Transactions on Computational and Scientific Methods | Vol. 4, No. 7, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Time-Aware and Multi-Source Feature Fusion for Transformer-Based Medical Text Analysis

Xiaokai Wang
Santa Clara University, Santa Clara, USA
shawnxkwang@gmail.com

Abstract: This study proposes an improved Transformer-based model for automatic classification and risk prediction using electronic medical records (EMRs). The goal is to address the limitations of traditional methods in semantic understanding when processing unstructured medical texts. The model introduces a hierarchical attention mechanism, a multi-source feature fusion structure, and time-aware embeddings. These enhancements improve the model's ability to capture semantic relationships and track patient condition progression. In the experimental section, the study uses the MIMIC-III dataset to design comparative experiments across three dimensions: embedding strategies, proportion of unstructured text, and prediction window length. These experiments validate the model's advantages in classification accuracy and risk prediction performance. In addition, a joint loss function is constructed to enable multi-task optimization. This further improves the model's adaptability to multi-objective prediction tasks. Experimental results show that the proposed model outperforms mainstream pre-trained language models across all evaluation metrics. It demonstrates strong performance stability and effective text understanding capabilities.

Keywords: Electronic medical records; Transformer; Risk prediction; Multi-task learning

1. Introduction

With the rapid development of healthcare informatization, electronic medical records (EMRs) have become a key data source for clinical diagnosis, scientific research, and health management [1]. EMRs contain both structured and unstructured information, including patient demographics, clinical notes, examination results, diagnoses, and treatment plans. Mining such data helps reconstruct the complete medical history of patients and supports intelligent healthcare services, such as clinical decision-making, disease prediction, and personalized treatment [2]. However, EMR texts often exhibit non-standardized formats, redundancy, and semantic complexity. Traditional information extraction and analysis methods struggle to process such heterogeneous, high-dimensional, and weakly structured data, highlighting the need for more efficient modeling approaches for automatic disease classification and risk prediction.

In recent years, deep learning has made remarkable advances in natural language processing. Transformer-based pre-trained models, such as BERT and RoBERTa, have achieved outstanding results in tasks like text classification, named entity recognition, and sentiment analysis. The Transformer architecture models dependencies between any two positions in a sequence via self-attention, overcoming the efficiency and capability bottlenecks of recurrent neural networks in processing long texts. As a result, Transformer models are gaining traction in medical text mining. While some studies have applied Transformer variants to medical named entity recognition and diagnosis code prediction, most remain at the stage of direct transfer learning or

fine-tuning. These studies often neglect the unique characteristics of medical texts, limiting performance gains in EMR-based classification and risk prediction tasks [3].

From the perspective of disease management, accurate classification of medical records can uncover hidden associations among pathological states and support more informed intervention strategies. This enhances both the efficiency of medical resource allocation and the effectiveness of treatment. Moreover, risk prediction models based on EMRs can issue early warnings before severe symptoms emerge, providing valuable support for chronic disease management and early intervention. This holds significant clinical and societal value, especially amid accelerating population aging and rising chronic disease prevalence. Building robust and interpretable models for automatic classification and prediction has thus become a critical requirement for smart healthcare. A major challenge lies in extracting meaningful patient information from multi-source EMRs and building predictive models that balance generalization with personalized expression.

This study proposes an enhanced Transformer model for EMR-based automatic classification and risk prediction. The model addresses the limitations of standard Transformer architectures by incorporating a hierarchical structure, medical prior knowledge integration, and attention enhancement modules. These improvements enable the model to capture essential semantic and structural information in EMR texts more effectively. Through deep semantic modeling and latent pattern recognition, the model can accurately classify disease types and dynamically predict progression risks. Combining model-driven and data-driven approaches improves interpretability and transferability while offering strong support for clinical decision-making [4].

In summary, this research aims to develop an intelligent model for EMR-based automatic classification and risk prediction. It integrates techniques from natural language processing, deep learning, and medical knowledge engineering to enable deep understanding and scientific modeling of EMR information. The study contributes to advancing smart healthcare from diagnostic assistance to proactive prediction and precise intervention. It holds both theoretical and practical significance by enriching model frameworks in medical AI and supporting the intelligent transformation of healthcare services. Future applications may extend to medical cost control, disease prognosis, and clinical pathway optimization, fostering deeper integration of digital and intelligent healthcare systems.

