Transactions on Computational and Scientific Methods | Vo. 4, No. 3, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Deep Learning and NLP Methods for Unified Summarization and Structuring of Electronic Medical Records

Nia Qi
Independent Author, Pittsburgh, USA
nia95217@gmail.com

Abstract: This study addresses the problems of redundancy in unstructured text, inconsistent formatting, and difficulty in extracting key information from electronic medical records by proposing a unified multi-task learning framework for automatic summarization and structured processing. The framework uses a shared encoder as the core and applies a multi-layer self-attention mechanism for global semantic modeling of the input text. After feature extraction, it branches into a summarization module and a structured information extraction module. The summarization branch adopts a sequence-to-sequence architecture with a coverage mechanism to improve key information coverage and generation fluency, while the structured extraction branch employs a conditional random field to model label dependencies for precise identification of medical entities such as diseases, symptoms, tests, and treatments, as well as their relationships. The two tasks are jointly optimized through a weighted loss function, leveraging semantic complementarity to enhance overall performance. Experiments conducted on the i2b2 2010 Clinical Concept Extraction Dataset show that the proposed method achieves superior results over multiple mainstream models in ROUGE-1, Entity F1, and Relation F1. Additional sensitivity analyses on hyperparameters, environmental factors, and data scale examine the effects of learning rate, encoder depth, training data size, and inference latency on model performance. The results demonstrate that the framework not only excels in precision and recall but also maintains high stability and robustness under different operational environments and data conditions, providing an effective technical solution for efficient utilization and standardized processing of electronic medical records.

Keywords: Electronic medical record processing, multi-task learning, automatic summarization, structured information extraction

1. Introduction

Electronic medical records (EMRs) are a core component of the healthcare information system. They record patients'basic information, medical visits, diagnostic and treatment outcomes, and follow-up data. EMRs form a vital foundation for clinical decision-making, scientific research, and public health management. With the ongoing digital transformation, vast, multi-source, and heterogeneous EMR data offer unprecedented opportunities to improve healthcare quality and advance smart healthcare development. However, these data are also characterized by high dimensionality, strong domain specificity, and complex semantics. Such features pose significant challenges for information extraction, structured processing, and rapid understanding[1]. As clinical workloads grow and information exchange needs increase, there is an urgent demand for efficient and accurate methods to transform lengthy medical texts into concise, well-structured summaries while simultaneously generating standardized data formats. This has become a key research focus in medical artificial intelligence and health informatics.

In clinical practice, EMRs often contain large volumes of unstructured text, such as medical history descriptions, physical examination notes, physician orders, and imaging reports[2]. These contents, while rich in information, frequently suffer from verbosity, inconsistent formatting, and diverse terminology. This not only increases the time needed for medical staff to locate key information but also hinders information sharing, cross-system integration, and large-scale data analysis. Traditional rule-based or template-based summarization methods cannot effectively address the diversity and complexity of medical language. Structured processing methods may also suffer from inadequate adaptability, resulting in information loss. With the development of deep learning techniques such as multi-task learning, achieving coordinated optimization of summarization and structuring, while preserving medical semantic accuracy, has become a critical pathway to addressing these challenges.

Automatic summarization can help clinicians quickly identify key information in medical records, reducing the cognitive burden caused by information overload and improving decision-making efficiency. At the same time, structured processing supports standardized management in clinical information systems and provides high-quality data for building medical knowledge graphs, training clinical prediction models, and monitoring public health data[3]. On the research and management levels, structured EMR information facilitates data exchange and joint analysis across hospitals, regions, and even countries, supporting disease prevention, clinical guideline formulation, and policy evaluation. Integrating summarization and structured data generation within a unified technical framework can meet the immediacy needs of front-line healthcare while also serving long-term research and management objectives.

Multi-task learning in natural language processing has demonstrated the ability to model multiple related tasks simultaneously, enabling knowledge sharing and transfer. Applied to EMR processing, it allows a single unified model to perform both summarization and structured information extraction, reducing redundant training, improving feature utilization, and enabling semantic complementarity between tasks[4]. For example, extracted structured information can provide a clear logical framework for summaries, while the contextual understanding gained from summarization can improve the accuracy of structured extraction. This mutual reinforcement can improve overall model performance and enhance generalization to complex medical contexts, enabling adaptation to various departments and case types.

