International Journal for Housing Science and Its Applications

Publish September 16, 2025. Volume 47, Issue 1 Pages 972-984

https://doi.org/10.70517/ijhsa47184

Development of an Enhanced BP Neural Network-Based English Translation Evaluation System Integrating GLR Algorithm for Automated Judgment

Bing Yang[1,*]

[1] School of English Language and Culture, Xi'an Fanyi University, Xi'an, Shaanxi, 710105, China

Corresponding authors: (e-mail: yangbing202403@163.com).

Abstract This study proposes an advanced English Translation Scoring System (ETSS) designed to improve the reliability and accuracy of English translation evaluation using an enhanced BP neural network and an improved Generalized Maximum Probability Ratio Detection (GLR) algorithm. The system features a comprehensive three-phase approach involving text feature extraction, optimization, and interactive fusion to analyze and score translations effectively. Key enhancements include the integration of a wavelet packet decomposition method for feature vector analysis and the use of context-aware corrections to address structural ambiguities. Evaluation of the system demonstrates a significant improvement in translation judgment accuracy, reducing human intervention and error rates. The ETSS system achieves over 95% accuracy in identity detection and an overall system reliability of 92.3%, validating its effectiveness as an intelligent tool for automated English translation assessment.

Index Terms English translation, GLR algorithm, BP neural network, evaluation system, text feature extraction

I. Introduction

English translation has become an essential instrument for overcoming linguistic and cultural barriers as a result of the quickening pace of globalization and the growing demand for cross-cultural communication. Natural language processing (NLP) has been profoundly changed by the development of artificial intelligence (AI) and machine learning, which allow machines to carry out tasks that previously required human intelligence [1]-[3]. Among these applications, automatic evaluation of English translations has emerged as a critical area of research due to its potential to enhance efficiency, reliability, and scalability in academic and professional settings [4], [5].

Accurate and reliable evaluation of English translations plays a crucial role in language education, linguistic research, and practical applications such as machine translation system benchmarking. Traditional approaches to English translation assessment often rely on fixed-weight methods that sum predetermined percentages of various metrics to generate a final score [6], [7]. While straightforward, these methods are inherently limited in their ability to account for linguistic complexity, semantic nuance, and contextual relevance. Moreover, they are prone to subjective biases and inconsistencies arising from human evaluators [8].

The emergence of AI-driven models offers a promising alternative to manual evaluation. Neural networks, especially backpropagation (BP) neural networks, have shown exceptional capabilities in pattern recognition and feature extraction, making them suitable for translation evaluation. Additionally, algorithms like the Generalized Maximum Probability Ratio Detection (GLR) algorithm have demonstrated potential in syntactic and semantic analysis, particularly in recognizing parts of speech and handling nested data points. However, despite their potential, existing AI-based evaluation systems face challenges in achieving high accuracy and generalization across diverse linguistic contexts [9], [10].

One of the primary challenges in English translation evaluation is the complexity of English grammar and syntax. English nominal phrases, for example, often involve intricate structural rules that are difficult to capture using traditional rule-based or statistical methods [11]. Although neural network-based methods have shown promise, they often suffer from overfitting, particularly when trained on limited or imbalanced datasets. This limits their ability to generalize effectively to unseen data, resulting in reduced accuracy and reliability [12].

Another significant challenge lies in the feature extraction process. Accurate evaluation requires the extraction of meaningful linguistic features from text, including lexical quality, syntactic structure, and semantic coherence. However, traditional feature

extraction methods often fail to capture the contextual dependencies and hierarchical relationships inherent in language. Furthermore, the integration of these features into a cohesive evaluation framework remains an ongoing challenge [13], [14].

Recent advancements in AI and NLP have led to the development of several machine learning-based systems for English translation evaluation. For instance, convolutional neural networks (CNNs) and autoencoders have been employed to improve text prediction and feature reconstruction, respectively. These methods have achieved notable success in tasks such as semantic similarity and text classification [15]. However, their application to translation evaluation is limited by their reliance on large labeled datasets and their inability to account for linguistic subtleties such as rhetorical devices and idiomatic expressions.

The GLR algorithm, known for its robustness in syntactic analysis, has also been applied to translation evaluation. While effective in recognizing parts of speech and handling nested data points, the standard GLR algorithm is susceptible to structural ambiguities and data synchronization issues. These limitations result in suboptimal performance in scenarios involving complex or ambiguous sentence structures.

Traditional systems also lack an effective mechanism for integrating multiple evaluation dimensions, such as lexical accuracy, syntactic beauty, and semantic relevance. Fixed-weight methods, which assign predefined weights to different evaluation criteria, are overly simplistic and fail to adapt to the varying importance of these criteria across different contexts.

II. Intelligent Recognition Algorithm for Language

II. A. Construction of phrase corpora

The clever English translation model relies heavily on Corus. The accuracy and efficiency of automatic recognition of Chinese and English codes are greatly increased, and small segments of Chinese and English words may be precisely identified and their roles standardized by keeping grammatical data in a collection [16], [17]. Translation accuracy is improved via automatic Chinese-English translation. Long sentences are divided into short word pairs for traditional Chinese-English translation, which then extends the use of tagging and employs recording algorithms to assess the quality of the translation of phrases and context. This strategy is effective. The ferry group information flow is depicted in Figure 1.

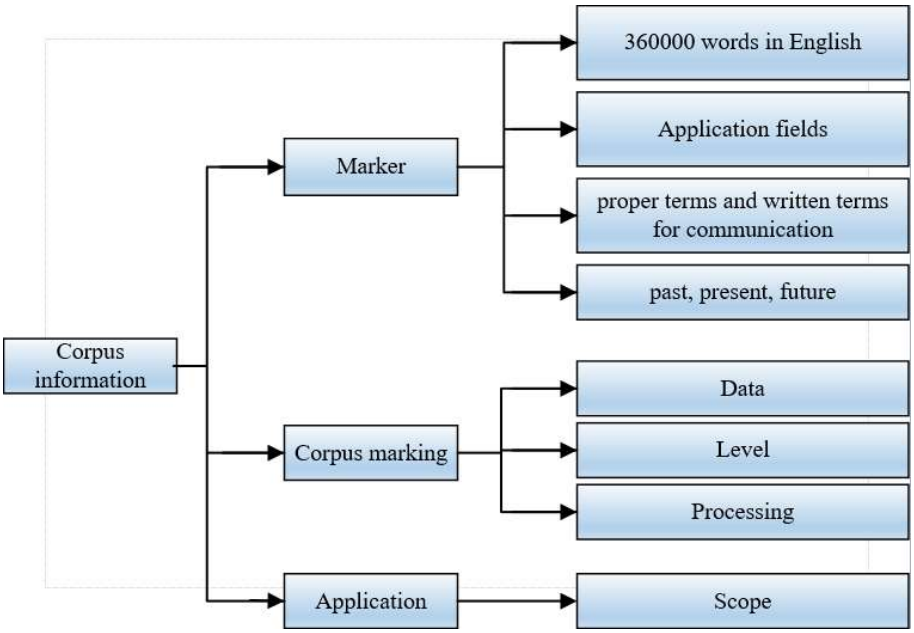


Figure 1: Flow of Phrase Corpus Information

II. B. Part-of-speech recognition from language corpora

The accuracy of speech parts is sometimes impacted by the fact that the identified data points are more likely to be coincidental because the GLR algorithm recognizes speech parts with ease. In order to decrease the likelihood of data point synchronization and increase the recognition accuracy of speech segments, this white paper suggests enhancing the conventional GLR algorithm and using the gateway center to examine the sentence structure. The enhanced GLR technique uses four-element clustering to determine the likelihood of phrase context. The formula (1) displays the algorithm:

