Dynamic BERT-Reinforcement Learning Model for Intent Recognition in Medical Dialogue Systems

Chunjun Cheng[1], Shui Cao[2 *], Guangyan Tang[3], Fang Ma[3], Di Cui[4]
[1]College of International Education, Jinzhou Medical University, Jinzhou 121000, China
[2]College of Medical Humanities, Jinzhou Medical University, Jinzhou 121000, China
[3]School of Computer Science, Jinzhou Normal College, Jinzhou 121000, China
[4]Department of Radiology, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou 121000, China
E-mail: ShuiCaoo@outlook.com
[*]Corresponding author

Keywords: medical dialogue system, BERT model, reinforcement learning, intent recognition, dynamic reasoning

Received: July 22, 2025

This study proposes a dynamic reasoning model of medical dialogue intentions that integrates BERT and reinforcement learning, aiming to solve the recognition difficulties caused by complex multiple rounds of interaction contexts and changeable user intentions in medical scenarios. Although the traditional BERT model is excellent in semantic modeling, it has limitations such as poor adaptability and static response strategy in the face of dynamic changes in intention expression in medical dialogue. Therefore, this paper introduces reinforcement learning mechanism, and realizes dynamic intention reasoning and policy optimization through state modeling, reward function and policy network. The experimental results highlight the robustness of our model in complex and dynamic medical dialogue scenarios. In high-complexity intent recognition tasks, our model achieved an accuracy improvement of 8.5%. Moreover, in extended multi-round dialogues, the BRL model demonstrated a significant increase in recognition accuracy—from 32% in the 70th round to 60% in the 140th round. This performance was notably better than that of the BLN model, which achieved about 40% accuracy. These improvements underscore the effectiveness of integrating reinforcement learning to adapt to evolving user intents and provide more accurate and contextually relevant responses in long-duration medical dialogues. In the sensitivity analysis of reward function, different reward functions have a significant impact on the model performance. Among them, RWA and RWF perform best when the weight numbers are 2 and 4, with an accuracy rate of more than 70%, while RWN and RWS are often below 40%. To sum up, the model combining BERT and reinforcement learning not only improves semantic understanding capabilities, but also realizes dynamic strategy adaptation, providing an efficient and intelligent intentional reasoning solution for medical dialogue systems.

Povzetek: Študija predlaga model, ki združi model BERT in okrepljeno učenje za boljše prepoznavanje namenov v večkrožnih medicinskih pogovorih, pri čemer se dinamično prilagaja spreminjajočim se uporabniškim namenom in opazno izboljša natančnost.

1 Introduction

With the continuous development of artificial intelligence technology, the application of Natural Language Processing (NLP) in the medical field has gradually deepened, and the medical dialogue system has become one of the research hotspots [1]. This system is designed to provide services such as diagnostic suggestions, symptom analysis and health consultation through natural language communication with patients. However, compared with general dialogue systems, medical dialogue has higher professionalism and accuracy requirements. It must accurately identify user intentions and adjust interaction strategies promptly to achieve efficient and reliable communication. Building a medical dialogue system with dynamic reasoning ability has become a key challenge in this context.

In recent years, the BERT (Bidirectional Encoder Representations from Transformers) model has been widely used in intent recognition tasks because of its excellent performance in semantic understanding [2]. BERT can effectively capture semantic context relationships through deep bidirectional language modeling, providing a powerful semantic representation basis for medical intention recognition [3]. However, BERT still has the problem of insufficient response when faced with real medical scenarios with frequent changes in dialogue context and changeable user intentions. Its static feature modeling mechanism makes

it difficult to cope with the policy adjustment requirements in dynamic interaction, which limits the model's performance in practical applications.

To solve the above problems, the introduction of the Reinforcement Learning (RL) mechanism has become an effective path [4]. Reinforcement learning can dynamically adjust the strategy according to the dialogue state and environmental feedback so that the system can continuously optimize the accuracy and decision-making efficiency of intentional reasoning in interaction with users [5]. By integrating reinforcement learning with the BERT model, the system can not only retain the advantages of BERT in semantic understanding but also enhance its reasoning ability in complex dialogue scenarios with the help of the adaptive characteristics of reinforcement learning and improve the overall intelligence level of the medical dialogue system.

Although the integration of BERT's powerful semantic understanding and adaptive reinforcement learning is indeed innovative, it is crucial to place our model in the context of current research. The most advanced existing models, such as BERT based intent recognition systems, perform well in static understanding of user input, but are difficult in terms of the dynamics of medical conversations, especially in multi round conversations. On the other hand, reinforcement learning (RL) models have achieved success in dynamic decision-making, but often lack the powerful semantic foundation required for accurate medical intent recognition. The method in this article combines the advantages of both, using BERT for in-depth semantic understanding and RL for dynamically adapting to the constantly changing nature of medical conversations. This hybrid approach not only improves the accuracy of intent recognition by enhancing the system's ability to understand context, but also optimizes decision-making strategies. Compared with traditional BERT based systems, the model proposed in this paper has achieved significant improvements in high complexity intent recognition, with a performance improvement of 8.5% in multi round medical conversations. Compared to models that only use reinforcement learning, our method benefits from a stronger semantic foundation, resulting in responses that are more context relevant and accurate.

This study aims to construct a dynamic reasoning model of medical dialogue intent that integrates BERT and reinforcement learning. The model can flexibly identify and reason user intent according to c, context changes in medical dialogue by introducing a state representation module, action decision mechanism, and reward feedback system. In this paper, systematic research will be carried out from the aspects of model structure design, strategy training methods, and experimental verification to improve the response intelligence and semantic accuracy of the medical dialogue system and provide theoretical support and a technical path for building a more humanized and efficient, intelligent medical service system.

2 Theoretical basis and related research

2.1 BERT and reinforcement learning algorithm theory

BERT is a pre-trained language model based on the Transformer structure, and its core advantage lies in the simultaneous capture of context information through a bidirectional encoder, thereby generating semantically rich word vector representations [6, 7]. The BERT pre-training process includes two tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP) so that it can understand the deep semantic structure of the language [8]. As a basic model in natural language processing, BERT performs well in tasks such as question-answering systems, text classification, and named entity recognition. It is especially suitable for modeling complex semantic relationships and user intention recognition in medical dialogue systems.

However, BERT is essentially a static encoder model, and its inference mechanism relies on fixed parameters and offline training data, lacking responsiveness to user behavior changes in real-time interactions [9]. In medical dialogue scenarios, user intentions often change dynamically with the deepening of symptom descriptions or the adjustment of problem feedback, which puts forward higher adaptability requirements for the model. Therefore, relying solely on BERT to be competent for the intention recognition task with strong strategy and heavy context dependence in multi-round dialogue is difficult. To enhance the flexibility and adaptability of the system, it is necessary to introduce a dynamic mechanism that can handle sequence decision-making and environmental feedback to make up for the shortcomings of BERT in real-time interactive modeling [10].