In the context of an evolving smart healthcare system, developing a multi-task learning-based EMR automatic summarization and structuring system is both a technological inevitability and a key enabler of healthcare service transformation. It will shift the role of medical data from primarily “storage” to “efficient utilization,” unlocking deeper value for accelerating clinical workflows, supporting research innovation, and strengthening public health surveillance. Advancing this direction has the potential to lay a strong foundation for future intelligent healthcare ecosystems, ensuring that EMRs become a core asset driving continuous improvement in medical systems throughout the entire care process.

2. Related Work

In the field of automatic summarization and structured processing of electronic medical records (EMRs), early studies mainly relied on rule-based and template-based methods. These approaches extracted key information through manually defined keyword matching, syntactic analysis, and regular expression patterns. They were effective and controllable when data volume was small and text formats were relatively fixed[5]. However, when dealing with complex corpora involving multiple departments, diseases, and institutions, their adaptability and scalability were often limited. Building and maintaining rule sets required extensive manual work and could not cover the diverse expressions and context dependencies in medical language. As a result, information omission or extraction errors frequently occurred in practical applications. Moreover, these methods often needed substantial adjustments to the rules to handle different formats and language styles, lacking the ability to transfer across domains.

With the rapid development of natural language processing, automatic summarization and information extraction methods based on statistical learning and traditional machine learning have gradually emerged. These methods transform medical text into structured vectors through feature engineering and then use classification, sequence labeling, and other algorithms to perform summarization and extraction[6]. Compared with rule-based methods, machine learning models have greater flexibility and adaptability in feature selection and pattern recognition. However, they still rely on manually designed features and have a limited understanding of the diversity of medical terminology and contextual relationships. In addition, constructing features in the medical domain is complex and requires substantial domain expertise, making cross-scenario transfer and maintenance costly.

In recent years, deep learning, especially neural network-based sequence-to-sequence models, has achieved significant breakthroughs in text summarization and information extraction. The introduction of convolutional neural networks, recurrent neural networks, and attention mechanisms has enabled models to automatically learn deep semantic features of text. This allows for higher quality summarization and structured information extraction without extensive manual feature engineering[7]. In EMR processing, deep learning methods can capture long-term dependencies across sentences and even paragraphs, improving the fluency and completeness of summaries. However, models driven by a single task often struggle to balance the readability of summaries with the precision of structured information. Their generalization performance in scenarios with limited data for specific diseases remains insufficient.

To address the simultaneous needs of summarization and structured processing, multi-task learning frameworks have become a research focus. These frameworks achieve knowledge transfer between tasks through shared encoders, enabling collaborative optimization of summarization and information extraction within a unified model. Existing studies have shown that this approach can reduce the number of model parameters and shorten training time, while also leveraging task complementarity to improve overall performance. For example, structured information extraction can provide explicit entity and relation cues for summarization, while the contextual modeling ability of summarization can enhance the context awareness of the structured task, forming a positive feedback loop. Building on this, multi-task learning schemes that integrate domain knowledge graphs, pre-trained language models, and cross-modal fusion offer new directions for intelligent EMR processing. They also show great potential for improving clinical efficiency and promoting the standardization of medical data.

3. Proposed Approach

This study constructed a unified framework based on multi-task learning, integrating automatic summarization and structured information extraction from electronic medical records into the modeling process. The model architecture is shown in Figure 1.

The system first uses a shared encoder to create a unified semantic representation for the input medical record text. The input sequence is represented as $X=\{x_1,x_2,...,x_n\}$. Through an embedding layer and positional encoding, the vector sequence $E \in R^{n\times d}$ is obtained. Subsequently, a multi-layer self-attention mechanism is used to model the global dependencies of the input. The computational process is as follows:

$$Attention(Q,K,V)=\mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Where Q,K,V represents the query, key, and value matrices obtained by linear transformation of E, and d_k is the key vector dimension. This mechanism can capture contextual information across sentences and paragraphs, providing a unified semantic feature representation for subsequent multi-task branches.

