Deep Learning Aided English-Chinese Translation Method And Effect Verification For Academic Texts

Lixia Shen

School of Foreign Languages, Zhengzhou University of Science and Technology, Zhengzhou 450064 China

Corresponding author. E-mail: hsiaoweiw@163.com

Received Jan. 2, 2026; accepted July 10, 2026

Academic text translation differs from general text translation due to its rigorous logical structure, specialized domain terminology, fixed syntactic patterns, and high requirements for semantic fidelity. Traditional neural machine translation (NMT) models based on vanilla Transformer suffer from two core defects in English-Chinese academic translation: insufficient capture of domain-specific term semantic features and weak long-distance logical dependency modeling, which easily cause terminology mistranslation, logical dislocation, and low readability of translated texts. To solve the above problems, this paper proposes a novel deep learning aided English-Chinese academic text translation model (ST-Transformer) based on semantic term enhancement and hierarchical gated attention mechanism. Firstly, a domain academic term embedding module is constructed to fuse term dictionary features and contextual semantic information to realize differentiated encoding of common words and professional terms. Secondly, a hierarchical gated attention unit is designed to optimize the multi-head attention structure of the encoder, suppress invalid redundant information in long academic sentences, and strengthen the correlation between core semantic components. Finally, a length penalty adaptive loss function is introduced to balance the translation quality of long and short academic sentences and avoid incomplete translation of complex clauses. Based on self-constructed multi-domain academic parallel corpus and public IWSLT2019 academic translation dataset, comparative experiments are conducted with mainstream baseline models including vanilla Transformer, BERT-NMT, mBART and LightSeq. The experimental results show that the proposed ST-Transformer model achieves 4.21 BLEU points and 3.86 COMET points higher than the vanilla Transformer on English-Chinese academic translation tasks. Meanwhile, the terminology translation accuracy reaches 96.35%, which effectively improves the overall translation quality and semantic consistency of academic texts. Ablation experiments verify the independent effectiveness of each improved module. This study provides an efficient technical solution for intelligent and high-quality translation of cross-language academic papers, monographs and research reports.

Keywords: Neural Machine Translation; Academic Text; English-Chinese Translation; Transformer; Gated Attention; Term Embedding

http://dx.doi.org/10.6180/jase.202611_34.003

1. Introduction

With the acceleration of global academic cooperation and knowledge circulation, English has become the dominant language for international academic exchanges. A large number of cutting-edge research achievements in various disciplines are initially published in English, while Chinese researchers need to efficiently obtain and digest English academic literature. Conversely, Chinese scholars also need to translate domestic innovative research results into standard

English to publish in international journals and conferences [1–3]. As a special type of formal text, academic texts cover multiple disciplines such as engineering, computer science, materials science and social science, featuring dense professional terminology, complex nested clauses, rigorous logical deduction and single semantic ambiguity [4, 5]. Different from daily conversational texts and general news texts, academic translation puts forward higher requirements for terminology accuracy, syntactic standardization and semantic logical integrity [6, 7].

In recent years, deep learning-based neural machine translation has gradually replaced traditional statistical machine translation (SMT) and become the mainstream technical scheme for cross-language text translation [8–10]. The Transformer model proposed by Vaswani et al. [11] in 2017 realizes global sequence dependency modeling through multi-head self-attention mechanism, breaking through the performance bottleneck of recurrent neural network (RNN) based translation models in long-sequence tasks [12]. On the basis of Transformer, pre-trained language models such as mBERT and mBART further integrate massive unlabeled corpus to enhance cross-language semantic representation ability, which has been widely applied in general text translation scenarios [13].

In view of the particularity of academic text translation, scholars have carried out targeted research in recent years. Tairas et al. [14] constructed a domain-specific corpus to optimize the Transformer model and verified that domain corpus pre-training could effectively improve academic translation quality. Some researchers integrate part-of-speech features and syntactic dependency trees into the encoder to enhance the model's perception of academic sentence structure [15, 16]. In terms of term optimization, a small number of studies add term constraint branches to the decoding end, but most methods only focus on single-discipline terminology and lack universal applicability to multi-domain academic texts [17]. In addition, most existing studies ignore the imbalance problem of long and short sentence translation in academic scenarios, resulting in limited overall optimization effect of the model.

Although existing NMT models have achieved excellent performance in daily text translation, there are still prominent shortcomings in English-Chinese academic text translation.