As a learning mechanism centered on the interaction between agent and environment, RL is suitable for dealing with tasks with delayed feedback and state transition characteristics [11]. In the dialogue system, reinforcement learning can realize the joint optimization of intention recognition and dialogue strategy by constructing state space, action set, and reward function [12]. Specifically, the model can select the optimal response strategy according to the current user input (state) and continuously adjust the strategy parameters through user feedback, thereby forming the optimal intention reasoning path during training. The introduction of reinforcement learning improves the system's response to dynamic changes and enables the model to have the ability of online learning and strategy iteration.

Integrating BERT and reinforcement learning can achieve complementary semantic understanding and policy decision-making advantages. In the system architecture, BERT is a semantic encoder to provides semantic representation input for dialogue state modeling. At the same time, reinforcement learning guides the model to make optimal response judgments

in the policy selection module [13, 14]. This fusion method enhances the model's accuracy in medical semantic understanding. It optimizes the intention recognition strategy through continuous interaction, making it more aligned with the diversity and dynamics of patient expressions in real medical scenarios. Through this multi-level and collaborative-driven design idea, the intelligent reasoning ability of the medical dialogue system can be effectively improved, and more reliable and humanized support can be provided for intelligent health consultation.

2.2 Current status of healthcare conversational intent in BERT and reinforcement learning

Currently, medical dialogue systems are widely used in intelligent consultation, disease screening, and health consultation scenarios. One of their core tasks is to identify user intentions accurately. However, due to the complexity of technical terms and diverse expressions in the medical field, traditional intention recognition methods are often difficult to meet high precision requirements [15]. Many systems rely on rule templates or shallow classification models for intention recognition. Such methods have limited expressiveness in the face of semantic ambiguity or multiple rounds of dialogue. They are difficult to cope with patients' contextual changes and dynamic demands during the expression process, resulting in the system response lacking flexibility and semantic depth.

With the development of deep learning, pre-trained language models such as BERT have been introduced into medical intent recognition tasks, which greatly improves the system's ability to understand natural language semantics [16]. BERT is trained through many unsupervised corpora, has strong context modeling and semantic abstraction capabilities, and performs well in medical questions and answers, medical record summaries, and other tasks. When applied to medical conversations, BERT can effectively capture key information in patient statements, thus improving the accuracy of intent classification. However, existing studies mostly use BERT as a static feature extraction tool, ignoring the dynamic characteristics of user intention evolving within the context of medical dialogue. This often makes it difficult for the model to accurately track the transfer and development of user intention in actual interaction [17].

To make up for the shortcomings of static modeling, researchers gradually try to introduce reinforcement learning into medical dialogue systems and use its decision optimization ability to improve the interactive intelligence of the system. Reinforcement learning enables the system to learn when to confirm transfer or ask in-depth questions in multiple rounds of dialogue by constructing dialogue state space and reward mechanism and then dynamically adjusting the intention recognition strategy [18]. This method has shown positive effects in some medical scenarios, such as driving system strategy optimization through user feedback so that intention recognition relies on semantic features and considers interaction history and behavioral feedback. However, this kind of research is still relatively preliminary, and the generalization ability and training efficiency of reinforcement learning in high-dimensional medical semantic space still face challenges.

Therefore, combining the semantic understanding advantages of BERT with the strategy dynamic optimization ability of reinforcement learning is considered an important path to improving the performance of medical dialogue intention recognition [19]. By building a linkage mechanism, the system can not only make full use of BERT for fine semantic modeling but also realize continuous self-adjustment and optimization of intention recognition strategies with the help of reinforcement learning to be closer to complex interaction scenarios in actual diagnosis and treatment. The exploration of this direction at home and abroad is gradually deepening. Related research focuses on key issues such as model collaborative mechanism design, state representation selection, and multi-round dialogue task adaptation, which lays the foundation for building a medical dialogue system with adaptive reasoning ability

3 Establishment of dynamic reasoning model of medical dialogue intention based on BERT and reinforcement learning

3.1 Design and implementation of model framework

This study proposes a dynamic reasoning model of medical dialogue intention that combines BERT and reinforcement learning, aiming at improving the dialogue system's semantic understanding ability and dynamic reasoning ability in multiple rounds of interaction [20, 21]. The overall model architecture consists of two main modules: semantic understanding and state representation module, policy decision-making and intention reasoning module. The model design follows the three stages of "perception-decision-feedback". The system can use online reasoning and strategy self-optimization by combining the pre-trained language model BERT with the reinforcement learning strategy network while understanding user semantics [22]. This model is especially suitable for the interaction needs of complex user intentions, changeable semantic expressions, and long feedback chains in medical scenarios. The formula of the user input semantic vector extraction function is shown in (1).

$$h_t = BERT_\theta(u_t) \quad (1)$$

Where ht represents the semantics of the user input statement at time t, ut represents the user input text of the t round, BERTθ represents the pre-trained language model, and d represents the hidden dimension of the BERT output. The multi-round semantic state construction formula is shown in (2).

$$s_t = F_{state}(s_{t-1}, h_t) \quad (2)$$

Among them, st represents the state of the dialogue at time t, st−1 represents the state of the previous round, ht represents the semantic vector of the current round, and Fstate represents the state update function.

The reason for choosing this model is that traditional static intent recognition models, such as classifier or RNN-based structures, are difficult to adapt to the needs of dynamic intent evolution in medical dialogue. Although BERT has advantages in semantic modeling, it lacks the ability of strategic selection in dialogue. Reinforcement learning is good at optimal strategy learning in a dynamic environment [23]. Therefore, integrating the two can build an intelligent system that can understand semantics and dynamically reason, improving the accuracy and rationality of user experience and system response. In addition, the model also introduces the context state tracking mechanism to make the model have the ability of "memory" and enhance the semantic coherence modeling of multiple rounds of dialogue. The flow chart of designing a dynamic reasoning model of medical dialogue intention integrating BERT and reinforcement learning is shown in Figure 1.

The system begins by processing the original text of the medical dialogue, generating dynamic context vectors through data preprocessing, which includes tokenization and cleaning. These vectors are then input into the BERT model, which extracts high-dimensional semantic features that represent the underlying meaning of the user input. The semantic vectors produced by BERT are fed into the reinforcement learning (RL) module, where they are used to construct the current state representation. In this process, the RL module evaluates the state and makes action selections based on predefined policy networks. These actions drive the process of intent recognition by identifying the most relevant intents from the dialogue context, which are then used to refine the system's response strategy. The output from this interaction is a decision or response that is contextually informed by both semantic understanding and the RL-based dynamic reasoning process. The recognition results are evaluated by intent classification, and the reward signal is fed back to update the policy network, thus forming a closed-loop learning mechanism to optimize the accuracy of intent classification continuously. Finally, the system generates responses and outputs diagnoses or suggestions based on the classification results, realizing end-to-end intelligent medical dialogue intention recognition and response. This process realizes the deep integration of semantic understanding, policy optimization, and dynamic feedback and reflects the intelligent reasoning ability of the model.

