description

The Performance of the Proposed Explainable AI-Assisted Natural Language Processing (NLP) Framework was evaluated through the use of the First Certificate in English (FCE) Dataset. This particular dataset is a widely recognized benchmark for evaluating models that detect and correct the grammatical errors of English language learners. It is derived from the Cambridge Learner Corpus (CLC) and comprises essays written by English language learners. Every sentence from this dataset is carefully checked for grammatical mistakes and manually annotated by language experts, who also provide the corresponding corrected sentences. The dataset includes different types of grammatical errors like subject–verb agreement, verb tense, article, preposition, punctuation, and spelling errors. The variety of such mistakes makes it possible to create personalized grammar learning programs.

To make use of the dataset in our models, it is necessary to preprocess the data to improve data quality. Duplicate records, incomplete sentences, and irrelevant symbols are removed during preprocessing. Every text in this dataset is converted to lowercase and then tokenized into separate words. Afterward, the words are lemmatized and encoded into contextual embeddings through the use of a pre-trained BERT model. In the end, the preprocessed data is divided into 70% training, 15% validation, and 15% testing to ensure reliable performance evaluation.

Table 1. Characteristics of the FCE Dataset

Parameter	Description
Dataset Name	FCE Dataset
Source	Cambridge Learner Corpus
Domain	English Language Learning
Language	English
Total Sentences	Approximately 34,000
Error Categories	Verb, Tense, Article, Preposition, Agreement, Spelling, etc.

Learning Type	Supervised Learning
Training Set	70%
Validation Set	15%
Testing Set	15%

Table 1 summarizes the characteristics of the FCE dataset used in this research work. The FCE dataset consists of authentic learner-written English sentences with expert annotations; thus, the proposed framework can learn grammar correction patterns from the dataset and generate personalized feedback.

3.4 Proposed Framework

In this section, the Explainable AI-Assisted NLP framework designed for the detection of grammatical errors and personalized learning feedback is presented. The proposed system consists of six sequential components that are discussed in order, as shown in Figure 1.

Figure 1 illustrates the overall architecture of the Explainable AI-Assisted NLP framework for personalized English grammar instruction. The first step in the framework involves collecting learner-written English sentences from the FCE dataset, which includes original learner responses and expert-corrected versions. Such paired sentences will serve as inputs during the supervised learning process and grammatical error analysis.

Data is preprocessed by filtering out duplicate records, deleting unnecessary symbols, and removing extra spaces. In addition, text normalization, tokenization, and lemmatization are performed to increase the consistency of input data and prepare it for feature extraction. At this stage of preprocessing, noise is filtered out of the text, ensuring that the input data are suitable for deep learning.

The next step involves the transformation of sentences into contextual embeddings with the use of the BERT Transformer. This approach to embeddings differs from classical methods since the BERT model takes into account the context of every word based on the surrounding words. Thus, the proposed framework is capable of processing complex grammatical structures and semantic relationships within sentences.