2. Related work

With the widespread adoption of medical big data, electronic medical records (EMRs) have become a critical data foundation for medical artificial intelligence research. Early EMR modeling studies mainly relied on rule-based text extraction methods, such as regular expressions and template matching, to extract key elements like diagnoses, medications, and examinations. However, these approaches depend heavily on expert knowledge to construct rules, resulting in limited adaptability and scalability. Later, statistical learning methods such as Conditional Random Fields (CRF) and Support Vector Machines (SVM) were introduced to enhance the automation of information extraction and text classification [5-7]. Despite this progress, these methods still rely heavily on feature engineering and struggle to handle complex semantic expressions and cross-sentence dependencies, especially when processing unstructured free-text and ambiguous statements in EMRs.

In recent years, deep learning techniques, particularly models based on the Transformer architecture, have demonstrated outstanding performance in natural language processing and have gradually been applied to medical text analysis. These models, with their self-attention mechanisms and parallel computation capabilities, offer significant advantages over traditional RNN- or CNN-based approaches, especially in handling long and complex sequences such as clinical narratives. Pre-trained language models, such as BERT, learn semantic representations from large-scale corpora through unsupervised learning and can be fine-tuned for downstream tasks with relatively small amounts of labeled data. These characteristics make them highly adaptable for various applications in medical natural language processing. They have achieved remarkable results in tasks such as medical named entity recognition, relation extraction, and automatic diagnosis

prediction, substantially improving the efficiency and accuracy of clinical information extraction. For instance, domain-adapted models like ClinicalBERT and BioBERT were specifically trained on large-scale medical corpora to enhance their understanding of professional terminology and domain-specific expressions, thereby improving their performance on specialized healthcare tasks [8-10]. Nevertheless, despite these advances, existing models still face notable limitations when applied to electronic medical record (EMR) classification and risk prediction. In particular, they often lack the capacity to effectively model long-range textual dependencies and fail to capture fine-grained contextual relationships, both of which are crucial for understanding disease progression and patient trajectories. Moreover, general-purpose models may struggle to adapt to the idiosyncrasies of EMRs, such as fragmented sentence structures, redundant information, and interleaved clinical events, which can severely limit their capacity to uncover the implicit logic embedded in the progression of complex medical conditions.

In the field of automatic disease classification and risk prediction, recent studies have explored a variety of architectures, including multi-task neural networks and sequential prediction models, to better capture the multifaceted nature of patient data. These models often integrate heterogeneous information sources such as temporal sequences of clinical visits, laboratory test results, and prior diagnoses in an attempt to model longitudinal patient development trajectories. For example, models such as RETAIN and Med-BERT incorporate attention mechanisms and sequential structures to better utilize historical patient records and highlight salient features relevant to prediction. Such approaches have shown promise in improving the interpretability and temporal sensitivity of predictions. However, many of these models predominantly focus on structured data formats like numerical indicators or coded diagnoses, and exhibit limited capacity in processing the rich and unstructured clinical narratives that often contain crucial contextual cues. Furthermore, challenges such as semantic redundancy, contextual jumps, and inconsistent data quality in EMRs often degrade the performance of traditional sequence models. To address these limitations, some recent studies have attempted to introduce hierarchical attention mechanisms, local-global semantic fusion strategies, and segment-level modeling to enhance the representation of unstructured text. While these methods have yielded incremental improvements, they often lack a systematic design that tightly integrates domain-specific medical knowledge with flexible deep learning architectures. Additionally, model interpretability and clinical applicability are frequently overlooked in favor of accuracy metrics. Therefore, there is an urgent need to develop an improved Transformer-based model that not only offers a well-structured architectural foundation and strong semantic representation capabilities but also aligns closely with clinical knowledge and real-world healthcare practices. Such a model would be better positioned to handle the complexity of EMR data and provide reliable support for disease classification and dynamic risk prediction tasks in intelligent medical systems.

3. Method

This study proposes an automatic classification and risk prediction model for electronic medical records based on an improved Transformer architecture. By fully leveraging the strengths of Transformer-based structures in capturing long-range dependencies and contextual semantics, the model is specifically tailored to address the challenges inherent in processing unstructured clinical texts. It incorporates a hierarchical attention mechanism, local-global semantic fusion, and medical knowledge-enhanced embedding strategies to enhance the representation capability of the input data. Furthermore, the model supports multi-source feature fusion by integrating both structured and unstructured components of electronic medical records, allowing it to generate comprehensive semantic representations that reflect patient status more accurately. The overall network architecture, as illustrated in Figure 1, demonstrates a modular yet tightly integrated framework where multi-task outputs—classification and risk prediction—are jointly optimized through a customized loss function. This enables the model to simultaneously improve predictive accuracy and maintain consistency across tasks, providing robust support for clinical decision-making and intelligent healthcare applications.