(1) Poor term translation performance. Vanilla Transformer and pre-trained models adopt unified embedding encoding for all words in the text, without distinguishing professional academic terms and ordinary vocabulary. Domain-specific terms have unique fixed semantic meanings in academic scenarios, and unified encoding is easy to cause semantic deviation and term mistranslation [18].

(2) Weak long-distance logical modeling capability. Academic texts contain a large number of long nested complex sentences. The traditional multi-head attention mechanism calculates the weight of all input words indiscriminately, which will introduce redundant noise information, interfere with the dependency capture of core semantic components, and lead to logical confusion of long sentences after translation [19].

(3) Unbalanced translation quality of sentences with different lengths. The standard cross-entropy loss function has equal weight for all tokens, which makes the model prone to incomplete translation and missing components when facing ultra-long academic clauses, and cannot adapt to the syntactic characteristics of academic texts [20].

Aiming at the above problems, this paper carries out targeted optimization for academic text translation scenarios, and the main research contributions are summarized as follows.

(1) An academic term enhanced embedding module is proposed, which fuses domain term dictionary features and contextual semantic embedding to realize hierarchical differentiated encoding of professional terms and common words, and improve the accuracy of domain terminology translation.

(2) A hierarchical gated multi-head attention mechanism (HG-MHA) is designed to add gated filtering units to the encoder attention layer, dynamically screen invalid redundant information in long sequences, and enhance the long-distance logical dependency modeling ability of the model for complex academic sentences.

(3) An adaptive length penalty loss function is constructed to dynamically adjust the loss weight according to sentence length, balance the translation optimization degree of long and short sentences, and solve the problem of missing translation of long academic clauses.

(4) A multi-domain English-Chinese academic parallel corpus covering 6 disciplines is constructed, and extensive comparative experiments and ablation experiments are carried out to verify the robustness and practical applicability of the proposed model.

2. Materials and methods

2.1. Overall Model Framework

The ST-Transformer model takes the improved Transformer as the basic architecture, including three core parts: term enhanced embedding module, encoder with hierarchical gated attention units, and decoder based on adaptive length penalty loss. The overall workflow is as follows. (1) The input English academic text is embedded by the term enhancement module to obtain fused semantic features containing term attribute information. (2) The encoder uses HG-MHA to filter redundant information and complete long-distance dependency modeling of academic sentences. (3) The decoder receives the encoder output features, generates Chinese translation sequences through multi-layer decoding. (4) The adaptive loss function is used to update model parameters to complete training optimization. The overall structure of the model is shown in Figure 1.

2.2. Academic Term Enhanced Embedding Module

To solve the problem of insufficient term semantic representation in traditional embedding methods, this paper constructs a term enhanced embedding module, which fuses word embedding, position embedding and term attribute embedding. For any input sequence $X = \{x_1, x_2, ., x_n\}$, the final fused embedding vector E_{final} is calculated by Formula (1).

$$E_{final} = \alpha E_{word} + \beta E_{pos} + \gamma E_{term} \tag{1}$$

Where E_{word} represents the basic word embedding obtained by dictionary lookup. E_{pos} represents the sinusoidal position embedding used to perceive sequence order information. E_{term} represents the term attribute embedding, which is generated by the pre-built multi-domain academic term dictionary. (α, β, γ) are learnable fusion weights, satisfying $\alpha + \beta + \gamma = 1$, initialized to 0.5, 0.2 and 0.3 respectively.

For words marked as professional terms in the term dictionary, the model increases the weight of E_{term} dynamically according to term frequency; for ordinary words, the term embedding weight is reduced to avoid redundant interference. This differentiated embedding strategy enables the model to effectively distinguish professional terms and common vocabulary, and enhance the semantic representation ability of domain terminology.

2.3. Hierarchical Gated Multi-Head Attention (HG-MHA)

Aiming at the problem that traditional multi-head attention introduces redundant noise in long academic sentences, this paper designs a hierarchical gated attention unit, which adds a feature filtering gate after the attention scoring stage. The calculation process of HG-MHA is as follows.