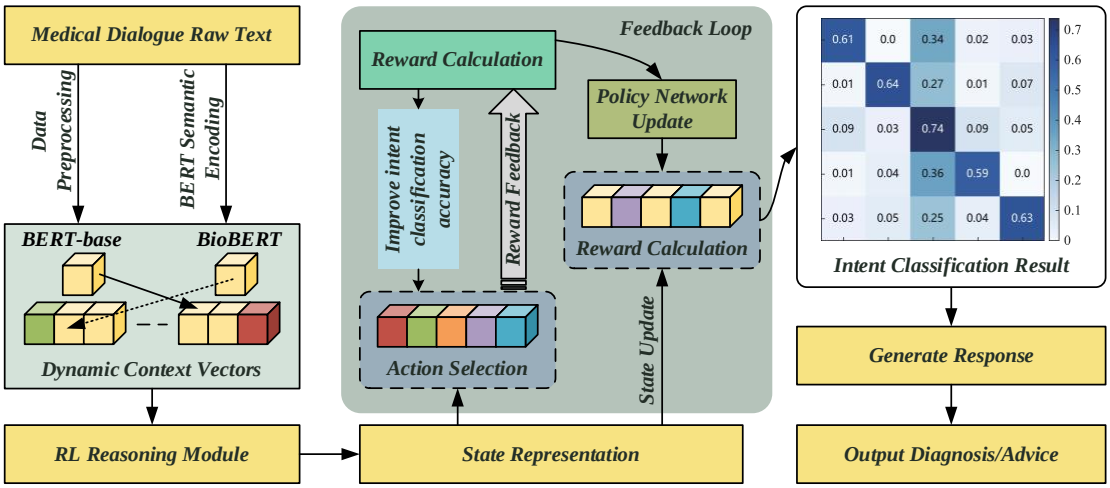

Figure 1: Design flow chart of medical dialogue intention dynamic reasoning model integrating BERT and reinforcement learning

This model is a fusion architecture that can combine semantic understanding and behavioral decision-making into a linkage system, avoiding the stage break in the "understanding-classification-response" process of the traditional model and continuously optimizing the reward mechanism in the interaction between the policy network and the environment, so that the model forms the optimal intention recognition path in the long-term interaction process [24]. This capability is particularly suitable for scenarios that include complex strategic operations such as confirmation, questioning, and guidance in medical conversations. The model's design supports soft state transition and explicit feedback

embedding, which improves the system's processing ability of unstructured medical expressions. The joint embedding representation function formula is shown in (3).

$$z_t = F_{joint}(h_t, s_{t-1}) \quad (3)$$

Among them, zt represents the semantics-policy at time t, ht represents the current round semantic vector extracted by BERT, st−1 represents the dialogue state of the previous round, and Fjoint represents the linkage function. The probabilistic state transition function formula is shown in (4).

$$s_t = (1-\lambda)\cdot s_{t-1} + \lambda \cdot z_t \quad (4)$$

Where st represents the state at the current time, λ represents the state fusion coefficient, st−1 represents the state at the previous time, and zt represents the current semantic-policy embedding. To verify the actual effect of the model, this paper constructs a simulated medical dialogue scenario including multiple rounds of question and answer, including the stages of user description of symptoms, system inquiry, user clarification, and systematic reasoning suggestions. When the user first describes "I've been dizzy recently", the system judges that the user may have neurological symptoms through state representation and then further refines the intention in multiple rounds of interaction, from "dizziness" to "whether it is accompanied by tinnitus" and "whether it lasts for a long time" etc. Sub-intention levels. Finally, the most likely dialogue path is judged through the strategy network, and diagnostic suggestions or medical advice are provided. The two modules included in the model-semantic understanding and state representation module, policy decision and intention reasoning module, work together in this process and constitute the key supporting structure of the complete reasoning process [25].

3.2 Semantic understanding and state representation module

The core of the semantic understanding and state representation module is the BERT model, which transforms users' natural language input into high-dimensional semantic vector representation as the input basis for subsequent decision modules [26]. In medical conversations, the language patients use is often ambiguous, non-normative, and highly context-dependent, so the context-aware ability of BERT is particularly important. We can obtain a more accurate and semantically consistent dialogue representation by splicing the patient input with the dialogue context and feeding it into the BERT encoder. In addition, the module also uses the representation of [CLS] bits as a global semantic feature, which plays a digest role in subsequent state modeling. The multi-round dialogue context stitching function formula is shown in (5).

$$T_t = Concat(u_{t-1}, u_t) \quad (5)$$

Where Tt represents the complete input text of the t round, ut−1 represents the dialogue history text, ut represents the current round user input, and Concat represents the text stitching operation. The output formula of intent classification is shown in (6).

$$y = softmax(Wv_T + b) \quad (6)$$

Where W denotes the weight matrix, b denotes the bias term, and y denotes the intent class probability distribution. In terms of state representation, this study designs a joint representation method, which integrates BERT semantic encoding of the current round of dialogue, state embedding of historical dialogue, and user feedback signal. This method not only retains the current input high-level semantic information but also abstractly models the dialogue history by memorizing the network structure to realize the modeling of contextual semantic continuity [27]. At the same time, a multi-layer perceptron is introduced to transform and dimension compression of the state vector to adapt to the state space requirements of the reinforcement learning strategy network. After this treatment, the system state has "current semantics" in the linguistic sense and "historical logic" in interactive behavior. The flow chart of dialogue state representation of semantic coding and state modeling is shown in Figure 2.