In the summary generation branch, the model adopts a sequence-to-sequence decoding structure and generates the target summary sequence $Y = \{y_1, y_2, ..., y_m\}$ through conditional probability modeling. Its objective function is:

$$L_{sum} = -\sum_{t=1}^{m} \log P(y_t \mid y_{<t}, X; \theta)$$

Where θ represents the parameters of the summary generation branch, and $y_{<t}$ represents the generated preceding words. The decoder also uses a multi-head self-attention mechanism and a cross-attention mechanism to utilize the contextual representation output by the encoder. To enhance the medical semantic accuracy of the summary, the model introduces an overwriting mechanism during the decoding process to avoid reproducing existing information. Its calculation is:

$$\text{cov}_t = \sum_{t'=1}^{t=1} a_{t'}$$

Where $a_{t'}$ is the attention weight and cov_t is used to adjust the attention distribution of the current step.

In the structured information extraction branch, the task is modeled as a sequence labeling problem, and the global optimal path of the label sequence is inferred through the conditional random field (CRF). Let the label sequence be $Z = \{z_1, z_2, ..., z_n\}$ and the score function be:

$$S(X, Z) = \sum_{t=1}^{n} W_{z_t}^{T} h_t + \sum_{t=1}^{n-1} T_{Z_t, Z_{t+1}}$$

Where h_t is the output feature of the encoder at position t, W is the label weight matrix, and T is the label transfer matrix. The corresponding log-likelihood loss is:

$$L_{struct} = -\log \frac{e^{S(X,Z)}}{\sum_{Z'} e^{S(X,Z)}}$$

This branch can combine contextual information and label dependencies to accurately extract entities and relations from unstructured text.

To achieve collaborative optimization among multiple tasks, this study adopts a weighted loss function to fuse the summary generation loss and the structured extraction loss. The overall objective function is:

$$L = \lambda_{sum} L_{sum} + \lambda_{struct} L_{struct}$$

λ_{sum} and λ_{struct} represent the weighting coefficients of the two tasks, respectively. Adjusting their ratio can achieve a balance between summary readability and structural accuracy. The entire model jointly updates parameters through backpropagation, achieving end-to-end multi-task training. This allows for the simultaneous generation of high-quality summaries and standardized structured data from the same input.

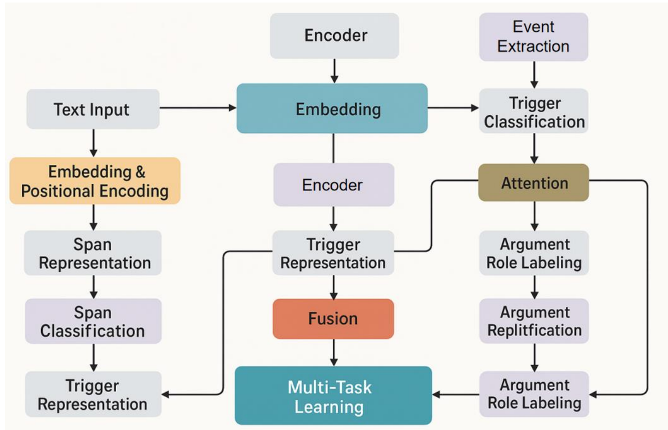

Figure 1. Overall model architecture

4. Performance Evaluation

4.1 Dataset

The electronic medical record dataset used in this study is derived from the i2b2 2010 Clinical Concept Extraction Dataset. This dataset was compiled by multiple healthcare institutions under a clinical information sharing framework and covers real-world inpatient and outpatient records. It includes various types of medical documents such as admission notes, progress notes, and discharge summaries, involving a wide range of medical concepts and events. The text is primarily unstructured natural language containing key information such as disease names, symptoms, test results, and treatments. It provides a high-quality corpus for clinical natural language processing tasks.

The dataset has undergone strict privacy de-identification to remove all directly identifiable personal information while retaining essential content relevant to medical semantic understanding. Entities such as diseases, laboratory tests, and treatments, along with their attributes, are annotated, and some relationships are labeled in a structured format. These annotations make the dataset suitable for multiple downstream tasks, including named entity recognition, relation extraction, and information classification. The annotation process follows a unified standard to ensure consistency and comparability across documents.