First, we map the input feature vector to query, key and value matrices through linear transformation, as shown in Formula (2).

$$Q = XW_Q, K = XW_K, V = XW_V \tag{2}$$

Where W_Q, W_K are trainable parameter matrices. The standard attention score is calculated by scaled dot-product attention as shown in Formula (3).

$$Attn(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

On this basis, the gated filtering unit is introduced, and the gate control coefficient G is calculated by the sigmoid activation function as shown in Formula (4).

$$G = \sigma(W_g \cdot Attn(Q, K, V) + b_g) \tag{4}$$

Where W_g and b_g are the weight and bias of the gate layer. σ represents the sigmoid function. The final gated attention output is shown in Formula (5).

$$HG\_MHA = G \odot Attn(Q, K, V) \tag{5}$$

The gate control coefficient can dynamically screen feature information: for redundant invalid tokens in long academic sentences, the gate coefficient is close to 0 to suppress noise interference. For core semantic tokens such as terms and predicate verbs, the gate coefficient is close to 1 to retain effective features, so as to optimize the long-distance dependency modeling effect of the model.

2.4. Adaptive Length Penalty Loss Function

The standard cross-entropy loss function treats all tokens equally, which leads to poor translation effect of long academic sentences. This paper proposes an adaptive length penalty loss function, which dynamically adjusts the loss weight according to the length of the input sentence. The improved loss function is shown in Formula (6).

$$L_{adaptive} = -\frac{1}{N}\sum_{i=1}^{N}\lambda(l_i)\sum_{t=1}^{T_i} logP(y_{i,t}|x_i;\theta) \tag{6}$$

where N is the total number of samples in a batch. l_i represents the length of the i-th input sentence. $\lambda(l_i)$ is the length penalty coefficient, and its calculation formula is as follows.

$$\lambda(l_i) = 1 + \mu \cdot tanh(\frac{l_i - l_{avg}}{l_{avg}}) \tag{7}$$

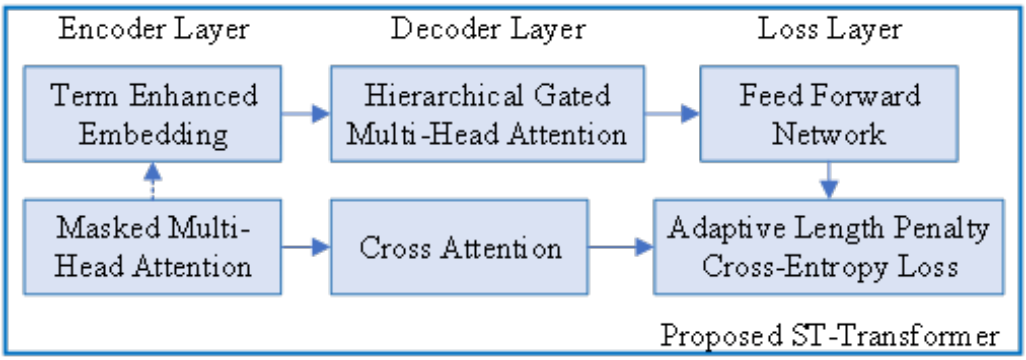

Fig. 1. Overall framework of the proposed ST-Transformer model

In Formula (7), l_{avg} is the average sentence length of the training set. $$\mu$$ is the penalty intensity hyperparameter (set to 0.2 through experimental debugging). For sentences longer than the average length, the penalty coefficient is increased to strengthen the model's optimization constraint on long sentences; for short sentences, the coefficient is appropriately reduced to avoid over-fitting, so as to balance the translation quality of sentences with different lengths.

3. Results and discussion

3.1. Experimental Datasets

The experiments are carried out based on two datasets: public dataset and self-constructed multi-domain academic parallel corpus. The detailed information of the datasets is shown in Table 1.

The self-built academic corpus is collected from high-level SCI/SSCI journals and Chinese core journals, covering 6 mainstream academic disciplines. All parallel sentence pairs are manually screened and proofread to eliminate low-quality noisy data such as machine translation garbage sentences, ensuring the authenticity and effectiveness of experimental data.

3.2. Experimental Settings

To verify the superiority of the proposed model, four mainstream state-of-the-art translation models are selected as baselines.

(1) **Vanilla Transformer.** The original standard Transformer model, the most classic basic framework of NMT tasks.

(2) **BERT-NMT** [21]. Integrate pre-trained BERT encoder into Transformer to enhance contextual semantic representation.

(3) **mBART.** Multilingual pre-trained translation model, which performs well in cross-language academic text translation.

(4) **LightSeq** [22]. Lightweight optimized Transformer model, with high training efficiency and excellent translation performance.

All models are trained on the same hardware and software environment to ensure the fairness of experiments. The hardware environment is NVIDIA RTX 4090 GPU (24G video memory). The deep learning framework is PyTorch 2.1. The key hyperparameters are set as shown in Table 2.