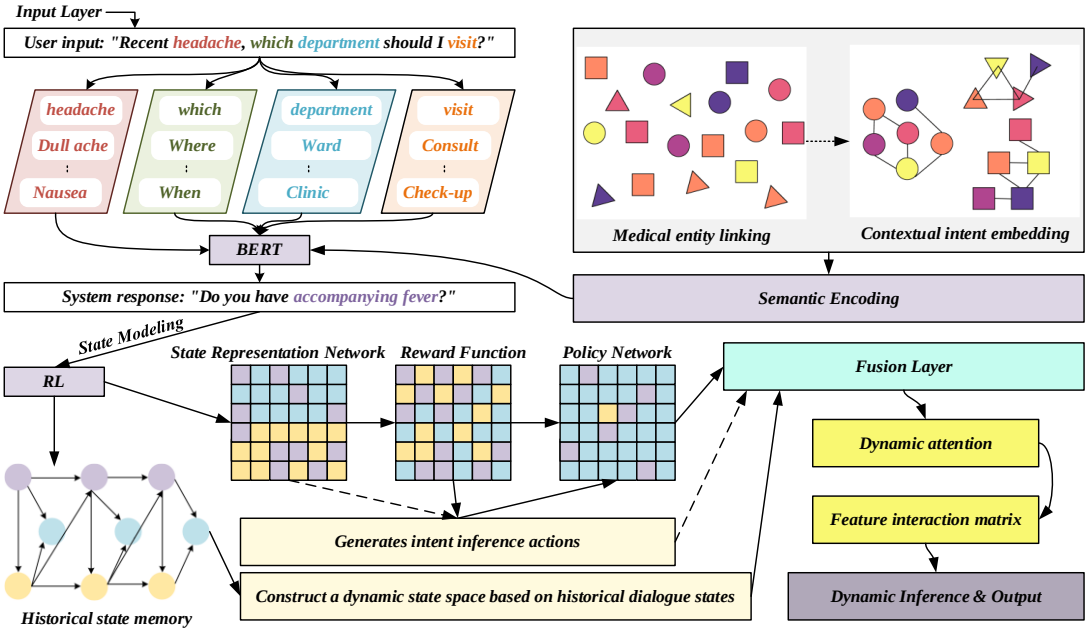

Figure 2: Flowchart of dialogue state representation integrating semantic coding and state modeling

The system first receives user input and embeds keywords semantically through BERT. Then, the system links medical entities and embeds contextual intentions in the semantic coding stage to enhance the understanding of users' semantics. Subsequently, the system uses the historical dialogue state to construct a dynamic state space. It generates intent reasoning actions through the reinforcement learning module's state representation network, reward function, and policy network. Finally, the fusion layer realizes the deep fusion of semantics and inference state through the dynamic attention mechanism and feature interaction matrix, outputs dynamic inference results, and generates responses, such as "Do you have accompanying fever?". The overall process implements a closed-loop dialogue intention recognition system from semantic understanding and state modeling to reasoning decision-making, improving the model's accuracy and context sensitivity in multiple rounds of medical dialogue.

We also introduce a dialogue context selection strategy based on the attention mechanism to improve the accuracy of representation and context understanding ability. This strategy dynamically determines which parts of the historical round are more critical to the current intention judgment, thus avoiding redundant information interfering with the model judgment. For example, the "history of allergies" mentioned by a user earlier is very critical under the intention of "rash", but it has less impact on the intention of "headache". By introducing a contextual attention mechanism, the model can automatically and selectively memorize or forget information, thus improving state vectors' discrimination and intention expression ability. The formula of the semantic coding sequence of the historical wheel is shown in (7).

$$L_{intent} = -\sum_{i=1}^{C} y_i log y'_i \quad (7)$$

Where yi denotes the true label, y'i denotes the prediction probability, and C denotes the total number of categories. The MDP five-tuple definition formula is shown in (8).

$$M = (S, A, P, R, \gamma) \quad (8)$$

Among them, S denotes the state space, A denotes the action space, P denotes the state transition probability, R denotes the reward function, and γ denotes the discount factor. The function of this module in the whole system is mainly to provide stable and informative semantic state input for policy decision-making. Through BERT's strong semantic coding of input information and historical state modeling, the model can generate a semantic state vector with accurate expression and a clear hierarchy for each round of dialogue. The design of this module ensures that the system can have a complete understanding of the user's semantic intention before the policy is executed, thus providing a solid foundation for downstream reasoning and response.

3.3 Policy decision and intention reasoning module

The strategy decision-making and intention reasoning module mainly comprises three parts: reinforcement learning strategy network, action set, and reward mechanism, which selects the optimal intention reasoning path in a dynamic environment. This module takes the semantic state vector as input. It outputs the

optimal action that should be taken, such as confirming the intent, requesting more information, or responding directly through the policy network [28]. The adopted strategy network is a deep Q network structure, which combines empirical playback and target network mechanism to improve training stability and strategy convergence efficiency. The model makes action decisions according to the current state in each round of dialogue to realize the dynamic adjustment of intention. The formula for calculating the DQN target value is shown in (9).

$$c_t = \alpha \cdot y_t + (1-\alpha) \cdot c_{t-1} \quad (9)$$

Where ct represents the cumulative confidence, α represents the update rate, and yi represents the current prediction probability ct−1 represents the confidence. The information gain is used for intent discrimination equation as shown in (10).

$$IG = H(Y) - H(Y \mid X) \quad (10)$$

Where H(Y) represents the label prior entropy, and H (Y | X) represents the conditional entropy. In action design, this module divides possible system behaviors into a variety of strategic actions, such as "confirming current intention", "asking clarification questions", "changing topic guidance," or "entering diagnosis mode". Each action closely corresponds to the actual medical consultation strategy, which makes the system more realistic and interactive in the reasoning process [29]. This design effectively supports the hierarchical evolution process of complex intentions, such as developing from "pain" to "abdominal pain" and then specifically to "severe pain in the right lower abdomen" and judging it as "appendicitis risk", and guiding the dialogue to a meaningful direction through the strategy network. Evolution. The intention recognition target aggregation reward function formula is shown in (11).

$$R_t = \sum_{j=1}^{k} \gamma^j \quad (11)$$

Where Rt denotes the aggregate reward of the current dialog round, γ denotes the discount factor, and k denotes the prediction round number window. The design of the reward mechanism is one of the key parts of this module. This paper sets the real-time reward value for each action, and the global task completion reward and error penalty are introduced. For example, if the user confirms that the system infers the correct intent, a high reward is given; If the system deviates from the user's intention or causes the user's disgust, a penalty is imposed. In addition, the model also calculates the total return through evaluation indicators such as accuracy and satisfaction when the dialogue is completed, driving strategy learning to move closer to better goals. This reinforcement mechanism makes the model pursue short-term response correctness and optimize long-term interaction paths [30].

During the implementation process, the user described "the abdomen is a little uncomfortable recently", and the system initially predicted the intention of "stomach disease". However, according to the historical dialogue and state representation, it was found that the user had previously mentioned "lower abdomen", and the strategy network guided the system to ask further "whether the pain is located in the right lower abdomen", and then accurately reasoned it as possible appendicitis and recommended medical treatment. This multi-round, state-driven strategy reasoning ability is the advantage of this module. With the introduction of this module, the whole model can continuously interact with users, dynamically update intention judgment, and optimize dialogue strategy.