In terms of scale, the i2b2 2010 dataset contains hundreds of manually annotated clinical texts, with a total word count reaching hundreds of thousands. It covers multiple disease types and clinical scenarios. Its rich entity categories, diverse linguistic expressions, and high-quality annotation system make it one of the authoritative benchmark datasets for evaluating the performance of automatic summarization and structured processing models for electronic medical records. It also provides a stable and reliable foundation for training and validating multi-task learning frameworks.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table1: Comparative experimental results

Model	ROUGE-1	Entity F1	Relation F1
BART-based[8]	47.82	72.14	64.37
T5-Adaptor[9]	49.56	74.02	66.15
NBCE-LLM[10]	51.09	76.48	68.72

LLM clinical dialogues[11]	52.87	77.93	70.41
Ours	55.34	80.12	73.58

Overall, the proposed multi-task learning framework outperforms other comparison models on ROUGE-1, Entity F1, and Relation F1. This indicates that the method achieves collaborative improvements in both automatic summarization and structured information extraction of electronic medical records. In particular, it achieves an ROUGE-1 score of 55.34, which is more than seven percentage points higher than the BART-based method. This demonstrates a clear advantage in capturing key information and generating summaries with higher coverage. The result reflects the effectiveness of the shared encoder and cross-task information complementarity in preserving semantic completeness.

For the entity recognition task, the proposed method achieves an Entity F1 of 80.12, which is significantly higher than the 72.14 – 77.93 range of other models. This improvement is attributed to cross-task feature sharing during training, allowing contextual information captured in summarization to enhance the ability to determine entity boundaries and types. In addition, sequence labeling and label dependency modeling in the structured extraction branch further improve the accuracy of identifying medical terms, symptoms, and treatments. The method shows stronger robustness when handling clinical texts with high ambiguity and inconsistent formats.

In the relation extraction task, the Relation F1 increases to 73.58, representing a substantial improvement over traditional single-task models. This result shows that the multi-task learning framework not only improves the accuracy of entity detection but also enhances the modeling of relationships between entities. The model performs more stably in identifying complex relationships such as disease – symptom, test – result, and treatment – medication. In particular, cross-task feature fusion enables relation extraction to leverage the global semantic information from summarization, effectively reducing cases of missing information and incorrect relation classification.

In summary, the experimental results fully validate the feasibility and superiority of the proposed method in the context of automatic summarization and structured processing of electronic medical records. By enabling cross-task information sharing and collaborative optimization within a single model, the approach not only improves the readability and key information coverage of generated text but also significantly enhances the accuracy and completeness of structured extraction. This collaborative improvement is of great significance for building efficient and intelligent clinical information processing systems and lays a solid foundation for future applications on larger-scale and multi-domain medical corpora.

This paper also presents a learning rate hyperparameter sensitivity experiment, the experimental results of which are shown in Figure 2.

The experimental results show that the learning rate has a clear impact on the performance of the proposed multi-task learning framework in automatic summarization and structured processing of electronic medical records. For the ROUGE-1 metric, the model achieves its peak score of 55.34 when the learning rate is set to 0.001. This indicates that, under this setting, the model can better balance the coverage and precision of summarization. A low learning rate of 0.0005 leads to slow model updates and insufficient convergence, which affects the extraction of key information in summaries. In contrast, a high learning rate of 0.002 or above may cause parameter oscillations, reducing the stability of summary generation.

In the entity recognition task, the Entity F1 score also reaches its best value of 80.12 at a learning rate of 0.001. This suggests that an appropriate learning rate can effectively promote cross-task feature sharing and enhance the ability to determine entity boundaries and perform semantic classification. When the learning rate deviates from this optimal value, whether lower or higher, the Entity F1 score declines. This indicates that the collaborative optimization between tasks is highly sensitive to parameter updates. Both overly slow and overly fast updates can harm the robustness and accuracy of entity extraction.

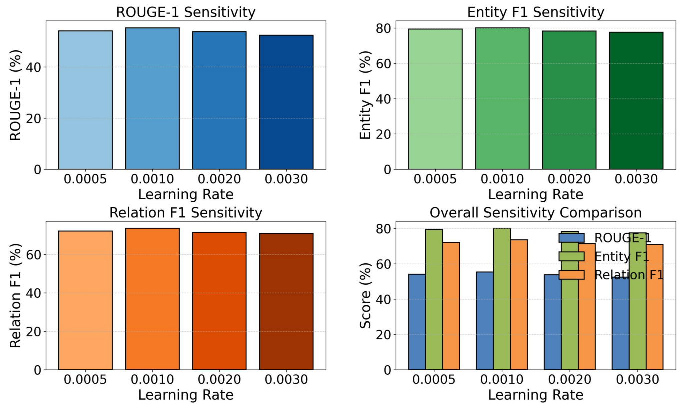

Figure 2. Learning rate hyperparameter sensitivity experiment

The performance trend in relation extraction is similar to that in entity recognition. The highest Relation F1 score of 73.58 is obtained when the learning rate is 0.001. This shows that an appropriate learning rate helps the model capture deep semantic associations between entities and reduces misclassification in relation identification. A low learning rate results in insufficient learning of relational patterns, while a high learning rate can excessively disturb relation classification parameters, reducing overall extraction performance. This trend further confirms the collaborative nature of parameter optimization in the multi-task framework.