Three authoritative evaluation metrics are adopted to evaluate the model from multiple dimensions.

(1) **BLEU.** The most widely used automatic evaluation metric for machine translation, which calculates the n-gram coincidence degree between the translated text and the reference text.

(2) **COMET.** A pre-training-based evaluation metric, which is more consistent with human subjective evaluation results.

(3) **Term-Acc.** Customized terminology translation accuracy metric, which is used to evaluate the translation effect of academic professional terms.

3.3. Comparative Experimental Results and Analysis

The comparative experimental results of all models on two datasets are shown in Table 3. The best performance data in each column is marked in bold.

It can be seen from Table 3 that the proposed ST-Transformer model achieves the optimal results on all evaluation metrics of the two datasets.

(1) On the IWSLT2019 dataset, the BLEU value is 4.21 higher than vanilla Transformer, and the terminology accuracy is increased by 5.09%.

Table 1. Statistical information of experimental datasets

Dataset	Domain	Training Set	Validation Set	Test Set	Average Sentence Length
IWSLT2019	Computer Science	186,520	5,200	8,500	28.6
Self-built Corpus	Engineering, Materials, Education, Medicine, Economics, Physics	258,360	6,800	10,200	35.2

Table 2. Key hyperparameter settings

Hyperparameter	Parameter Value
Encoder/Decoder Layers	6
Multi-head Attention Number	8
Embedding Dimension	512
Feed Forward Hidden Dimension	2048
Batch Size	32
Learning Rate	5e-4
Dropout Rate	0.15
Epoch	100
Optimizer	AdamW

(2) On the multi-domain self-built corpus with higher difficulty, the model still maintains obvious advantages, which proves that the improved module has strong multi-domain adaptability. The pre-trained models such as mBART are better than the original Transformer, but their optimization effect is limited because they do not carry out targeted design for academic text characteristics.

In order to intuitively compare the performance differences of each model, the BLEU value distribution of different models on the self-built corpus is visualized in the form of histogram, as shown in Figure 2.

First, the vanilla Transformer achieves the lowest BLEU score of 35.18 among all compared models, which serves as the baseline performance benchmark. This result indicates that the standard Transformer architecture, despite its success in general-domain translation tasks, still faces significant challenges when handling academic texts with dense terminology and complex syntactic structures.

Second, the pre-trained models demonstrate moderate improvements over the baseline. Specifically, BERT-NMT achieves a BLEU score of 37.06, representing a 1.88-point improvement compared to the vanilla Transformer. This improvement can be attributed to the enhanced contextual semantic representation capability introduced by the pre-trained BERT encoder. Similarly, mBART and LightSeq achieve BLEU scores of 37.85 and 37.52, respectively, outperforming the vanilla Transformer by 2.67 and 2.34 points. The performance gains of these pre-trained and optimized models confirm that leveraging large-scale pre-training or architectural optimization can partially alleviate the difficulties in academic text translation. However, their improvement margins remain limited, primarily because these models are not specifically designed to address the unique characteristics of academic texts, such as domain-specific terminology and long-distance logical dependencies.

Most notably, the proposed ST-Transformer model achieves the highest BLEU score of 39.39, substantially surpassing all baseline models. Compared with the vanilla Transformer, the ST-Transformer improves the BLEU score by 4.21 points, representing a relative improvement of approximately 12.0%. When compared with the strongest baseline model (mBART), the proposed model still achieves a 1.54-point advantage. This significant performance gap demonstrates the effectiveness of the three core improvements integrated into the ST-Transformer.

(1) The academic term enhanced embedding module, which enables differentiated semantic encoding of professional terminology.

(2) The hierarchical gated multi-head attention mechanism (HG-MHA), which effectively suppresses redundant noise in long academic sentences.

(3) The adaptive length penalty loss function, which balances the optimization degree for sentences of varying lengths.