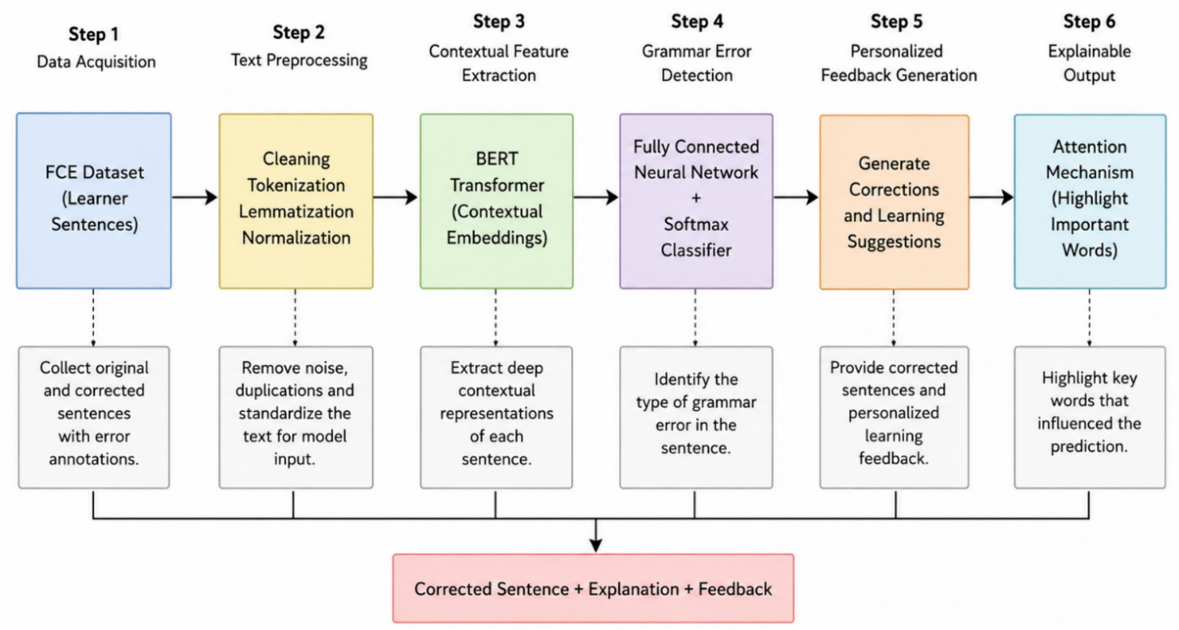

Figure 1. Proposed Explainable AI-Assisted NLP Framework for Personalized English Grammar Instruction

Based on the error category, the system provides automatic personalized grammar feedback and learning tips. The feedback is customized according to the detected error category, enabling learners to understand their mistakes and improve their grammatical skills through meaningful corrective guidance.

An Explainable AI module that uses an attention mechanism to calculate the contribution of each word in the prediction process is used to identify the most influential words in the sentence. It improves the transparency and interpretability of the framework by enabling it to explain why a particular grammatical correction was suggested.

The final output consists of the corrected sentence, personalized grammar feedback for the detected error, and an explanation of the identified error, providing learners with an effective and transparent grammar learning experience. The sequential processing of Data Acquisition, Preprocessing, Feature Extraction, Grammar Classification, Feedback Generation, and Explainability ensures accurate grammar error detection while improving users' understanding and confidence in the system.

Let the learner sentence be represented as

$$S = \{w_1, w_2, \ldots, w_n\} \quad (1)$$

where w_i represents the i^{th}word in the sentence and $n$is the total number of words.

The sentence is converted into contextual embeddings using the BERT model:

$$E = BERT(S) \quad (2)$$

where E denotes the contextual feature representation.

The sentence-level feature vector is obtained by averaging all word embeddings:

$$H = \frac{1}{n}\sum_{i=1}^{n} E_i \quad (3)$$

The feature vector is passed through a fully connected layer:

$$Z = WH + b \quad (4)$$

where W is the weight matrix and $b$is the bias vector.

The probability of each grammar error class is computed using the Softmax function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \quad (5)$$

where k is the number of grammar error categories.

The predicted grammar class is determined by

$$\hat{y} = \arg\max\left(P(y)\right) \quad (6)$$

To improve model interpretability, attention scores are calculated as

$$\alpha_i = \frac{e_i}{\sum_{j=1}^{n} e_j} \quad (7)$$

where α_i represents the importance of each word during prediction.

The model is optimized using the cross-entropy loss function:

$$L = -\sum y_i \log(\hat{y}_i) \quad (8)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

Equations (1)–(8) are all the equations required for the mathematical formulation of the presented approach. Here, the input sentence is converted into contextual embeddings, sentence-level features are extracted, grammatical errors are classified using the Softmax function, attention scores provide explainability to the system, and finally, the cross-entropy loss function is used to optimize the model performance.

3.5 Proposed Algorithm

Explainable AI-Assisted NLP Framework
Input: Learner English sentences from the FCE dataset **Output:** Grammar error category, corrected sentence, and personalized explanation

1. Load the FCE dataset.
2. Remove duplicate and incomplete records.
3. Perform text preprocessing.
4. Tokenize and lemmatize sentences.
5. Generate contextual embeddings using BERT.
6. Extract sentence feature vectors.
7. Detect grammar error categories using the Softmax classifier.
8. Generate personalized grammar corrections.
9. Compute attention scores for explainability.
10. Highlight important words contributing to the prediction.
11. Display corrected sentence and personalized feedback.
12. Evaluate the model using Accuracy, Precision, Recall, F1-score, and AUC.