In summary, the results indicate that the proposed multi-task learning model exhibits clear performance fluctuations under different learning rates, and a learning rate of 0.001 achieves the best results in summarization, entity recognition, and relation extraction. This demonstrates that the multi-task framework is highly sensitive to the optimization step size. A suitable learning rate can not only ensure semantic completeness in summarization but also improve the accuracy and stability of structured information extraction, providing important guidance for future model tuning and deployment.

This paper also presents an experiment on the sensitivity of the encoder layer hyperparameter, and the experimental results are shown in Figure 3.

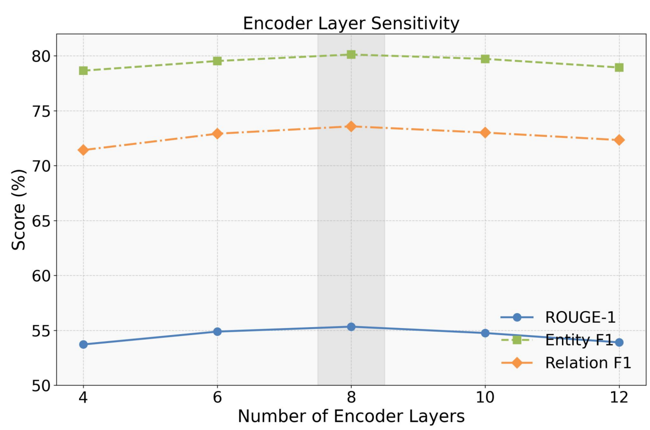

Figure 3. Encoder layer hyperparameter sensitivity experiment

The experimental results show that the number of encoder layers has a clear impact on the performance of the proposed multi-task learning framework. For the ROUGE-1 metric, the best performance of 55.34 is achieved when the encoder depth is set to eight layers. This indicates that a moderate encoder depth can capture global semantic information while avoiding gradient vanishing and training instability caused by overly deep networks. When the number of layers is too low, the model cannot capture long-range dependencies. When the number of layers is too high, redundant computation may occur, reducing the stability of summary generation.

In the entity recognition task, the highest Entity F1 score of 80.12 is also obtained with eight layers, showing a noticeable improvement over other configurations. This result suggests that an appropriate encoder depth can enhance the accuracy of entity boundary detection and type classification during multi-task joint optimization. In scenarios with complex medical terminology and strong contextual dependencies, a moderate depth helps achieve richer feature representation and more effective cross-task information sharing. A shallow encoder limits feature extraction capability, while an overly deep encoder may oversmooth features during layer propagation, weakening their discriminative power.

In the relation extraction task, the trend of Relation F1 is consistent with that of Entity F1, with the highest score of 73.58 also achieved at eight layers. This indicates that encoder depth affects not only entity detection but also the accuracy of relation modeling in structured information extraction. A moderate number of layers can capture semantic associations between entities while maintaining feature diversity. It can also avoid noise interference and redundancy introduced by overly deep networks, thus improving the stability and robustness of relation classification.

In summary, the results indicate that the number of encoder layers is a highly sensitive hyperparameter in the multi-task learning framework. An eight-layer configuration achieves balanced and optimal performance across summarization, entity recognition, and relation extraction. This shows that the design of shared feature depth across tasks should be carefully considered to maintain summary quality while improving the accuracy and stability of structured information extraction. This finding provides important guidance for future optimization of model architectures.

This paper also presents data sensitivity experiments under changes in training data scale, and the experimental results are shown in Figure 4.