3.4. Ablation Experiments

To verify the independent effectiveness of the three improved modules in ST-Transformer, ablation experiments are carried out on the self-built corpus. We remove a single improved module in turn and compare the performance changes of the model. The experimental results are shown

Table 3. Comparative experimental results of different models

Model	IWSLT2019 Dataset			Self-built Academic Corpus		
	BLEU	COMET	Term-Acc (%)	BLEU	COMET	Term-Acc (%)
Vanilla Transformer	36.75	78.23	91.26	35.18	77.05	90.12
BERT-NMT	38.62	80.15	93.05	37.06	78.92	92.35
mBART	39.18	80.68	93.72	37.85	79.46	92.88
LightSeq	38.95	80.32	93.41	37.52	79.11	92.56
ST-Transformer	**40.96**	**82.09**	**96.35**	**39.39**	**81.17**	**95.82**

Table 4. Results of ablation experiments

Model Variant	Term Enhanced Embedding	HG-MHA	Adaptive Loss	BLEU	Term-Acc (%)
Baseline(Transformer)	×	×	×	35.18	90.12
Variant 1	✓	×	×	37.25	94.68
Variant 2	×	✓	×	36.89	91.53
Variant 3	×	×	✓	36.42	90.86
ST-Transformer(Full)	✓	✓	✓	**39.39**	**95.82**

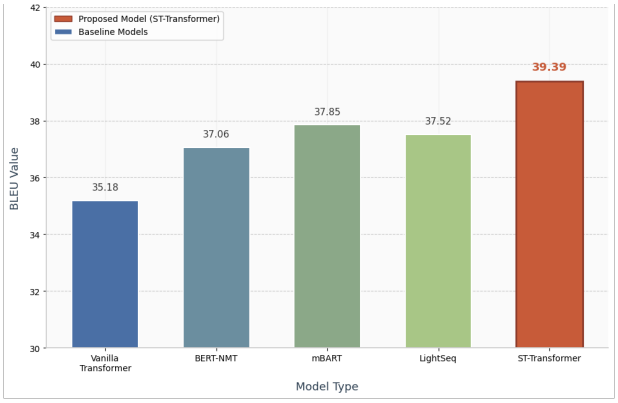

Fig. 2. BLEU value comparison of different models on self-built academic corpus

in Table 4.

The ablation experimental results show that all three modules have positive optimization effects on the model.

(1) The term enhanced embedding module contributes the most to the improvement of terminology accuracy, which directly solves the core pain point of term mistranslation in academic translation.

(2) The HG-MHA module significantly improves the model's ability to process long complex sentences.

(3) The adaptive loss function effectively reduces the missing translation rate of ultra-long sentences.

The simultaneous use of the three modules can produce a synergistic optimization effect, and the overall performance of the model is significantly better than that of single-module improved models.

3.5. Sentence Length Adaptability Analysis

To further verify the adaptability of the proposed model to sentences with different lengths, we divide the test set samples into three categories according to sentence length: short sentences (<20 tokens), medium sentences (20-40 tokens), long sentences (>40 tokens). The BLEU values of vanilla Transformer and ST-Transformer under different sentence lengths are compared, and the results are shown in Figure 3.

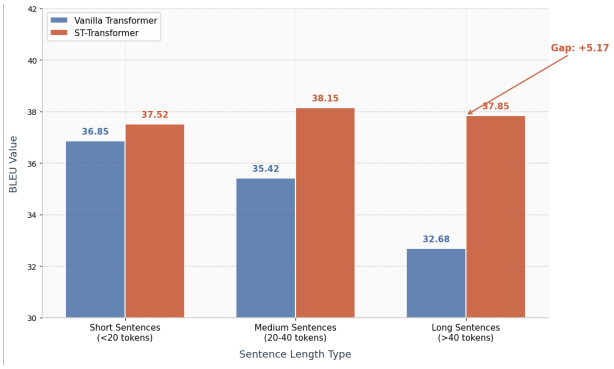

Fig. 3. BLEU performance of two models under different sentence lengths

As shown in Figure 3, the performance gap between ST-Transformer and the baseline model gradually increases with the growth of sentence length. For long academic

sentences with more than 40 tokens, the BLEU value of the proposed model is 5.17 higher than that of the vanilla Transformer, which fully proves that the hierarchical gated attention and adaptive length penalty loss can effectively solve the long sentence translation problem in academic texts.

3.6. Discussion

Based on the above experimental results, this paper further discusses the internal optimization mechanism and application limitations of the ST-Transformer model.

First of all, the term enhanced embedding module breaks the unified encoding mode of traditional NMT models for all vocabulary. By introducing domain term dictionary features, the model can accurately identify professional terms in different disciplines and match fixed academic translation expressions, which is the core reason for the significant improvement of Term-Acc metrics. Different from the simple term constraint method at the decoding end adopted in previous studies, the embedding-level fusion strategy in this paper can realize end-to-end synchronous optimization of terms and contextual semantics.