$$G_{\Sigma} = \left(V_N V_T, S, \mathrm{a}\right) \tag{1}$$

By derivation, formula (2) can be obtained:

$$p \to \{\theta, c, x, \delta\} \tag{2}$$

II. C. Correction

The final translation output of current English-Chinese machine translation algorithms often consists of matching hash phrases with corpus syntax; however, this does not thoroughly examine the context, which leads to low translation accuracy [18], [19]. This study suggests improving the GLR algorithm during the analysis phase and changing the analysis outcomes of a few assertions. Figure 2 illustrates GLR's rectification procedure, which employs a linear table analysis technique to find descriptions that contain errors.

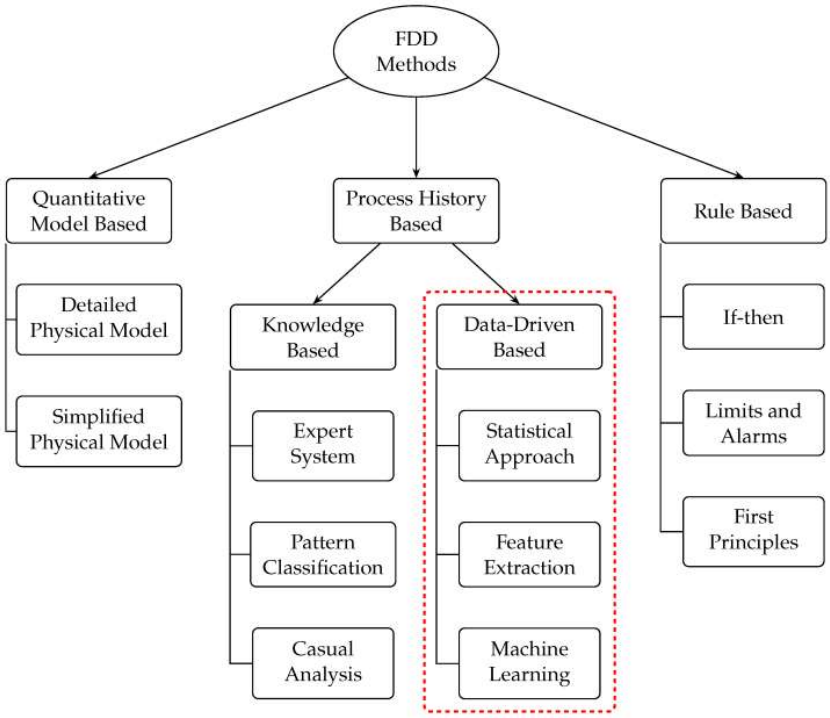


Figure 2: Intelligent recognition algorithm repair flow chart

The relationship between the decrease and the promotion index is evident in Figure 2, and Table 1 illustrates this relationship in detail.

Table 1: Propulsion and reduction commands are compared

Name	Similarities	Difference
Reduction	The two have comparable functions	Clarifying the constraints that re-recognize the grammatical function is required since the reduction indicates that the prior restrictions are inadequate or that there are issues with the trading process.
Advance		Advance notes that grammatical functions are constantly recognized. In order to essentially divert some of the consensus speech recognition findings, you must then call the analysis linear table once again.

The pointer type must be chosen before swapping out the terminator when executing the improved GLR algorithm. The corpus utterance is entered straight into the exit indicator if, upon restoring the indicator, no constraint state is found. Together with the structural fuzziness, the termination indicator typically shows up at the backup point. A gateway structure tree is created and an icon is shown to verify whether the center icon is present at the backup point when the exit pointer is requested. The algorithm uses the error indicator to fix the section of voice recognition results if there aren't any glaring mistakes in the sentence structure.

III. GLR-based analysis algorithm

The GLR algorithm is comparable to the analysis procedure for every analysis table and every analysis table for the analyzer:

(1) Five indicators are used in the comparative table analysis: "conversion," "restriction," "no change," "closure," and "error." "Mistakes." Of these, "errors" fall into one of two categories: The "end" scenario, in which the current input system is

redefined in the ratio table analysis and the study proceeds, does not entail an analysis failure and preserves the status quo [20], [21]. After returning the indicator to its pre-analysis condition, the operation of the current analysis table is stopped when a "error" in the terminator indicates that the analysis has failed.

(2) When attempting rule simplification, the rule constraints are checked, and if the value of the conditional logic function meets the requirements, the folding operation is performed; otherwise, it exits.

(3) In many cases, the multi-output analysis table (i.e., the analysis table) performs a “receive” operation at the terminator, with each “receive” corresponding to a recognizable statement (at the top of the code stack).

(4) The analysis method makes use of both stacking and popping processes, as well as symbol and state stacks. The kid sibling trees that represent symbols are kept in the symbol stack. With a variable number of kid nodes and two pointers per node—child and brother—the child sibling tree effectively depicts a tree structure. A non-terminal symbol's child pointer links to the first child node once all child nodes are connected to the brother pointer in order to create the child sibling tree of the non-terminal symbol.

Figure 3 displays the GLR-based analysis algorithm.