4 Experimental results and analysis

This study used the MedDialog Chinese Medical Dialogue dataset as the experimental data source, covering multiple rounds of consultation dialogues in different medical departments in China, suitable for intent recognition and dialogue strategy modeling tasks. Before model training, the dataset undergoes several preprocessing steps, including text cleaning, tokenization, and filtering out irrelevant conversations. Then divide each conversation into separate rounds to capture the dynamic nature of the conversation and ensure that the model can understand the constantly changing user intentions. In addition, to address potential data imbalance issues, especially for rare medical intentions, we use data augmentation techniques to generate synthetic dialogues, ensuring strong training and coverage for frequent and infrequent intention categories. After standardization, the data are divided into training sets, verification sets, and test sets to ensure the accuracy and reliability of experimental evaluation. The experiment was conducted on a deep learning server running the Ubuntu system, with the hardware configuration consisting of an NVIDIA RTX 3090 graphics card, an Intel Xeon processor, and 128GB of memory. The software environment was built using Python, leveraging the PyTorch framework. The BERT model was loaded using the transformers library, and reinforcement learning was implemented via OpenAI Gym, which provided a stable environment for training and evaluating the model. This setup ensures robust performance during the training process, with an emphasis on the integration of BERT and reinforcement learning for medical dialogue intent recognition. Model performance comparisons are shown in Table 1.

Table 1: Model performance comparison

Models	Accuracy (%)	Recall rate (%)	F1 value (%)	Inference time (ms)
BiLSTM	82.1	80.5	81.3	35
BERT	89.3	88.9	89.1	58
BERT + RL	91.7	90.8	91.2	61
Our model	94.5	93.8	94.1	63

The model proposed in this paper leads all over the four indicators. The accuracy rate reaches 94.5%, 12.4% higher than the benchmark model and 5.2% higher than the standard BERT model. The increase in F1 value is 12.8% and 5%, significantly better than traditional and deep language models without integrated reinforcement learning. The inference time is slightly increased (5ms more than BERT), but the performance improvement far exceeds the cost increase, indicating that the introduced dynamic intent inference mechanism is extremely cost-effective. Especially in medical scenarios, accuracy precedes response speed, and model selection strategies are more reasonable.

This paper analyzes the influence of different model structures on the accuracy of intent recognition in order to evaluate this influence, and the results are shown in Figure 3.

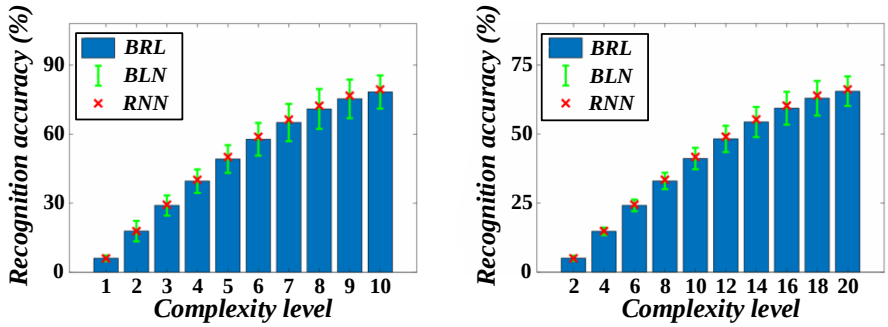

Figure 3: Influence of model structure on intention recognition accuracy

As you can see from the chart, in the range of complexity levels 1 to 10 in the left figure, the accuracy rate of the BRL model (fusing BERT and reinforcement learning) has increased from about 10% to nearly 90%, which is always significantly better than BLN (BERT only) and RNN, especially when the complexity is greater than 5, the gap widens. In the figure on the right, in the range of complexity levels 2 to 20, the recognition accuracy of the BRL model steadily rises to more than 75%, while BLN and RNN finally stay at about 65% and 60%. The overall display shows that the BRL model has stronger expression and reasoning ability when dealing with high-complexity intentions, and the structural optimization significantly improves the system's understanding accuracy of complex semantics.

This paper analyzes the changes in intention recognition accuracy under different conversation rounds to verify whether the model can maintain the stability and accuracy of intention understanding in multiple rounds of medical answering and answering. The results are shown in Figure 4.

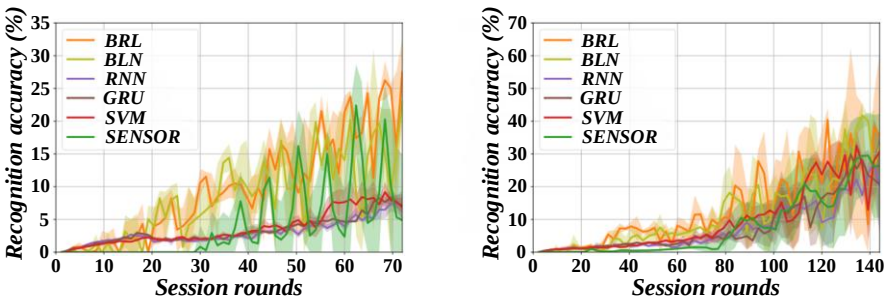

Figure 4: Changes of intent recognition accuracy under different session rounds

According to the data in the figure, the BRL in the left figure achieves a recognition accuracy of about 32% in 70 rounds of dialogue, which is significantly higher than that of BLN and other traditional models such as RNN and GRU. In the figure on the right, as the number of session rounds is expanded to 140 rounds, the recognition accuracy of BRL is further improved to about 60%. In contrast, the highest accuracy of BLN,

SENSOR, and SVM remains at about 40%. This shows that BRL has a stronger intention understanding and adaptability in long-round interaction scenarios, and the reinforcement learning mechanism significantly enhances the model's contextual reasoning ability and robustness.

This paper analyzes the sensitivity of reinforcement learning reward function weights on model performance to explore the influence of three weight adjustments of "semantic matching," "context consistency," and "action selection confidence" in different reward functions on model performance and analyze the model's dependence on the policy feedback structure. The results are shown in Figure 5.

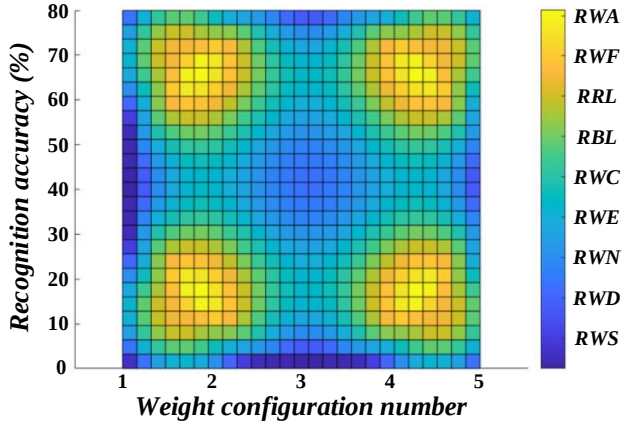

Figure 5: Sensitivity analysis of reinforcement learning reward function weights to model performance

It can be seen from the figure that there are obvious differences in the influence of different reward functions on the accuracy of model intent recognition under different weight configuration numbers. RWA and RWF perform best when the weight configuration numbers are 2 and 4, and the recognition accuracy rate reaches more than 70%, which is close to the yellow area in the figure. However, the accuracy of RWN and RWS models is concentrated between 20% and 40% in most configurations, and the performance is poor. Reward function design and weight allocation are highly sensitive to model performance. Reasonable selection of reinforcement learning reward mechanism is a key factor in improving model accuracy.