Secondly, the hierarchical gated attention mechanism filters redundant noise in long sequences from the feature dimension, which makes the model focus more on core semantic components such as subject, predicate and professional terms of academic sentences. Compared with the sparse attention mechanism, the gated filtering method has lower computational complexity and does not increase the training and inference cost of the model, which has better engineering practicability.

In addition, the adaptive length penalty loss function solves the inherent defect of the standard cross-entropy loss function in long-sequence academic translation. The variable penalty coefficient can dynamically adapt to the sentence length distribution characteristics of different disciplines, avoiding the phenomenon that the model overfits short sentences and ignores long complex clauses.

Meanwhile, this research still has certain limitations. The multi-domain term dictionary used by the model is manually constructed, and the expansion efficiency of new discipline terms is low. In the follow-up research, we will introduce automatic term extraction algorithm based on large-scale unlabeled academic corpus to realize real-time dynamic update of the term dictionary. In addition, this paper only focuses on sentence-level translation, and document-level global logical coherence optimization of academic papers will be the next research direction.

4. Conclusions

Aiming at the problems of low terminology translation accuracy, poor long-distance logical modeling ability and unbalanced translation quality of sentences with different lengths existing in current deep learning translation models for English-Chinese academic texts, this paper proposes an improved Transformer translation model ST-Transformer integrating term enhancement, hierarchical gated attention and adaptive length penalty loss. The main conclusions are summarized as follows.

(1) The academic term enhanced embedding module realizes differentiated semantic encoding of professional terms and ordinary vocabulary, effectively improving the translation accuracy of multi-domain academic terms, and the Term-Acc index of the model reaches 96.35% on public datasets.

(2) The hierarchical gated multi-head attention mechanism can suppress invalid redundant information of long academic sentences, strengthen the capture of core semantic dependencies, and significantly improve the translation performance of the model for complex nested clauses.

(3) The adaptive length penalty loss function balances the optimization weight of the model for sentences with different lengths, solves the problem of missing translation of ultra-long academic sentences, and improves the overall robustness of the model.

(4) Comparative experiments on public dataset and self-built multi-domain academic corpus verify that the comprehensive performance of ST-Transformer is significantly better than vanilla Transformer, BERT-NMT, mBART and other mainstream baseline models, which has good universality in different academic disciplines.

In future work, we will optimize the automatic construction technology of domain term dictionaries, expand the coverage of academic disciplines, and further explore document-level end-to-end translation methods to realize high-quality full-text translation of complete academic papers, so as to provide more convenient intelligent translation services for global academic researchers.

References

[1] A. Chidlow, E. Plakoyiannaki, and C. Welch, (2014) *"Translation in cross-language international business research: Beyond equivalence"* **Journal of international business studies *45***(5): 562–582. DOI: 10.1057/jibs.2013.67.

[2] M. Todorova, (2018) *"Civil society in translation: innovations to political discourse in Southeast Europe"* **The Translator *24***(4): 353–366. DOI: 10.1080/13556509.2019.1586071.

[3] B. Heinisch, (2021) *"The role of translation in citizen science to foster social innovation"* **Frontiers in sociology *6***: 629720. DOI: 10.3389/fsoc.2021.629720.

[4] J. Yu, L. Zhao, S. Yin, and M. Ivanović, (2024) *"News recommendation model based on encoder graph neural network and bat optimization in online social multimedia art education"* **Computer Science and Information Systems *21***(3): 989–1012. DOI: 10.2298/CSIS231225025Y.

[5] Y. Jiang and S. Yin, (2023) *"Heterogenous-view occluded expression data recognition based on cycle-consistent adversarial network and K-SVD dictionary learning under intelligent cooperative robot environment"* **Computer Science and Information Systems *20***(4): 1869–1883. DOI: 10.2298/CSIS221228034J.

[6] D. O'Neil, (2025) *"Standardization, power, and purity: Ideological tensions in language and scientific discourse"* **Education Sciences *15***(4): 489. DOI: 10.3390/educsci15040489.

[7] V. Stepanova, (2015) *"Legal drafting and editing in academic studies"* **Procedia-Social and Behavioral Sciences *214***: 1116–1124. DOI: 10.1016/j.sbspro.2015.11.715.