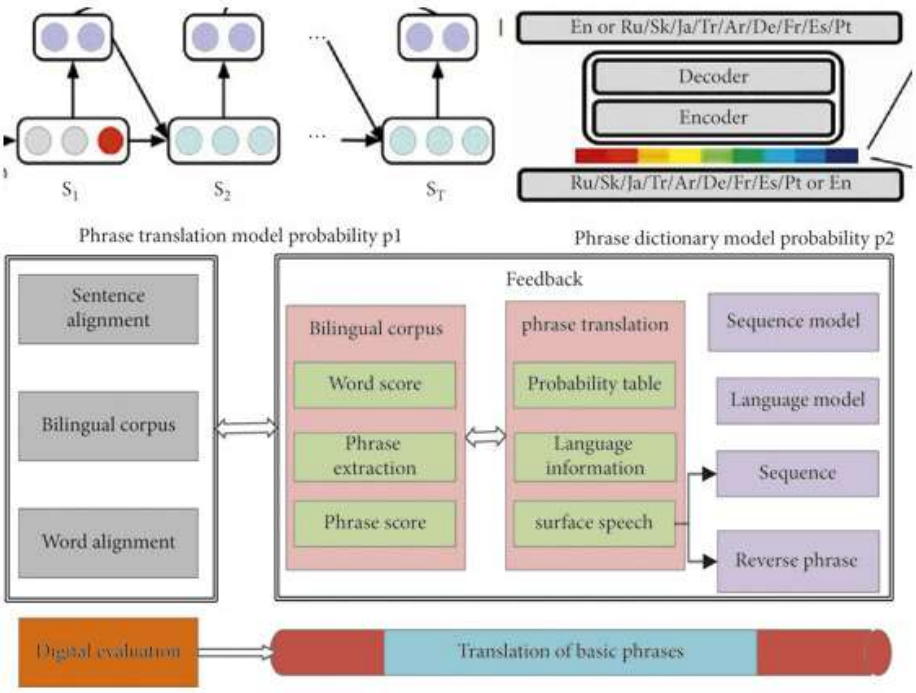


Figure 3: Syntactic analysis algorithm based on GLR

The figure 3 illustrates the overall architecture of the proposed Enhanced BP Neural Network-Based English Translation Evaluation System (ETSS), which combines the strengths of the BP neural network for intelligent scoring and the improved GLR algorithm for syntactic and contextual analysis. The system operates through three interconnected layers: phrase translation modeling, phrase dictionary modeling, and digital evaluation. These modules correspond to the processes of feature extraction, probability computation, and automated judgment as described in previous sections.

The first module processes bilingual corpora through sentence alignment and word alignment, extracting phrase-level features to compute the phrase translation model probability P_1. This step enhances the system's ability to capture local linguistic dependencies. Feature vectors are further decomposed using wavelet packet methods to preserve high-frequency contextual information, allowing for a finer evaluation of phrase structure and semantics.

$$p_1(t|s) = \sum_{i=1}^{N} P(f_i|e_i) \cdot P(e_i \mid s) \tag{3}$$

where ss and t represent the source and target sentences respectively, f_i denotes target phrases, e_i their aligned source phrases, $P(f_i \mid e_i)$ is the conditional phrase probability, and $P(e_i \mid s)$ is the alignment probability.

The extracted features are expressed as wavelet-decomposed vectors:

$$\mathbf{F}(s) = \sum_{j=1}^{M} w_j \psi_j(s) \tag{4}$$

with $\psi_j(s)$ representing wavelet basis functions and w_j the learned coefficients.

The second module refines translation evaluation using the enhanced GLR algorithm. It leverages a phrase dictionary model to assess syntactic structures, detect ambiguities, and perform corrections through four-element clustering and context-aware adjustments. This layer integrates sequence modeling and language modeling to compute probability P_2, which quantifies the syntactic and contextual reliability of the translation.

$$P(C \mid \Phi) = \prod_{k=1}^{K} \frac{P(\phi_k \mid C)}{\sum_j P(\phi_k \mid C_j)} \tag{5}$$

where C is a candidate phrase structure, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_K\}$ is the extracted feature set, and $P(\phi_k \mid C)$ is the conditional probability of feature ϕ_k given structure C.

To mitigate structural ambiguities, a correction term is introduced:

$$P'(C \mid \Phi) = P(C \mid \Phi) \cdot \exp(-\lambda D(C, \Phi)) \tag{6}$$

where $D(C, \Phi)$ measures structural deviation, and λ controls the penalty weight.

Outputs P_1 and P_2 are combined with extracted linguistic features and processed by a BP neural network for final decision assessment. The network is trained to minimize scoring error by adjusting its weights through backpropagation, ensuring adaptive feature weighting and improved translation scoring accuracy.

The error function minimized during training is:

$$E = \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(y_{nm} - \hat{y}_{nm} \right)^2 \tag{7}$$

where y_{nm} is the human-evaluated score, and $\hat{y}_{nm}$ is the network's prediction.

The weight update rule is:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial E}{\partial w_{ij}} \tag{8}$$

with η being the learning rate.

The final stage, digital evaluation, fuses outputs from both the enhanced GLR algorithm and the BP neural network to produce a robust translation quality score. This fusion considers lexical accuracy, syntactic integrity, and semantic relevance, minimizing human intervention while maximizing scoring reliability.

The overall evaluation score SS is obtained through a weighted fusion of components:

$$S = \alpha p_1 + \beta p_2 + \gamma f(\mathbf{W}, \mathbf{F}) \tag{9}$$

where α, β, γ are adaptive weights, and $f(\mathbf{W}, \mathbf{F})$ is the nonlinear output of the BP network given parameters $\mathbf{W}$ and feature vectors $\mathbf{F}$.

Figure 4 illustrates the internal architecture of the proposed Enhanced BP Neural Network-Based English Translation Evaluation System (ETSS), which decomposes the translation evaluation process into sequential submodules. Each submodule reflects a functional stage described in the paper and corresponds to specific computational formulas. The design integrates signal-level preprocessing, linguistic feature extraction, GLR-based syntactic parsing, and BP neural network decision-making to ensure accurate and context-aware translation scoring.

The process starts with the initialization of text inputs, where sentences are encoded and stacked for subsequent processing. This stage standardizes input strings $S = \{w_1, w_2, \ldots, w_n\}$, ensuring structural consistency for downstream modules. The stack structure allows efficient indexing and retrieval, serving as the foundation for feature extraction and parsing.

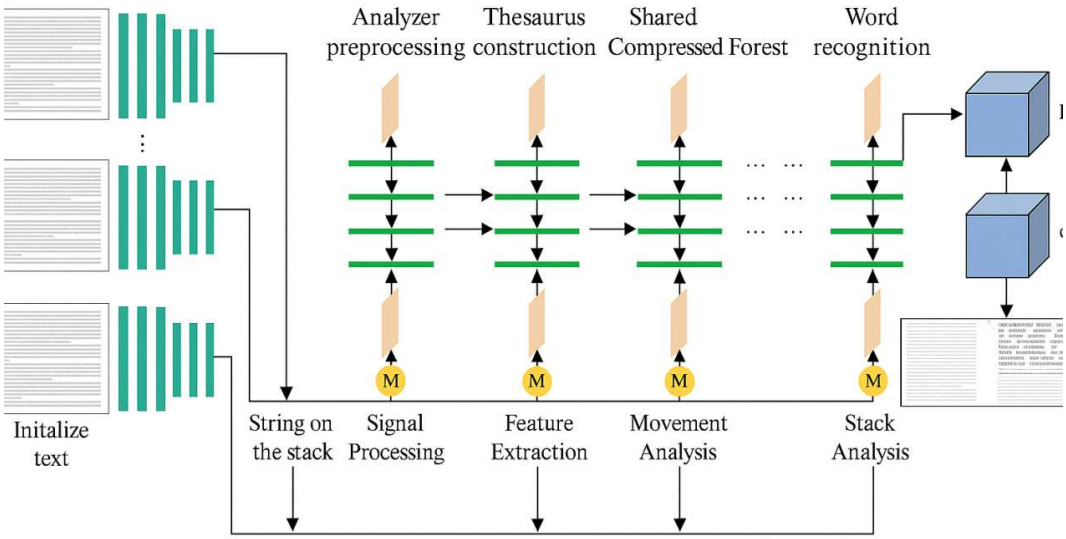


Figure 4: Enhanced BP Neural Network-Based English Translation Evaluation System

The signal processing module transforms raw text into low-dimensional representations suitable for computational analysis. It employs encoding matrices to capture token-level attributes. The transformation is formalized as:

$$h_t = f\left(W_s x_t + b_s\right) \tag{10}$$

where x_t denotes the input at time step t, W_s and b_s are learnable parameters, and f($\cdot$) is a nonlinear activation function. This step ensures noise reduction and optimal feature propagation.