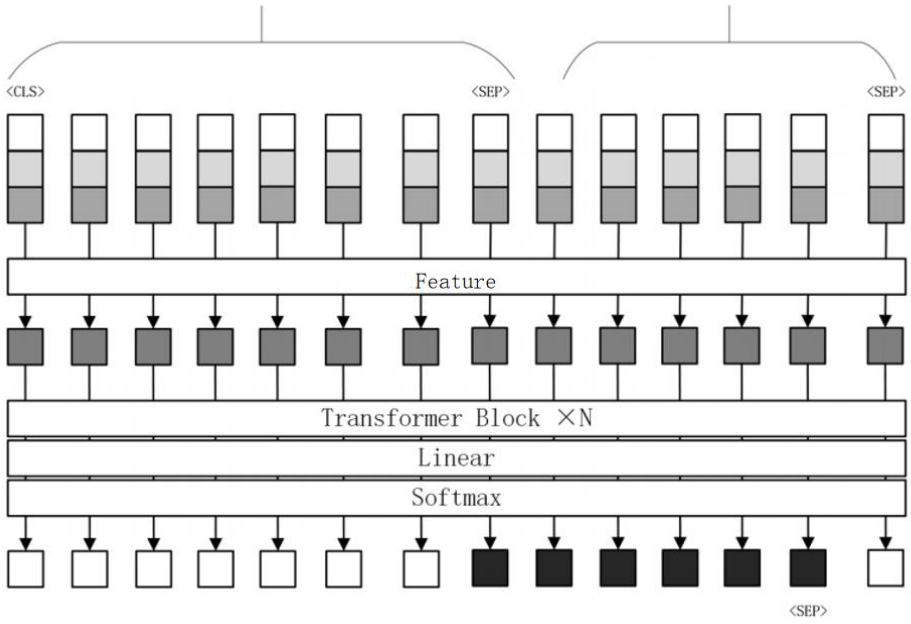


Figure 1. Model network architecture

Figure 1 shows the overall network architecture of the improved Transformer model proposed in this study. The model input is the structured and unstructured text information in the electronic medical record. After unified encoding through the embedding layer, multi-source feature representations are constructed respectively, and the integration of different semantic information is realized in the "Feature" fusion layer. The fused representation is passed into the stacked multi-layer Transformer [11] module to fully explore the long-distance dependencies and contextual semantics in the medical record. The high-dimensional semantic representation is then compressed through the linear transformation layer and finally sent to the Softmax classification layer to realize the automatic classification and risk level determination of the electronic medical record. The distribution of black and white blocks in the figure represents the multi-task output structure, corresponding to the label mapping process of the classification and risk prediction tasks respectively. The overall architecture realizes efficient modeling from raw medical record data to clinical decision-making elements through the layer-by-layer evolution and semantic aggregation of information flow.

Considering that there is a large amount of unstructured text information in electronic medical records, in order to fully extract potential semantic associations and contextual features, the original medical record text is first represented as a word vector sequence $X = \{x_1, x_2, ..., x_n\}$, where each $x_i \in R^d$ is a word vector with an embedding dimension of d. The model uses a pre-trained Transformer as the encoder structure and introduces a hierarchical position encoding mechanism to enhance the perception of clinical paragraph structure. In the self-attention mechanism, the output of each layer is calculated by the following formula:

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

Where Q, K, and V represent query, key, and value vectors, respectively, and d_k is the scaling factor. In order to avoid the problem of semantic dilution when the standard Transformer processes redundant medical information, a local-global fusion module is introduced to regulate the flow and retention of semantics at different layers through a gating mechanism.

When modeling the disease classification task, based on the global representation $H = \{h_1, h_2, ..., h_n\}$ output by the Transformer encoder, the medical record semantic aggregation vector h' is obtained through multi-

head attention weighted pooling for disease category discrimination. Multi-classification cross entropy is introduced as the loss function, which is defined as follows:

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

Where C is the total number of disease categories, y'_i is the predicted probability of the i-th category, and y_i is the one-hot encoding of the true label. In order to improve the model's discrimination boundary for different types of cases, a label smoothing regularization term is further introduced to effectively alleviate the risk of overfitting caused by sample imbalance.