[8] S. C. Siu. "Revolutionising translation with AI: Unravelling neural machine translation and generative pre-trained large language models". In: *New advances in translation technology: Applications and pedagogy*. Springer, 2024, 29–54. DOI: 10.1007/978-981-97-2958-6_3.

[9] S. Bo, Y. Zhang, J. Huang, S. Liu, Z. Chen, and Z. Li. "Attention mechanism and context modeling system for text mining machine translation". In: *2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS)*. IEEE. 2024, 857–863. DOI: 10.1109/DOCS63458.2024.10704434.

[10] T. G. Fantaye and G. Abera. "Developing Bidirectional English-Anuak Machine Translation Using a Deep Learning Approach". In: *Pan African Conference on Artificial Intelligence*. Springer. 2024, 293–308. DOI: 10.1007/978-3-032-05063-2_13.

[11] V. Ashish, S. Noam, P. Niki, U. Jakob, J. Llion, N. Aidan, K. Łukasz, and P. Illia. "Attention is all you need: Advances in neural information processing systems". In: *NeurIPS Proceedings*. 2017. DOI: 10.65215/r5bs2d54.

[12] K. Hu, A. A. Laghari, Y. Liu, J. Li, and H. Li, (2026) *"Application of Adaptive Step Size Runge-Kutta Method in Solving Ordinary Differential Equations"* **IFS/ACM Transactions on Machine Learning *3***(1): 1–8. DOI: 10.70891/TML.2026.040001.

[13] M. I. Ragab, E. H. Mohamed, and W. Medhat. "Multilingual propaganda detection: Exploring transformer-based models mBERT, XLM-RoBERTa, and mT5". In: *Proceedings of the first International Workshop on Nakba Narratives as Language Resources*. 2025, 75–82.

[14] R. Tairas and J. Cabot, (2015) *"Corpus-based analysis of domain-specific languages"* **Software & Systems Modeling *14***(2): 889–904. DOI: 10.1007/s10270-013-0352-6.

[15] M. M. Prajapati and H. A. Patil. "Speech-to-Tree: Cultivating Dependency Structures from Spoken English". In: *2025 International Conference on Asian Language Processing (IALP)*. IEEE. 2025, 79–84. DOI: 10.1109/IALP68296.2024.11156314.

[16] P. Li, Q. Zhu, and W. Zhang. "A dependency tree based approach for sentence-level sentiment classification". In: *2011 12th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. IEEE. 2011, 166–171. DOI: 10.1109/SNPD.2011.20.

[17] D. H. Whetstone, L. E. Ridenour, and H. Moulaison-Sandy, (2022) *"Questionable authorship practices across the disciplines: Building a multidisciplinary thesaurus using evolutionary concept analysis"* **Library & Information Science Research *44***(4): 101201. DOI: 10.1016/j.lisr.2022.101201.

[18] H. Liu, Z. Dong, R. Jiang, J. Deng, J. Deng, Q. Chen, and X. Song. "Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting". In: *Proceedings of the 32nd ACM international conference on information and knowledge management*. 2023, 4125–4129. DOI: 10.1145/3583780.3615160.

[19] S. Yin, L. Wang, A. A. Laghari, L. Teng, G. Srivastava, A. Almadhor, and T. R. Gadekallu, (2026) *"FGM-MLSD: A Fuzzy Region Competition and Gaussian Mixture Segment Model Via Modified Line Segment Detector Model for Airport Object Saliency Detection in Remote Sensing Images"* **IEEE Transactions on Fuzzy Systems** ***34***(4): 1175–1186. DOI: 10.1109/TFUZZ.2026.3650858.

[20] G. Polat, Ü. M. Çağlar, and A. Temizel, (2025) *"Class distance weighted cross entropy loss for classification of disease severity"* **Expert Systems with Applications** ***269***: 126372. DOI: 10.1016/j.eswa.2024.126372.

[21] S. Xiong, Z. Pang, and Y. Cheng. "Low-Resource Machine Translation Model Based on Hybrid Retrieval Enhancement and Dual Feedback Memory Mechanism". In: *2025 5th International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*. IEEE. 2025, 1–4. DOI: 10.1109/CEI66465.2025.11398435.

[22] H. Huang, H. Zhang, Y. Wang, H. Liu, X. Chen, Y. Chen, and Y. Liang, (2025) *"Integrated Robust Optimization for Lightweight Transformer Models in Low-Resource Scenarios"* **Symmetry** ***17***(7): 1162. DOI: 10.3390/sym17071162.