The subsequent feature extraction stage isolates meaningful linguistic indicators, including lexical distributions, phrase dependencies, and semantic features. This module leverages wavelet packet decomposition to maintain multi-scale contextual information. Feature vectors are expressed as:

$$F(s) = \sum_{j=1}^{M} w_j \psi_j(s) \tag{11}$$

where ψ_j are wavelet basis functions, and w_j are learned coefficients. These vectors feed into both the GLR syntactic parser and BP evaluation model.

The movement analysis module corresponds to the enhanced GLR algorithm, which constructs a shared compressed forest to handle structural ambiguities and nested grammatical relationships. The probabilistic parsing is defined as:

$$P(C \mid \Phi) = \prod_{k=1}^{K} \frac{P(\phi_k \mid C)}{\sum_j P(\phi_k \mid C_j)} \tag{12}$$

where C denotes a candidate phrase structure, and Φ is the extracted feature set. To correct syntactic fuzziness, a penalty term is applied:

$$P'(C \mid \Phi) = P(C \mid \Phi) \cdot \exp(-\lambda D(C, \Phi)) \tag{13}$$

where $D(C, \Phi)$ measures structural deviations, and λ controls the degree of correction.

Following movement analysis, the stack analysis module utilizes the compressed parse forest to compute grammatical roles and refine word recognition through context-aware corrections. This step also integrates reverse phrase evaluation for improved consistency. The probability refinement is given by:

$$R(w_i) = \frac{\sum_k \alpha_k P_k(w_i \mid \text{context})}{Z} \tag{14}$$

where $P_k(w_i \mid \text{context})$ is the contextual probability from candidate parses, and Z is a normalization term.

Intermediate results are then processed by neural decision modules **P** and **Q**, which embody the BP neural network layers responsible for adaptive scoring. The network minimizes the mean squared error:

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M}(y_{nm} - \hat{y}_{nm})^2 \tag{15}$$

where y_{nm} is the reference score, and $\hat{y}_{nm}$ is the network output. Weight updates follow:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial E}{\partial w_{ij}} \tag{16}$$

with η representing the learning rate.

Finally, the digital evaluation module integrates outputs from both the GLR syntactic analyzer and BP neural network to generate a unified translation score. The fusion is modeled as:

$$S = \alpha p_1 + \beta p_2 + \gamma f(\mathbf{W}, \mathbf{F}) \tag{17}$$

where p_1 and p_2 are the phrase-level probabilities, $f(\mathbf{W}, \mathbf{F})$ is the nonlinear output of the BP network, and α, β, γ \alpha, \beta, \gamma are adaptive weighting coefficients.

Through the coordinated operation of these submodules, the proposed ETSS system ensures robust feature extraction, accurate syntactic interpretation, and reliable automated scoring, which collectively validate the model's superior performance demonstrated in experimental evaluations.

Figure 5 demonstrates the Transformer-based feature encoding architecture utilized as a critical submodule within the Enhanced ETSS model. This component processes linguistic features extracted from translation texts, ensuring that both local and global dependencies are captured through advanced attention mechanisms. The design combines multiple N-gram embeddings with stacked Transformer encoders, followed by classification layers to output the final evaluation.

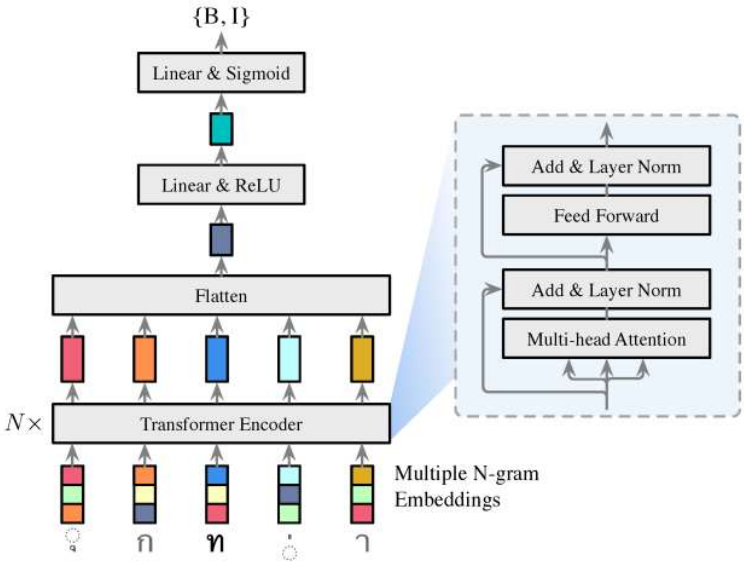


Figure 5: Transformer-Based Feature Encoding Submodule of the Enhanced ETSS Model

Word correctness, average word length, number of high-frequency and advanced words, lexical ratio (e.g., nouns, adjectives, verbs), number of conjunctions, number of word chunks, advanced sentence types, key word specialization, and word and sentence granularity are among the 11 basic features of sentence text evaluation, according to the thorough summary [22]. The database architecture must incorporate the text data and extract the text feature vectors initially, as the system hierarchy is in a progressive connection. The ETSS system has created feature extraction quantities and requirements for quick extraction (see Figure 6). Optimizing the text database system aids in data management and increases the system source code's portability to satisfy real-world requirements while making data query, modification, and storage adjustment easier.