When making risk predictions, the model introduces a historical state modeling mechanism. By modeling the patient's multiple medical record coding sequences for time perception, the improved position embedding function $P_t = PE(t)$ is introduced to map the timestamp t to a time vector of the same dimension as the coding output, and fuse it at the feature layer and the semantic layer. Finally, the multiple visits representation $\{h'_1, h'_2, ..., h'_T\}$ is input into the prediction subnetwork to generate the risk score r'. The weighted mean square error is used as the risk regression objective function:

$$L_{risk} = \frac{1}{T}\sum_{t=1}^{T} w_t (r'_t - r_t)^2$$

w_t is the time decay weight and r_t is the real risk score. The final model jointly optimizes the objective functions of the two subtasks of classification and risk:

$$L = \lambda L_{cls} + (1-\lambda) L_{risk}$$

$\lambda \in [0,1]$ is a hyperparameter used to balance the learning weights of classification and prediction tasks. Through this joint training mechanism, the model can not only achieve accurate automatic disease classification but also has the ability to dynamically predict the progression of the patient's condition, thus improving the practical value of electronic medical records in clinical decision support.

4. Experiment

4.1 Datasets

This study uses the publicly available medical database MIMIC-III (Medical Information Mart for Intensive Care III) as the experimental data source. MIMIC-III is jointly released by the Massachusetts Institute of Technology and the Beth Israel Deaconess Medical Center. It contains detailed electronic medical records of over 40,000 ICU inpatients, covering data from 2001 to 2012. The dataset includes information on hospital admissions, laboratory tests, examinations, medications, and diagnoses. As a comprehensive medical database with multimodal and multi-structured features, MIMIC-III is widely used in clinical predictive modeling and medical natural language processing research.

In this study, we mainly use clinical notes from MIMIC-III—such as admission notes, progress notes, and discharge summaries—as model input to perform automatic patient classification and risk level prediction. All texts are preprocessed, including removal of invalid symbols, terminology normalization, and sentence segmentation. In addition, structured patient information such as age, gender, and length of stay is used as auxiliary input. To ensure data quality and consistency, we include only samples with complete hospitalization records and clear diagnostic labels in our modeling experiments.

This dataset is large in scale, diverse in structure, and clearly labeled. It supports effective evaluation and validation for disease classification and risk modeling tasks. Moreover, MIMIC-III provides highly reliable real-world clinical data, which serves as a strong foundation for testing model performance in complex

medical environments. This helps improve the medical applicability and practical value of the research findings.

4.2 Experimental Results

This paper first gives a performance comparison experiment with the mainstream pre-training model. The experimental results are shown in Table 1.

Table 1: Experimental results

Model	ACC	Precision	Recall	F1-Score
BERT[12]	85.12	83.97	84.55	84.26
RoBERTa[13]	86.03	85.21	85.07	85.14
ClinicalBERT	87.42	86.75	86.33	86.54
BioBERT[14]	88.10	87.48	86.90	87.19
Ours	89.63	89.05	88.72	88.88

As shown in the experimental results in Table 1, mainstream pre-trained models such as BERT, RoBERTa, ClinicalBERT, and BioBERT all demonstrate good performance in EMR-based automatic classification and risk prediction tasks. This indicates the strong modeling capability of Transformer architectures for processing medical texts. Among these models, BioBERT achieves the best overall results, with an F1-score of 87.19%, due to its pre-training on large-scale biomedical corpora. This reflects its enhanced ability to understand domain-specific semantics compared to general-purpose language models.

In comparison, the proposed improved Transformer model outperforms all baseline models across evaluation metrics. Specifically, it achieves an accuracy of 89.63%, a precision of 89.05%, and an F1-score of 88.88%. These results highlight the model's superior ability to capture and represent information in clinical texts. The performance gains can be attributed to the model's architectural enhancements, including multi-layer semantic aggregation, local-global attention fusion, and integration of medical prior knowledge. These components enable more fine-grained modeling of clinical features and contextual dependencies, thereby improving the model's capacity to discriminate complex cases.

Moreover, the comparison of precision and recall shows that our method maintains high accuracy while significantly improving stability and recall. This suggests stronger practical value in real-world clinical applications. Overall, the proposed model achieves a good balance between accuracy and completeness, validating the effectiveness of our structural optimization and feature enhancement strategies.