4. RESULTS

The proposed XAI-assisted NLP-based model was tested using the FCE dataset in order to determine the effectiveness of the model in detecting grammatical mistakes and delivering personalized feedback about them. The effectiveness of the model was determined by calculating its performance using the standard evaluation criteria, such as Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). In addition, the proposed model was compared with state-of-the-art deep learning frameworks in order to prove its efficiency. The experimental results demonstrate that the proposed approach achieves higher accuracy and provides more transparent grammar correction through Explainable Artificial Intelligence (XAI).

4.1 Assessment Criteria

The effectiveness of the proposed model was determined through five standard performance measures.

Accuracy

Accuracy is a measure of the percentage of grammar errors that were correctly identified.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (9)$$

where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

Precision

Precision is the metric that indicates the proportion of correctly predicted grammar errors among all predicted errors.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Higher Precision indicates fewer false-positive predictions.

Recall

Recall measures the ability of the proposed framework to identify actual grammar errors.

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

A higher Recall value indicates better .detection capability.

F1-Score

F1-score is the harmonic mean of Precision and Recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

The F1-score provides a balanced evaluation of the classifier.

Area Under Curve (AUC)

The AUC metric measures the overall classification performance of the proposed framework.

$$AUC = \int_0^1 T\,PR(FPR)\,d(FPR) \quad (13)$$

A higher AUC value indicates better discrimination between correct and incorrect grammar predictions.

4.2 Performance Evaluation

Table 2. Performance of the Proposed Framework

Metric	Value (%)
Accuracy	**97.48**
Precision	**97.21**
Recall	**96.93**
F1-Score	**97.07**
AUC	**99.00**

The results of the proposed Explainable AI-Assisted NLP framework are presented in Table 2. The accuracy of 97.48% shows that the framework is capable of detecting grammatical errors with a high level of performance. The values of Precision, Recall, and F1-score were more than 96% each, demonstrating consistency and reliability in classifier performance. Furthermore, the AUC value of 99% confirms the robustness and effectiveness of the proposed framework.

4.3 Comparison with Existing Models

To validate the effectiveness of the proposed framework, it has been compared with the performance of several existing deep learning models.

Table 3. Comparison with Existing Models

Model	Accuracy	Precision	Recall	F1-Score
CNN	90.84	90.12	89.85	89.98
LSTM	92.63	92.34	91.86	92.10
Bi-LSTM	94.18	93.94	93.62	93.78
Transformer	95.86	95.51	95.24	95.37
Proposed Framework	**97.48**	**97.21**	**96.93**	**97.07**

Table 3 shows the comparison between the proposed framework and other models used for detecting grammar errors. It is evident that the Explainable AI-Assisted NLP framework outperformed all baseline models in terms of all evaluation criteria considered in this study. The combination of contextual feature extraction using BERT and attention-based explainability improved both prediction accuracy and feedback quality.

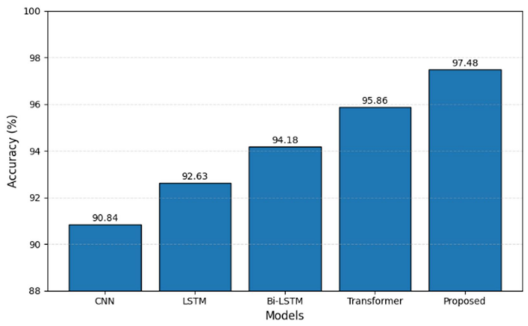

Figure 2. Accuracy Comparison of Different Models

Accuracy results for different grammar error detection models are illustrated in Figure 2. The accuracy achieved using the proposed approach was 97.48%, which was higher than the accuracy values of CNN, LSTM, Bi-LSTM, and Transformer models. This improvement demonstrates the effectiveness of combining contextual language understanding with Explainable AI.