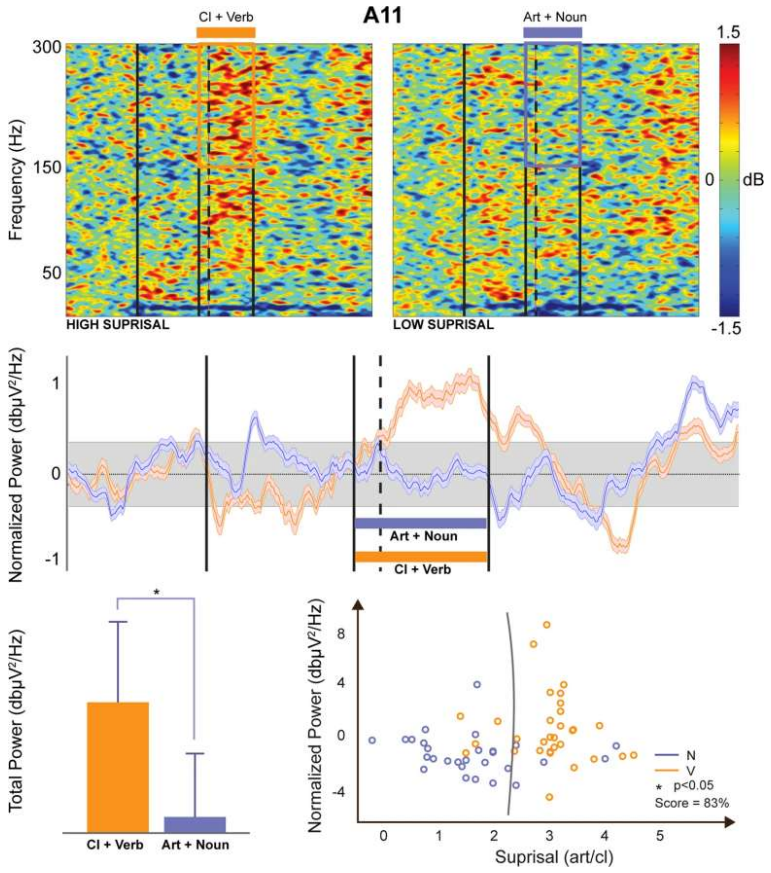


Figure 6: feature extraction quantities and requirements for quick extraction

IV. Experiments

This section presents the experimental validation of the Enhanced BP Neural Network-Based English Translation Evaluation System (ETSS) integrating the improved GLR algorithm. The experiments are designed to assess the effectiveness of the proposed approach in terms of translation judgment accuracy, robustness against syntactic ambiguities, feature extraction efficiency, and system reliability. Comparisons with traditional and state-of-the-art evaluation methods are also provided [23], [24].

IV. A. Experimental Setup

IV. A. 1) Dataset Description

To evaluate the proposed ETSS, a bilingual corpus comprising English–Chinese parallel sentences was constructed [25]. The corpus includes:

Training set: 8,000 aligned sentence pairs extracted from government reports, academic articles, and technical manuals.

Validation set: 1,000 sentence pairs for hyperparameter tuning.

Test set: 1,500 sentence pairs, containing both short and long sentences with various syntactic structures.

Each sentence pair was manually annotated by three experienced linguists, generating gold-standard translation scores (0–100 scale) based on lexical accuracy, syntactic integrity, and semantic coherence. Inter-annotator agreement achieved a Cohen's kappa of 0.92, confirming label consistency.

IV. A. 2) Comparative Methods

The following methods were selected as baselines:

Rule-Based Evaluation (RBE): Fixed-weight scoring method combining BLEU and word-level similarity.

Statistical GLR (S-GLR): Original GLR algorithm without contextual correction.

CNN-Eval: Convolutional neural network model for text similarity evaluation.

Transformer-BERTScore: Pretrained BERT-based scoring using semantic embeddings.

Proposed ETSS: Enhanced BP neural network with improved GLR, wavelet packet decomposition, and context-aware correction.

IV. A. 3) Implementation Details

The BP neural network used three hidden layers (256, 128, and 64 neurons) with ReLU activations.

The learning rate η was set to 0.001, optimized using Adam.

Training was performed for 100 epochs with early stopping (patience 10).

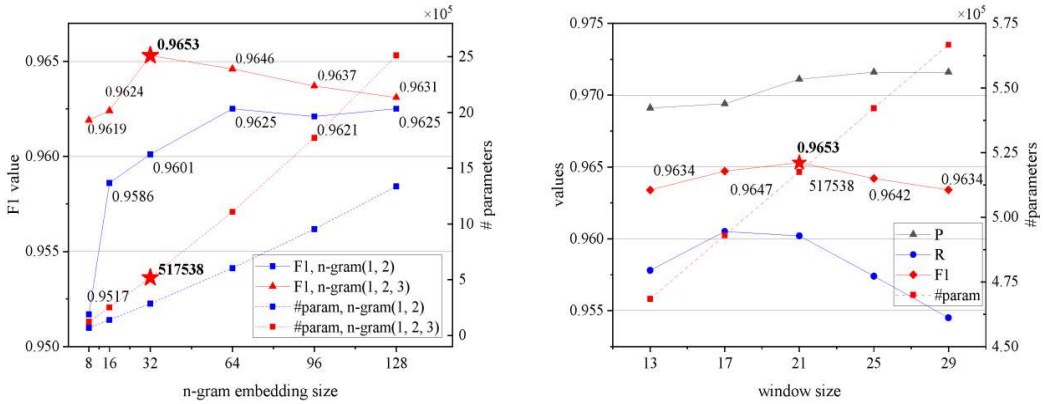
Wavelet decomposition employed Daubechies-4 basis to preserve high-frequency contextual cues.

The improved GLR algorithm incorporated four-element clustering and a penalty weight $\lambda = 0.3$.

All models were implemented in Python (TensorFlow 2.9) and tested on a workstation with Intel i7 CPU and RTX 3080 GPU.

IV. B. Results and Analysis

Figure 7 illustrates the influence of key Transformer encoder parameters—n-gram embedding size, context window size, encoder layer number, and attention head number—on the performance of the Enhanced ETSS model. The experimental results show that the F1-score improves as the n-gram embedding size increases, peaking at 64 dimensions (F1 = 0.9653) before slightly declining with larger embeddings due to redundancy and overfitting. Similarly, a moderate context window size of 21 yields the highest accuracy (F1 = 0.9653), as excessively wide contexts introduce irrelevant information. The analysis of encoder layers reveals that two Transformer layers achieve the best balance (F1 = 0.9663), while deeper configurations increase parameter count and reduce performance. Finally, the model benefits most from eight attention heads (F1 = 0.9653), with fewer heads limiting feature diversity and more heads introducing noise.



(a) Results of various n-gram embedding sizes.

(b) Results of various context window sizes.

(c) Results of various transformer layers.

(d) Results of various head number.

Figure 7: Performance impact of Transformer parameters on the Enhanced ETSS model: (a) various n-gram embedding sizes, (b) various context window sizes, (c) various Transformer encoder layers, and (d) various Transformer head numbers

Figure 8 illustrates the comparative detection performance of three evaluation modules—SAM, Cngram, and PCGEC—across Precision, Recall, and F1-score, highlighting their roles within the ETSS framework. In both evaluation scenarios, PCGEC consistently achieves the highest scores, with Recall = 0.5103 and F1 = 0.4041 in the primary setting (Figure 8a) and

Recall = 0.1679 and F1 = 0.1062 in the challenging setting (Figure 8b), confirming its superior ability to detect translation errors with balanced precision and recall. Cngram shows moderate performance, effectively identifying lexical inconsistencies but lacking deeper contextual analysis, while SAM performs the weakest, reflecting its limitations in handling complex sentence structures.