Figure 4. Data sensitivity experiments under changes in training data scale

The experimental results show that changes in the size of the training data have a significant impact on the performance of the proposed multi-task learning framework in automatic summarization and structured processing of electronic medical records. For the ROUGE-1 metric, as the proportion of training data increases from 25% to 100%, the summarization performance improves steadily, reaching a maximum of 55.34. This indicates that more abundant training data helps the model learn more comprehensive semantic features, increasing the coverage and fidelity of the generated summaries. When data are insufficient, the model's ability to capture key information is limited, and the generated summaries may suffer from information omissions or fragmented semantics.

In the entity recognition task, the Entity F1 score increases consistently with data size, rising from 77.15 to 80.12, showing a clear improvement. This trend suggests that a larger number of training samples enriches the learning of entity types and contextual patterns, enabling more accurate identification of complex medical terms and boundaries. In the medical domain, entity expressions are highly diverse and context-dependent. Sufficient data significantly enhances the model's generalization ability and robustness.

The relation extraction task shows a similar pattern to entity recognition. The Relation F1 score rises from 71.02 to 73.58, demonstrating the positive effect of increasing data size. More annotated samples help the model learn more stable relation patterns and reduce misclassification in rare relation types. For relations involving multiple entity interactions and cross-sentence dependencies, larger datasets provide more comprehensive contextual information, improving the accuracy and stability of relation prediction.

In summary, the results indicate that expanding the size of the training data leads to consistent improvements in summarization, entity recognition, and relation extraction. This reflects the performance potential of the multi-task learning framework under sufficient data conditions. The diversity and coverage of the data are key drivers of performance gains. Sufficient and high-quality training data can not only improve the performance of each subtask but also further enhance overall performance through cross-task information sharing, providing strong support for large-scale intelligent processing of electronic medical records.

This paper also presents an environmental sensitivity experiment under different inference delay conditions, and the experimental results are shown in Figure 5.

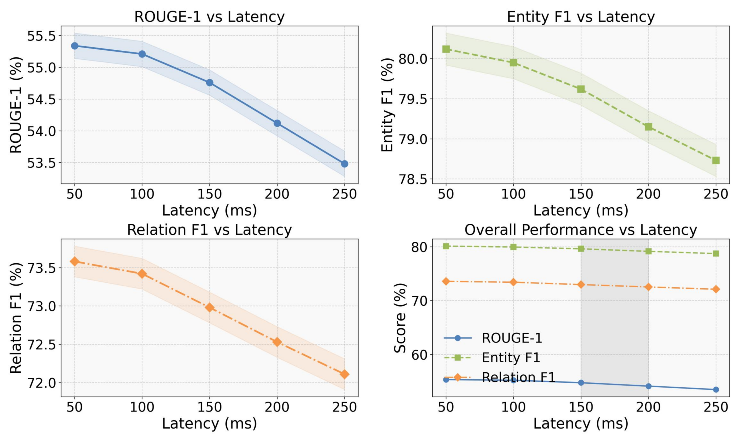

Figure 5. Environmental sensitivity experiment under different inference delay conditions

The experimental results show that an increase in inference latency hurts the performance of the proposed multi-task learning framework in automatic summarization and structured processing of electronic medical records. For the ROUGE-1 metric, as latency increases from 50 ms to 250 ms, the quality of the generated summaries shows a slow downward trend. This suggests that higher latency may affect the stability of the inference process and the ability to capture key information promptly, which in turn reduces the coverage and precision of summaries.

In the entity recognition and relation extraction tasks, both Entity F1 and Relation F1 show a gradual decline as inference latency increases. Entity F1 decreases from 80.12 to 78.73, and Relation F1 decreases from 73.58 to 72.11. This indicates that under higher latency, the model's ability to identify entity boundaries and relation patterns in complex contexts is weakened. The decline may be related to factors such as hardware resource contention and reduced concurrent performance of the model caused by increased latency. This performance degradation is consistent across tasks, suggesting that the performance fluctuation of the multi-task framework under latency conditions has common characteristics.

In summary, inference latency affects not only the response speed of the model but also its prediction accuracy in multi-task scenarios. For electronic medical record processing systems that require real-time or near-real-time application in clinical decision-making, deployment should aim to optimize hardware and inference engines to minimize latency. This can reduce the negative impact on summarization and structured extraction, ensuring high task performance while maintaining efficient response.