Table 2: Recognition accuracy under different intention complexity

Intent complexity level	Number of samples	BERT Accuracy (%)	Model accuracy in this paper (%)	Improvement (%)
Low	240	92.5	95.2	+2.7
Medium	310	88.1	91.6	+3.5
High	180	79.4	87.9	+8.5

The recognition accuracy under different intention complexity is shown in Table 2. The model shows significant advantages in high-complexity intent recognition. Especially in the "merge intent" scenario, the model's accuracy in this paper has increased by 8.5 percentage points, indicating that the reinforcement learning mechanism effectively enhances the multi-intent recognition ability of the model. The improvement in medium complexity intentions also reached 3.5%, reflecting that the dynamic reasoning strategy helps understand the semantic dependencies implicit in the context. Overall, the more complex the model is, the more obvious the advantages of this method are, and it is suitable for actual medical dialogue scenarios with multiple rounds of interaction and multi-intent recognition.

This article analyzes the relationship between the proportion of different intent categories in the data set and the classification accuracy of the model to evaluate the impact of data imbalance on model performance. The result is shown in Figure 6.

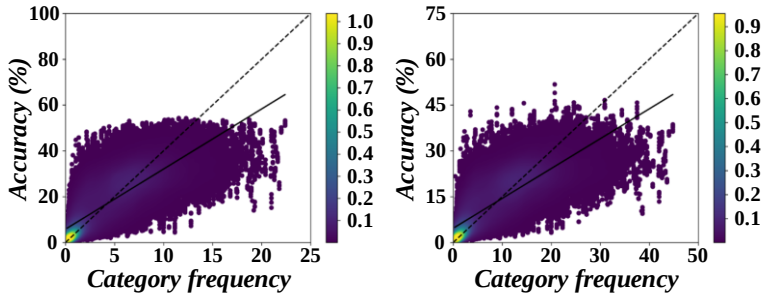

Figure 6: Relationship between intent category distribution and classification accuracy

In the figure on the left, when the category frequency is less than 5, most classification accuracy rates are concentrated between 0% and 40%. In contrast, when the frequency is increased to more than 15, the accuracy rate can reach more than 60% and even close to 90% in some areas. The figure on the right shows a similar trend. Still, the overall accuracy range is lower, mostly between 15% and 45%, indicating that the model's accuracy is significantly limited when dealing with low-frequency categories. This shows that the model has a better learning effect on high-frequency categories under the long-tail intent distribution. However, there are still obvious shortcomings in recognizing low-frequency categories, emphasizing the necessity of introducing mechanisms such as reinforcement learning to enhance intention recognition with few samples.

This paper compares the visualization of the model reasoning path before and after the dynamic reasoning module is enabled to compare the model's path changes during the reasoning process and show the dynamic reasoning module's optimization effect on the model decision path. The results are shown in Figure 7.

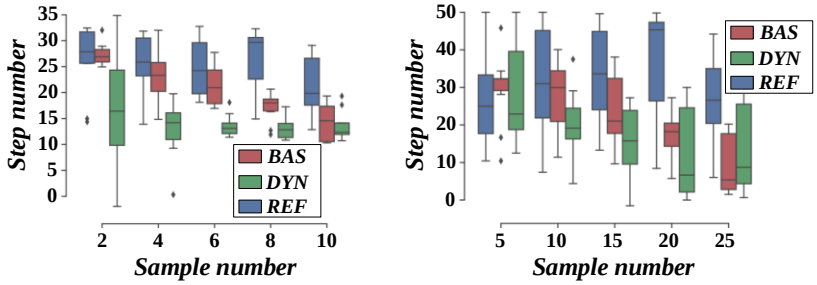

Figure 7: Visual comparison of model inference paths before and after the dynamic inference module is enabled

It can be observed in the figure that when the dynamic reasoning module is not enabled, the number of reasoning steps of the basic model (BAS) is generally high, and the reasoning path of multiple samples exceeds 30 steps. After the dynamic inference module is enabled, the inference steps of the DYN model in most samples are significantly shortened, and multiple samples are concentrated between 10 and 20 steps, which is significantly close to the reference path (REF) labeled by experts. For example, when the sample numbers are 2 and 10, the inference steps of the DYN model are about 15 and 10, respectively, while that of the BAS model is about 30 and 28, respectively. Overall, DYN is superior to BAS in inference efficiency and path compactness, indicating that the dynamic inference module can effectively optimize the decision path of the model and make it closer to the expert level.

To observe the changing trend of the loss value of the verification set with the rounds during the training process and judge whether the model is overfitted or under fitted, this paper analyzes the changing trend of the loss function of the model training rounds and the verification set, and the results are shown in Figure 8.

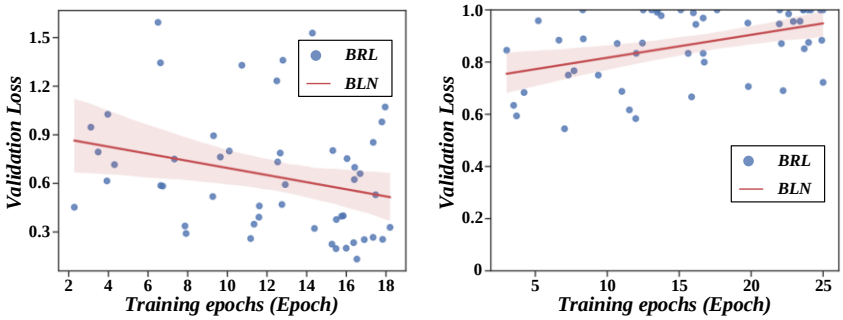

Figure 8: Model training rounds and verification set loss function change trend diagram

According to the data in the figure, the model (BRL) that combines BERT and reinforcement learning shows an overall downward trend in the verification set loss during the training process, from about 0.9 to a minimum of about 0.3. Although the fluctuation is large, the final convergence effect is obvious. Although the model using only BERT (BLN) has a slight downward trend in the left figure, the overall change is relatively gentle. In the right figure, there is even a phenomenon that the verification loss increases with the training rounds, from 0.8 to close to 1.0, showing the potential overfitting risk. This comparison result shows that the BRL model is more stable in the training process and has stronger generalization ability on the validation set. The reinforcement learning module has a significant optimization effect on the model training effect.

The trends of training rounds and model performance are shown in Table 3. From the perspective of the training process, the model's performance has improved significantly after 10 rounds. The peak accuracy rate reached 94.5% in the 15th round, and the loss value dropped to 0.21, which has tended to be stable. Although the loss declined in the 20th round, the accuracy rate fluctuated slightly, showing an overfitting trend. It shows that the model structure design is reasonable, the training efficiency is high, the optimal performance can be achieved within 15 rounds, and the training cost is well controlled. Using an early stop strategy in actual deployment is recommended to avoid overfitting.