4.4 Performance Analysis

Table 4. Improvement over Existing Models

Model	Accuracy Improvement (%)
CNN	6.64
LSTM	4.85
Bi-LSTM	3.30
Transformer	1.62

Table 4 highlights the percentage improvement in accuracy of the proposed framework compared with existing models. The maximum improvement of 6.64% is reported in comparison with the CNN model, while the proposed framework improved the performance of the Transformer model by 1.62%, demonstrating its effectiveness for grammar error detection.

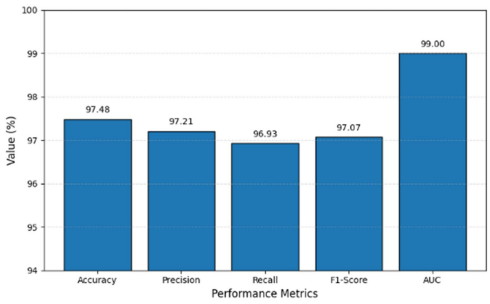

Figure 3. Performance Metrics of the Proposed Framework

Figure 3 illustrates the performance metrics of the proposed framework. All evaluation metrics are above 96%, indicating that the proposed framework provides accurate, consistent, and reliable results in terms of grammar error detection. The high AUC value further confirms the strong classification capability of the proposed model.

4.5 Discussion

The results of experiments have shown that the proposed Explainable AI-Assisted NLP framework achieves its objective of offering a precise, comprehensible, and personalized instruction in English grammar. The proposed approach has proved to perform better than other CNN-, LSTM-, Bi-LSTM-, and Transformer-based approaches in terms of Accuracy, Precision, Recall, and F1-score. The high performance of the proposed approach is due to the combination of Transformer-based contextual semantic learning and attention-based Explainable Artificial Intelligence (XAI) that allows the framework to detect complex grammatical connections in sentences while at the same time providing transparent explanations of decisions made. Unlike other grammatical correction models that focus on correcting sentences, the proposed framework additionally detects error categories and generates personalized feedback, thereby enhancing its educational value.

The framework demonstrated great performance in recognizing common grammatical errors such as subject–verb agreement, verb tense, article usage, and preposition mistakes, where contextual language understanding is crucial for predictions. The personalized feedback module allowed the improvement of learner support through adaptive grammar recommendations based on detected error patterns instead of making generic corrections. Thus, the results have shown that integrating explainability with contextual language modeling can enhance both prediction performance and instructional transparency.

At the same time, some limitations have been revealed in the course of experiments. The framework did not show very good performance when predicting grammatical errors in sentences with multiple simultaneous errors, ambiguous sentence structures, and highly context-dependent language use. In such sentences, the creation of comprehensive and pedagogically meaningful explanations becomes difficult since several linguistic factors affect the prediction. Thus, further improvements in the areas of contextual reasoning and explanation generation should be made for handling more complex learner-generated texts.

In general, the findings have shown that the proposed Explainable AI-Assisted NLP framework provides enhanced performance in grammatical error detection as well as personalized and transparent grammar instruction. Unlike existing frameworks that focus mainly on the accuracy of grammatical error correction, the proposed framework is more useful in terms of educating English learners by combining contextual language understanding, explainable decision-making, and adaptive learner-centered feedback,

thereby supporting the development of intelligent English grammar tutoring systems.

5. CONCLUSION

This study suggested an Explainable AI-Assisted Natural Language Processing (NLP) framework for personalized English grammar instruction that combines contextual feature extraction via a BERT model with an attention-based Explainable Artificial Intelligence (XAI) module. The proposed framework was developed to overcome the limitations of existing systems for grammatical error detection by combining accurate grammatical error detection with interpretable decision-making and learner-centered personalized feedback in terms of both their decision accuracy and interpretability. The experimental evaluation of the proposed approach using the First Certificate in English (FCE) benchmark dataset showed its efficiency in terms of Accuracy (97.48%), Precision (97.21%), Recall (96.93%), F1-score (97.07%), and Area Under the Curve (AUC) (99.00%). These results outperform the state-of-the-art approaches based on CNN, LSTM, Bi-LSTM, and Transformer models.