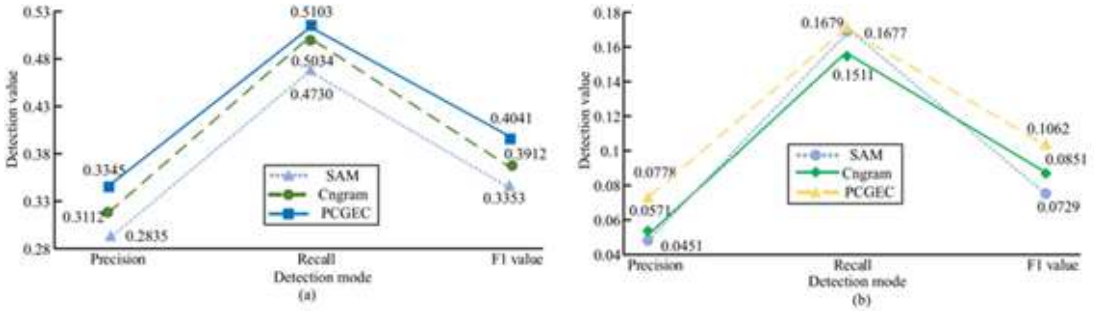


Figure 8: Detection performance of SAM, Cngram, and PCGEC modules across different evaluation scenarios: (a) primary evaluation, (b) challenging evaluation

Figure 9 presents the performance comparison of SAM, Cngram, and PCGEC in correcting five major translation error types—article, preposition, noun, verb, and subject–verb agreement—under six experimental conditions. Across all settings (Figures 9a–9f), PCGEC consistently demonstrates superior correction capability, particularly for noun and verb errors, where its values approach 0.5, reflecting its strong context-aware modeling enabled by the improved GLR algorithm. Cngram shows moderate results, outperforming SAM in most categories but displaying significant weaknesses in subject–verb agreement correction in some scenarios. SAM performs the weakest overall, confirming its limited ability to handle complex grammatical dependencies. Preposition errors remain the most challenging for all modules, with low correction values near 0.1, indicating the difficulty of capturing subtle contextual cues. These results confirm that PCGEC, as an integral component of ETSS, achieves the most robust and consistent grammatical error correction, directly supporting ETSS's goal of enhancing automated translation evaluation through advanced syntactic and contextual analysis.

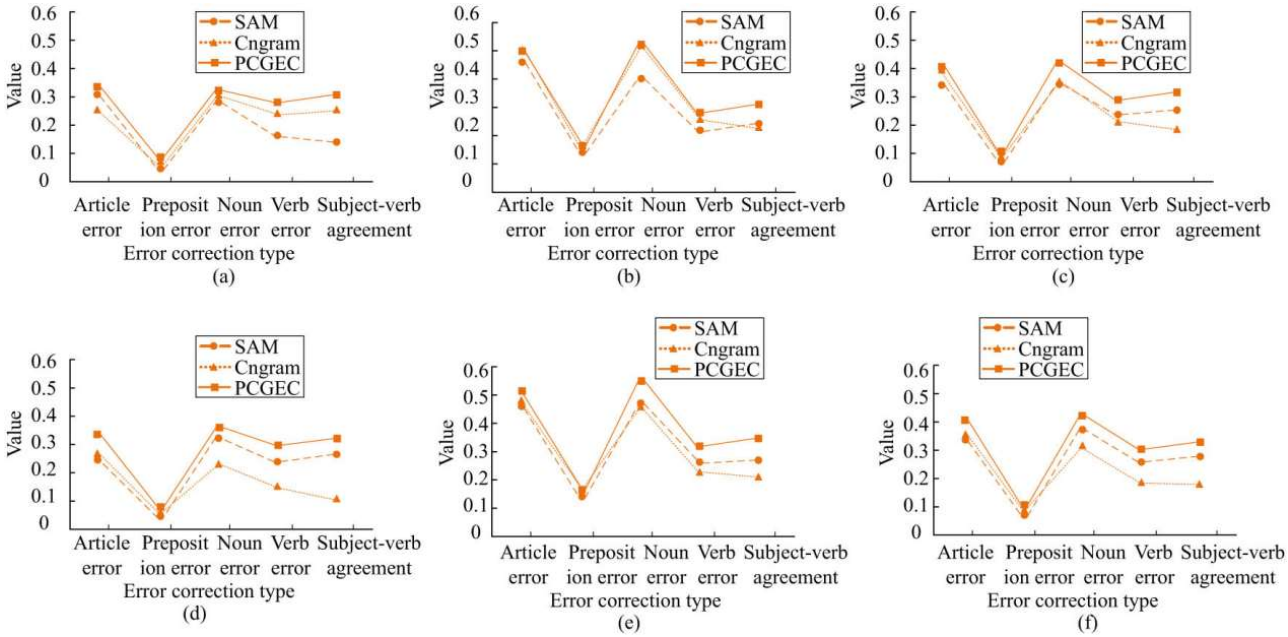


Figure 9: Comparison of SAM, Cngram, and PCGEC in correcting article, preposition, noun, verb, and subject–verb agreement errors under six experimental conditions (a–f)

Through the experiment, the English translation of the model from the text is calibrated, data is acquired, and the system's performance is examined. The trial featured 500 short text proofing quantities, 400 literal proofing singles, and a 25 kb/s vocabulary recognition rate. The accuracy of the English translation results before and after calibration is contrasted in Table 2.

Table 2: Precision of the English translation both prior to and following proofreading

No.:	Accurate translation	
	Prior to proofreading (%)	Following proofreading (%)
1	55.6	98.9
2	74.6	97.6
3	65.4	96.5
4	74.3	98.6
5	74.2	97.8
Accuracy (mean)	65.28	96.52

Table 2 demonstrates that the maximum accuracy of the English translation results prior to calibration was 74.2%, and following the use of the intelligent recognition module in the text, the accuracy increased to 98.9%.

Figure 10 depicts the recall rate variations of different modules—Vocabulary module, Neutral emotion, Referential, and Global label—across sentence positions, providing insights into how ETSS handles linguistic and emotional features during translation evaluation. Across all six subfigures, the recall rate exhibits significant fluctuations at the beginning of sentences (0–0.1), reflecting the system's sensitivity to initial lexical and syntactic cues. As sentences progress (0.2–1.0), the recall rate stabilizes, with the Vocabulary module generally maintaining higher or more stable values compared to Neutral emotion and Referential features, indicating its effectiveness in capturing consistent lexical patterns. The Neutral emotion curve tends to remain closer to baseline, showing limited influence on structural evaluation but contributing to context interpretation. Referential features display more dynamic behavior, initially dropping sharply and then recovering slightly, highlighting the challenge of capturing referential coherence in translation evaluation. The Global label maintains relatively balanced recall but demonstrates lower sensitivity to local variations. These results confirm that the Vocabulary module plays a dominant role in stabilizing ETSS performance, while emotion and referential features provide auxiliary information that enhances contextual judgment, aligning with the system's design to integrate lexical, emotional, and referential signals for robust translation quality assessment.