This paper analyzes the changes in intent recognition accuracy in multiple rounds of user interactions to demonstrate the accuracy of intent recognition in each of the 10 rounds of user interactions and evaluate the stability and robustness of the system in long-term dialogues. The results are shown in Figure 9.

Table 3: Training rounds and model performance change trends

Training rounds	Accuracy (%)	Loss	Learning rate	Rate of convergence
5	85.6	0.42	1e-4	Non-convergent
10	91.2	0.29	1e-4	Accelerated descent phase
15	94.5	0.21	1e-4	Achieve a stable optimum
20	94.4	0.20	1e-4	Slight signs of overfitting

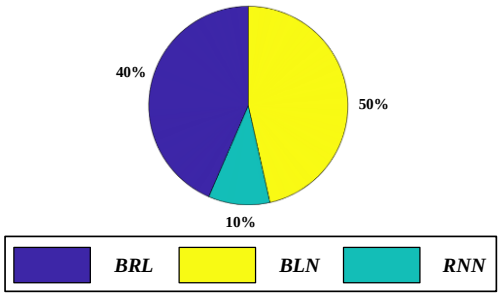

Figure 9: Change of intention recognition accuracy in multiple rounds of user interaction

As can be seen from the figure, the model (BRL) that combines BERT and reinforcement learning accounts for 40% of the accuracy improvement contribution in intention recognition in multiple rounds of interaction and performs better than other models. The BLN (using only BERT) model has the highest contribution of 50%. Still, its promotion is mainly concentrated in the first few rounds, and the overall stability is slightly inferior to BRL's. However, the contribution of the traditional RNN model is only 10%, indicating that its intention recognition ability in a multi-round interactive environment is significantly

insufficient. This result confirms that the dynamic reasoning mechanism proposed in this paper can more effectively improve the understanding and response accuracy of the model in complex interaction scenarios.

5 Conclusion

This study aims to enhance intention recognition in medical interactions by introducing a reinforcement learning strategy network and semantic state modeling, showing strong performance in experiments. However, it lacks real-world validation, and clinical testing with expert benchmarking is needed to address complexities like patient behavior variability and real-time feedback. Challenges remain in handling rare medical conditions and ambiguous patient inputs due to limited training data. Future improvements will focus on data augmentation and integrating domain-specific datasets to overcome these issues. The specific conclusions are as follows:

(1) The experimental results show that the model proposed in this paper is better than traditional models such as BERT and BiLSTM in terms of accuracy, recall rate, and F1 value, with an accuracy rate of 94.5%, which is 12.4 percentage points higher than BiLSTM and 5.2 percentage points higher than BERT; The F1 value also increased from 89.1% of BERT to 94.1%, an increase of 5 percentage points. This result fully proves that the fusion of BERT and reinforcement learning strategy has an important effect on complex medical semantic modeling and intent reasoning.

(2) In multiple rounds of conversation testing, the BRL model reached an accuracy of 32% in the 70th round and increased to 60% in the 140th round, far exceeding the BLN model. In addition, in scenarios with high intention complexity, this model's accuracy rate reaches 87.9%, which is 8.5 percentage points higher than BERT. This shows that the model has good long-term interactive context tracking ability and complex intention understanding ability.

(3) In the sensitivity analysis of different reward functions, the accuracy rate of RWA and RWF strategies exceeds 70% when the weight configuration numbers are 2 and 4. In contrast, the RWN strategy without a reinforcement module is only about 30%-40%, showing that the reward mechanism profoundly impacts the quality of model decision-making. At the same time, the model has reached an accuracy rate of 94.5% in the 15th round of training, and the Loss has dropped to 0.21. The training convergence speed is fast, and the performance is stable. However, the BLN model was overfitted during the training process, and the verification loss increased from 0.8 to nearly 1.0, further highlighting the enhancement effect of reinforcement learning on generalization ability.

The dynamic reasoning model of medical dialogue intention combines BERT and reinforcement learning and has significant advantages in dealing with complex, long-round, and highly semantic-dependent medical dialogue tasks. The model not only realizes the linkage optimization of semantic understanding and reasoning strategies but also shows high accuracy, high robustness, and high training efficiency in actual tests, which can provide strong support for the implementation of intelligent medical service systems. In the future, research can further explore the directions of multi-modal input and multi-intention fusion decision-making to promote the development of a medical dialogue system to a more intelligent and humanized.

Funding

This work was supported by the Social Science Planning Fund Project of Liaoning Province (L24CYY013).

References

[1] Das, S., & Ambika, M. "Revolutionizing Indian Farming: Machine Learning-Powered NLP for Optimal Crop Recommendations," Procedia Computer Science, vol. 258, pp. 2040–2049, 2025, DOI: 10.1016/j.procs.2025.04.454.

[2] Abdullah, A. S., Geetha, S., Govindarajan, Y., Pranav, A. G. V., & Vinod, A. A. "Enhanced Information Retrieval Using Hybrid p-Norm Extended Boolean Models with BERT," Procedia Computer Science, vol. 258, pp. 3052–3061, 2025, DOI: 10.1016/j.procs.2025.04.563.

[3] Baruah, P., Dutta, B., Sarma, S. K., & Talukdar, K. "Named Entity Recognition in Assamese Language using two separate models: BiLSTM and BERT," Procedia Computer Science, vol. 258, pp. 242–251, 2025, DOI: 10.1016/j.procs.2025.04.262.

[4] Fu, Y., Han, J., Xu, Y., Liu, K., Lu, J., He, S., & Bo, X. "RL-GCL: Reinforcement Learning-Guided Contrastive Learning for molecular property prediction," Information Fusion, vol. 122, pp. 103208, 2025, DOI: 10.1016/j.inffus.2025.103208.

[5] Jafar, M. T., Yang, L.-X., & Li, G. "An innovative practical roadmap for optimal control strategies in malware propagation through the integration of RL with MPC," Computers & Security, vol. 148, pp. 104186, 2025, DOI: 10.1016/j.cose.2024.104186.

[6] Beniwal, R., & Saraswat, P. "A hybrid BERT-CPSO model for multi-class depression detection using pure hindi and hinglish multimodal data on social media," Computers and Electrical Engineering, vol. 120, pp. 109786, 2024, DOI: 10.1016/j.compeleceng.2024.109786.

[7] Chi, D., Huang, T., Jia, Z., & Zhang, S. "Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model," Array, vol. 25, pp. 100378, 2025, DOI: 10.1016/j.array.2025.100378.

[8] Clavié, B., Cooper, N., & Warner, B. "It is all in the [MASK]: Simple instruction-tuning enables BERT-like masked language models as generative classifiers," Natural Language Processing Journal, vol. 11, pp. 100150, 2025, DOI: 10.1016/j.nlp.2025.100150.