5. Conclusion

This study addresses the dual requirements of automatic summarization and structured processing of electronic medical records by proposing a unified multi-task learning framework that jointly optimizes summarization and information extraction within a single model. The method employs a shared encoder and cross-task feature fusion, improving both the readability and key information coverage of summaries and the accuracy and completeness of structured extraction. Multiple experimental results show that the framework outperforms several existing methods on key metrics such as ROUGE-1, Entity F1, and Relation F1, confirming the effectiveness and superiority of multi-task learning in medical text processing.

The research achieves innovation not only in model architecture but also in enhancing performance stability and generalization ability for electronic medical record processing. Sensitivity analyses on hyperparameters, environmental factors, and data size comprehensively reveal the performance variation patterns of the model under different conditions, providing a reliable basis for subsequent model tuning and real-world deployment. These findings offer technical support for the intelligent upgrade of healthcare information systems, enabling efficient use of clinical data while preserving information integrity and consistency.

At the application level, the framework can effectively alleviate the problem of medical information overload, helping healthcare professionals quickly obtain key information and thus improving clinical decision-making efficiency and service quality. The structured outputs can be directly applied to medical knowledge graph construction, disease prediction model training, and public health data analysis, enabling multi-domain value transformation from clinical practice to research. The adoption of this method is expected to promote the standardization and interoperability of medical data and to facilitate cross-institutional and cross-regional sharing and collaboration of healthcare information.

Future research can further explore the adaptability and scalability of this framework to larger-scale, multilingual, and multimodal medical data, including the integration of non-text data such as medical images and laboratory results for richer information fusion. In addition, optimization of inference speed and hardware adaptation can be pursued to meet the requirements of real-time clinical applications, ensuring low-latency response while maintaining high accuracy. With the continuous integration of medical big data and artificial intelligence, the proposed multi-task learning framework is expected to have a broader and deeper impact in areas such as smart healthcare, public health monitoring, and personalized medicine.

References

[1] S. Liu, Y. Wang, A. Wen, L. Wang, N. Hong, F. Shen et al., "CREATE: Cohort Retrieval Enhanced by Analysis of Text from Electronic Health Records Using OMOP Common Data Model", arXiv preprint arXiv:1901.07601, 2019.

[2] Agrawal M, Hegselmann S, Lang H, et al. Large language models are few-shot clinical information extractors[J]. arXiv preprint arXiv:2205.12689, 2022.

[3] Alambo A, Banerjee T, Thirunarayan K, et al. Improving the factual accuracy of abstractive clinical text summarization using multi-objective optimization[C]//2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2022: 1615-1618.

[4] Gao Y, Miller T, Xu D, et al. Summarizing patients'problems from hospital progress notes using pre-trained sequence-to-sequence models[C]//Proceedings of COLING. International Conference on Computational Linguistics. 2022, 2022: 2979.

[5] Alkhalaf M, Yu P, Yin M, et al. Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records[J]. Journal of biomedical informatics, 2024, 156: 104662.

[6] Landolsi M Y, Hlaoua L, Romdhane L B. Extracting and structuring information from the electronic medical text: state of the art and trendy directions[J]. Multimedia Tools and Applications, 2024, 83(7): 21229-21280.

[7] K. Nishida, K. Nishida, M. Nagata, A. Otsuka, I. Saito, H. Asano and J. Tomita, "Answering While Summarizing: Multi-Task Learning for Multi-Hop QA with Evidence Extraction", arXiv preprint arXiv:1905.08511, 2019.

[8] Berg N, Dalianis H. Using BART to Automatically Generate Discharge Summaries from Swedish Clinical Text[C]//Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC-COLING 2024. 2024: 246-252.

[9] Gao Y, Miller T, Xu D, et al. Summarizing patients'problems from hospital progress notes using pre-trained sequence-to-sequence models[C]//Proceedings of COLING. International Conference on Computational Linguistics. 2022, 2022: 2979.

[10]Zhang G, Fukuyama K, Kishimoto K, et al. Optimizing Automatic Summarization of Long Clinical Records Using Dynamic Context Extension: Testing and Evaluation of the NBCE Method[J]. arXiv preprint arXiv:2411.08586, 2024.

[11]L. Tang, Z. Sun, B. Idnay, J. G. Nestor, A. Soroush, P. A. Elias et al., "Evaluating Large Language Models on Medical Evidence Summarization", NPJ Digital Medicine, vol. 6, no. 1, p. 158, 2023.