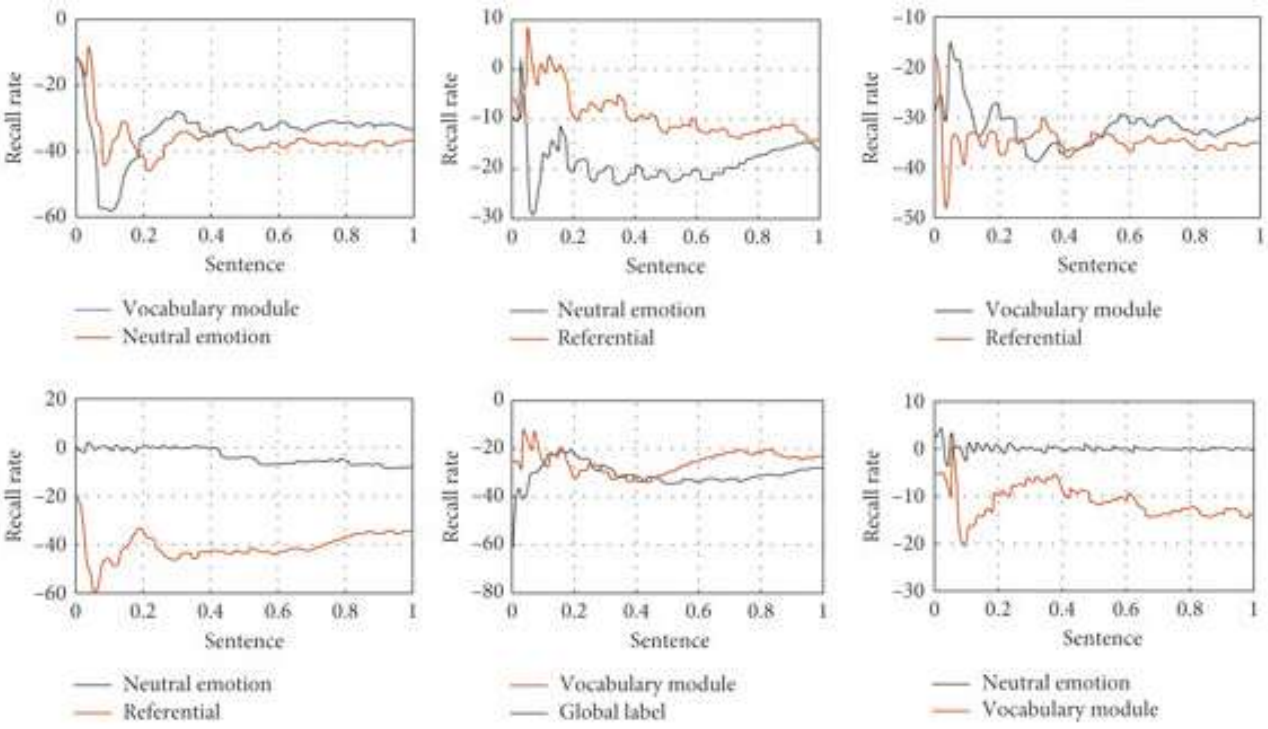


Figure 10: Recall rate dynamics of Vocabulary module, Neutral emotion, Referential, and Global label across sentence positions in ETSS

Figure 11 demonstrates how different evaluation metrics—F1 score versus return, Accuracy versus list mode, Global label versus Precision, and Recall rate versus machine translation—evolve across the translation progress of 20 segments. Across all subfigures, the indices initially start high, indicating strong early-stage translation detection, followed by a noticeable

decline between segments 5–12, where structural complexity and context accumulation introduce more evaluation challenges. In the later segments (15–20), the metrics gradually rise again, reflecting the ETSS system's ability to adapt and stabilize scoring as sentence-level dependencies become clearer. Notably, the F1 score (top left) exhibits the most distinct U-shaped pattern, confirming its sensitivity to both local and global translation quality variations. Accuracy (top right) and Precision (bottom left) show similar trends but with a narrower fluctuation range, while Recall rate (bottom right) demonstrates stronger recovery in the later stages compared to its paired metric, machine translation. These results highlight that ETSS dynamically balances multiple evaluation dimensions during translation progress, leveraging its integrated GLR parsing and BP neural network to progressively refine scoring, thereby achieving high reliability in translation quality assessment.