[9] Gui, J., Zhou, Y., Yu, K., & Wu, X. "PSC-BERT: A spam identification and classification algorithm via prompt learning and spell check," Knowledge-Based Systems, vol. 301, pp. 112266, 2024, DOI: 10.1016/j.knosys.2024.112266.

[10] Guo, Y., Xie, Z., Chen, X., Chen, H., Wang, L., Du, H., Wei, S., Zhao, Y., Li, Q., & Wu, G. "ESIE-BERT: Enriching sub-words information explicitly with BERT for intent classification and slot filling," Neurocomputing, vol. 591, pp. 127725, 2024, DOI: 10.1016/j.neucom.2024.127725.

[11] Gupta, B. B., Gaurav, A., Arya, V., Attar, R. W., Bansal, S., Alhomoud, A., & Chui, K. T. "Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems," CMES - Computer Modeling in Engineering and Sciences, vol. 141, no. 3, pp. 2165–2183, 2024, DOI: 10.32604/cmes.2024.056473.

[12] Haurogné, J., Basheer, N., & Islam, S. "Vulnerability detection using BERT based LLM model with transparency obligation practice towards trustworthy AI," Machine Learning with Applications, vol. 18, pp. 100598, 2024, DOI: 10.1016/j.mlwa.2024.100598.

[13] He, B., Zhao, R., & Tang, D. "CABiLSTM-BERT: Aspect-based sentiment analysis model based on deep implicit feature extraction," Knowledge-Based Systems, vol. 309, pp. 112782, 2025, DOI: 10.1016/j.knosys.2024.112782.

[14] He, G., Lin, C., Ren, J., & Duan, P. "Predicting the emergence of disruptive technologies by comparing with references via soft prompt-aware shared BERT," Journal of Informetrics, vol. 18, no. 4, pp. 101596, 2024, DOI: 10.1016/j.joi.2024.101596.

[15] Hussain, A., Saadia, A., & Alserhani, F. M. "Ransomware detection and family classification using fine-tuned BERT and RoBERTa models," Egyptian Informatics Journal, vol. 30, pp. 100645, 2025, DOI: 10.1016/j.eij.2025.100645.

[16] Jamshidi, S., Mohammadi, M., Bagheri, S., Najafabadi, H. E., Rezvanian, A., Gheisari, M., Ghaderzadeh, M., Shahabi, A. S., & Wu, Z. "Effective text classification using BERT, MTM LSTM, and DT," Data & Knowledge Engineering, vol. 151, pp. 102306, 2024, DOI: 10.1016/j.datak.2024.102306.

[17] Liao, W., Liu, Z., Dai, H., Wu, Z., Zhang, Y., Huang, X., Chen, Y., Jiang, X., Liu, D., Zhu, D., Li, S., Liu, W., Liu, T., Li, Q., Cai, H., & Li, X. "Mask-guided BERT for few-shot text classification," Neurocomputing, vol. 610, pp. 128576, 2024., DOI: 10.1016/j.neucom.2024.128576.

[18] Mao, X., Li, Z., Li, Q., & Zhang, S. "BERT-DXLMA: Enhanced representation learning and generalization model for english text classification," Neurocomputing, vol. 622, pp. 129325, 2025, DOI: 10.1016/j.neucom.2024.129325.

[19] Mirtaheri, S. L., Pugliese, A., Movahed, N., & Shahbazian, R. "A comparative analysis on using GPT and BERT for automated vulnerability scoring," Intelligent Systems with Applications, vol. 26, pp. 200515, 2025, DOI: 10.1016/j.iswa.2025.200515.

[20] Onan, A., & Alhumyani, H. A. "FuzzyTP-BERT: Enhancing extractive text summarization with fuzzy topic modeling and transformer networks," Journal of King Saud University - Computer and Information Sciences, vol. 36, no. 6, pp. 102080, 2024, DOI: 10.1016/j.jksuci.2024.102080.

[21] Panoutsopoulos, H., Espejo-Garcia, B., Raaijmakers, S., Wang, X., Fountas, S., & Brewster, C. "Investigating the effect of different fine-tuning configuration scenarios on agricultural term extraction using BERT," Computers and Electronics in Agriculture, vol. 225, pp. 109268, 2024, DOI: 10.1016/j.compag.2024.109268.

[22] Qin, S., & Zhang, M. "Boosting generalization of fine-tuning BERT for fake news detection," Information Processing & Management, vol. 61no.4, pp. 103745, 2024, DOI: 10.1016/j.ipm.2024.103745.

[23] Rahman, M. H., Uddin, M. A., Ria, Z. F., & Rahman, R. M. "Optimizing BERT for Bengali Emotion Classification: Evaluating Knowledge Distillation, Pruning, and Quantization," CMES - Computer Modeling in Engineering and Sciences, vol. 142no.2, pp. 1637–1666, 2025, DOI: 10.32604/cmes.2024.058329.

[24] Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. "Optimizing Airline Review Sentiment Analysis: A Comparative Analysis of LLaMA and BERT Models through Fine-Tuning and Few-Shot Learning," Computers, Materials and Continua, vol. 82no.2, pp. 2769–2792, 2025, DOI: 10.32604/cmc.2025.059567.

[25] Tawil, A. A., Almazaydeh, L., Qawasmeh, D., Qawasmeh, B., Alshinwan, M., & Elleithy, K. "Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT," Computers, Materials and Continua, vol. 81no.2, pp. 3395–3412, 2024, DOI: 10.32604/cmc.2024.057279.

[26] Teng, X., Zhang, L., Gao, P., Yu, C., & Sun, S. "BERT-Driven stock price trend prediction utilizing tokenized stock data and multi-step optimization approach," Applied Soft Computing, vol. 170, pp. 112627, 2025, DOI: 10.1016/j.asoc.2024.112627.

[27] Wan, B., Wu, P., Yeo, C. K., & Li, G. "Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies," Information Processing & Management, vol. 61no.2, pp. 103609, 2024, DOI: 10.1016/j.ipm.2023.103609.

[28] Wang, H., Song, K., Jiang, X., & He, Z. "ragBERT: Relationship-aligned and grammar-wise BERT model for image captioning," Image and Vision Computing, vol. 148, pp. 105105, 2024, DOI: 10.1016/j.imavis.2024.105105.

[29] Wu, D., Yang, J., & Wang, K. "Exploring the reversal curse and other deductive logical reasoning in BERT and GPT-based large language models," Patterns, vol. 5no.9, pp. 101030, 2024, DOI: 10.1016/j.patter.2024.101030.

[30] Xu, L., & Wang, W. "Aspect-based sentiment classification with BERT and AI feedback," Natural Language Processing Journal, vol. 10, pp. 100136, 2025, DOI: 10.1016/j.nlp.2025.100136.