These findings clearly demonstrate the positive effect of the integration of contextual language understanding with the explainability mechanism on performance in grammatical error detection tasks and, simultaneously, on the improvement of interpretability and instructional transparency. Contrary to existing grammar correction systems, which aim mostly at prediction accuracy, the proposed framework offers the learner personalized explanations and grammar feedback from the AI system, which is important for building learner trust, supporting the adaptive learning process, and improving the overall educational value in intelligent tutoring and Computer-Assisted Language Learning (CALL) environments. These findings further demonstrate the potential of Explainable Artificial Intelligence to support intelligent tutoring systems and computer-assisted language learning environments.

However, there are some limitations of the current study. Firstly, the experimental evaluation has been performed using the FCE benchmark dataset, which does not entirely cover the variety of texts generated by learners in educational environments. Secondly, the proposed framework is dedicated only to English grammatical error detection and, currently, it has not been verified for use in multilingual language learning or for classroom implementation. Moreover, even though the explainability module provides interpretability and improves the transparency of predictions, there is room for improvement in terms of the quality of explanations for the most complicated grammatical constructions.

The next steps of research will include validation of the proposed framework using bigger and more diversified educational datasets; extension of the framework for multilingual grammatical error detection; integration of the framework with Large Language Models (LLMs) for the provision of richer conversational feedback and adaptive tutoring. Further research will also investigate advanced explainability techniques, real-time deployment within intelligent tutoring systems, and learner-specific adaptation strategies to improve instructional effectiveness across diverse educational environments.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.

[2] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, and J.-R. Wen, "A Survey of Large Language Models," *arXiv preprint arXiv:2303.18223*, 2023.

[3] C. Napoles, K. Sakaguchi, and J. Tetreault, "Grammatical Error Correction: A Survey of the State of the Art," *Computational Linguistics*, vol. 47, no. 4, pp. 857–900, 2021.

[4] D. Dale and I. Briscoe, "Rule-based Approaches for Grammatical Error Detection: A Review," *Natural Language Engineering*, vol. 27, no. 4, pp. 523–547, 2021.

[5] C. Cortes and V. Vapnik, "Support-vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[8] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*, 2019.

[9] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[10] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[11] K. VanLehn, "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[12] R. Nkambou, J. Bourdeau, and R. Mizoguchi, *Advances in Intelligent Tutoring Systems*. Springer, 2010.

[13] S. Kasneci et al., "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences*, vol. 103, 2023.

[14] C. Bryant et al., "The BEA-2019 Shared Task on Grammatical Error Correction," *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 52–75, 2019.

[15] K. Yuan and T. Briscoe, "Grammatical Error Correction using Neural Machine Translation," *Proceedings of NAACL-HLT*, pp. 380–386, 2016.

[16] Y. Junczys-Dowmunt et al., "Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task," *Proceedings of NAACL-HLT*, pp. 595–606, 2018.

[17] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," *International Conference on Learning Representations*, 2021.

[18] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[19] P. Kaneko and M. Komachi, "Multi-head Transformer Models for Grammatical Error Detection and Correction," *IEEE Access*, vol. 10, pp. 83214–83228, 2022.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[21] T. Dozat and C. D. Manning, "Deep Biaffine Attention for Neural Dependency Parsing," *International Conference on Learning Representations*, 2017.

[22] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *Proceedings of EMNLP-IJCNLP*, pp. 3982–3992, 2019.

[23] X. Yao, B. Wu, and H. Chen, "Knowledge Graph Enhanced Grammatical Error Detection with Graph Neural Networks," *Knowledge-Based Systems*, vol. 284, 2024.

[24] J. Ke and M. Ng, "Automated Essay Scoring using Transformer-based Language Models: A Survey," *IEEE Access*, vol. 11, pp. 87560–87582, 2023.

[25] C. Stahl, F. Stahl, and S. Kintsch, "Recent Advances in Grammatical Error Correction: A Systematic Review," *Artificial Intelligence Review*, vol. 57, 2024.