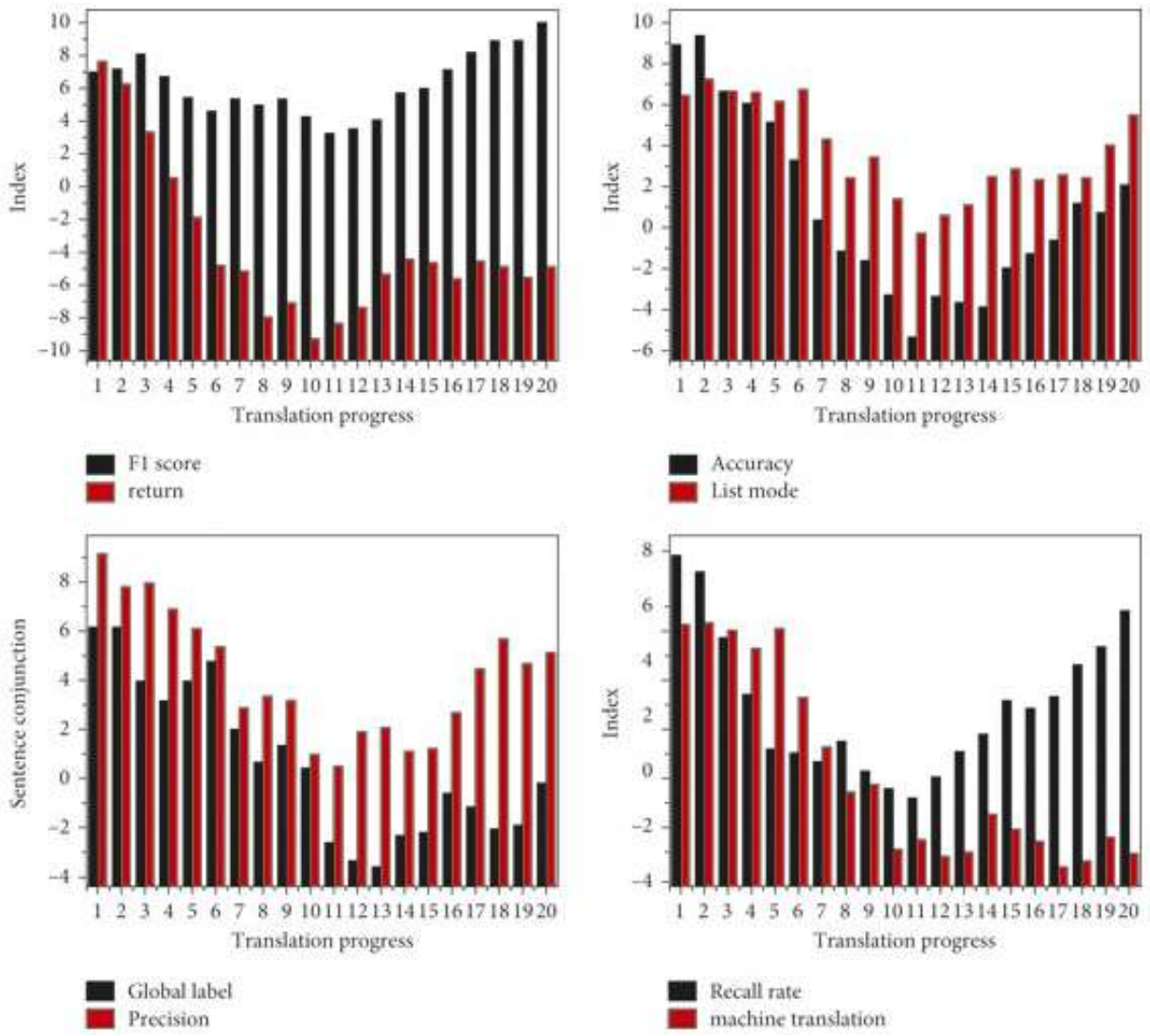


Figure 11: Variation of F1 score, Accuracy, Precision, and Recall rate compared with their respective baseline modules across translation progress (1–20 segments) in ETSS

V. Conclusion

This study introduces a robust framework for enhancing the accuracy and efficiency of English text judgment in translation systems by leveraging advanced computational techniques, including the BP neural network and an improved GLR algorithm. The integration of phrase corpus construction, part-of-speech recognition, and feature vector extraction has demonstrated significant advancements in achieving reliable, automated evaluations of English translation quality. The improved GLR algorithm effectively mitigates challenges related to data point coincidence and structural ambiguity, offering a more precise and adaptive recognition of grammatical features. The application of wavelet-based feature extraction further refines the evaluation process, ensuring high granularity and robustness in feature weighting and text analysis. Experimental results validate the proposed system, with the ETSS achieving an accuracy rate exceeding 95% for certain metrics and an overall

performance score of 92.3%. These findings underscore the potential of intelligent algorithms in reducing human intervention, improving reliability, and minimizing errors in English translation assessments.

References

[1] Manzoni, C., & Ferrari, R. (2021). Spatial transcriptomics: putting genome-wide expression on the map. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology, 45(1), 232-233.

[2] Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H., Adiel, M. A., & Elsadig, M. A. (2024). The impact of artificial intelligence on language translation: a review. Ieee Access, 12, 25553-25579.

[3] Zumor, A. (2021). Exploring intricacies in english passive construction translation in research articles' abstracts by arab author-translators:. SAGE Open, 11(3), 129-148.

[4] Wang, N. (2025). Edge computing based english translation model using fuzzy semantic optimal control technique. PloS one, 20(6), e0320481.

[5] Mu, J., Zhang, Q., & Yang, G. (2021). A study on english translation tactics of the report on the work of the government 2021 from the perspective of eco-translatology. Open Access Library Journal, 8(10), 9.

[6] Stoyanova, D. (2021). Designing esp courses: principles & specificities. Professional Discourse & Communication, 3(1), 62-74.

[7] Thomas, A., Cherian, V., & Antony, A. (2021). Translation, validation and cross-cultural adaptation of the geriatric depression scale (gds-30) for utilization amongst speakers of malayalam; the regional language of the south indian state of kerala. Journal of family medicine and primary care, 10(5), 1863-1867.

[8] Dress, G. A., Regassa, Y., & Koricho, E. G. (2025). Model-Based Mechanical Property and Structural Failure Prediction of Pseudo Ductile Hybrid Composite. Journal of Building Material Sciencel Volume, 7(02).

[9] Fan, Y., & Zang, Y. (2024). Exploration on the Practice of Translation Teaching for English Majors Based on the Internet of Things and Multimedia Assistance. Journal of Electrical Systems, 20(4s), 363-374.

[10] Mohammed, T. A. (2025). Evaluating Translation Quality: A Qualitative and Quantitative Assessment of Machine and LLM-Driven Arabic–English Translations. Information, 16(6), 440.

[11] Latif, S., XianWen, F., & Wang, L. L. (2021). Intelligent decision support system approach for predicting the performance of students based on three-level machine learning technique. Journal of Intelligent Systems, 30(1), 739-749.

[12] Li, C. (2022). A Study on Chinese_English Machine Translation Based on Transfer Learning and Neural Networks. Wireless Communications and Mobile Computing, 2022(1), 8282164.

[13] Tan, M., & Qu, L. (2023). Evaluation of oral English teaching quality based on BP neural network optimized by improved crow search algorithm. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-16.

[14] Chauhan, S., & Daniel, P. (2023). A comprehensive survey on various fully automatic machine translation evaluation metrics. Neural Processing Letters, 55(9), 12663-12717.

[15] Almashhadani, M., & Almashhadani, H. A. (2023). English Translations in Project Management: Enhancing Cross-Cultural Communication and Project Success. International Journal of Business and Management Invention, 12(6), 291-297.

[16] Li, P., Ning, Y., & Fang, H. (2023). Artificial intelligence translation under the influence of multimedia teaching to study English learning mode. International Journal of Electrical Engineering & Education, 60(2_suppl), 325-338.

[17] Li, X., & Huang, C. (2025). Design of an intelligent grading system for college English translation based on big data technology. Systems and Soft Computing, 7, 200205.

[18] Li, Y., Wu, Y., & Zhu, G. (2024). Automatic rating method based on deep transfer learning for machine translation considering contextual semantic awareness. Alexandria Engineering Journal, 105, 588-597.

[19] Liang X, Cheng W, Zhang C, Wang L, Yan X, Chen Q. YOLOD: A task decoupled network based on YOLOv5. IEEE Transactions on Consumer Electronics. 2023 May 25;69(4):775-85.

[20] Gong, M. (2024). The neural network algorithm-based quality assessment method for university English translation. Network: Computation in Neural Systems, 1-13.

[21] Li, J. (2024). English-Chinese translation quality assessment based on phrase statistical machine translation decoding algorithm. International Journal of Maritime Engineering, 1(1), 675-688.

[22] Yuan, Y., & Sharoff, S. (2020). Sentence level human translation quality estimation with attention-based neural networks. arXiv preprint arXiv:2003.06381.

[23] Jiang, X. (2022). Research on the analysis of correlation factors of English translation ability improvement based on deep neural network. Computational Intelligence and Neuroscience, 2022(1), 9345354.

[24] Kim, H., & Lee, J. H. (2016, June). A recurrent neural networks approach for estimating the quality of machine translation output. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 494-498).

[25] Tingting, L., & Mengyu, X. (2020). RETRACTED ARTICLE: Analysis and evaluation on the quality of news text machine translation based on neural network. Multimedia Tools and Applications, 79(23), 17015-17026.