This paper then gives a comparison of the effects of different embedding strategies in medical record text modeling, and the experimental results are shown in Figure 2.

Furthermore, this paper presents an experiment on the impact of the proportion of unstructured text in electronic medical records on the model effect. The experimental results are shown in Table 2.

As shown in the experimental results in Figure 2, different embedding strategies exhibit noticeable performance differences in EMR text modeling tasks. Traditional static word embedding methods such as Word2Vec, GloVe, and FastText achieve accuracy rates of 81.24%, 82.76%, and 83.15%, respectively. These results indicate a moderate level of performance. Such methods face limitations in processing medical texts, which are rich in domain-specific semantics and context-dependent expressions. In particular, they struggle to capture polysemy and semantic dependencies across sentences.

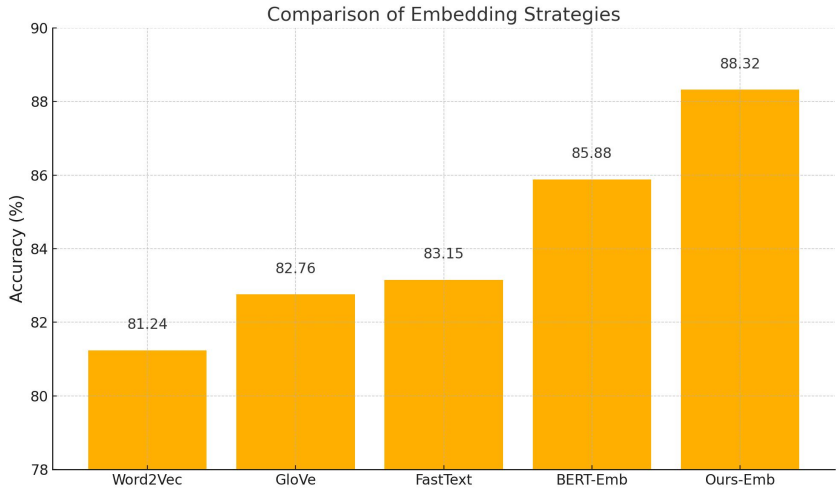


Figure 2. Comparison of Embedding Strategies

In contrast, the model using BERT embeddings achieves an accuracy of 85.88%, demonstrating stronger modeling capacity for medical texts. BERT leverages a bidirectional context encoding mechanism, which enables better understanding of complex semantic structures in clinical narratives. This leads to a clear performance improvement over traditional word embedding. However, BERT still has some limitations when dealing with unstructured formats and domain-specific terms in EMRs, which may affect its final classification accuracy.

Building on this, the embedding method proposed in this study further integrates prior medical knowledge and a hierarchy-aware mechanism. It achieves the highest accuracy of 88.32%. This result suggests that embedding strategies guided by clinical knowledge and structural awareness can significantly enhance the model's ability to represent and discriminate medical texts. It provides a more reliable semantic foundation for downstream tasks such as disease classification and risk prediction.

Furthermore, this paper presents an experiment on the impact of the proportion of unstructured text in electronic medical records on the model effect, and the experimental results are shown in Figure 3.

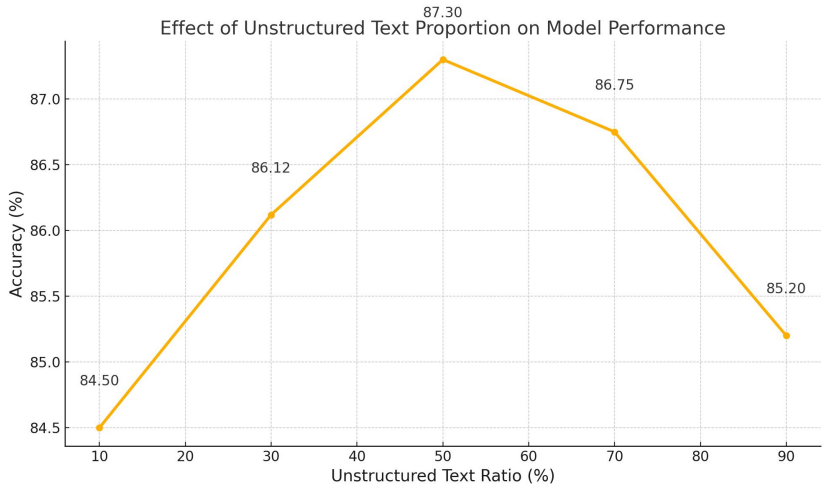


Figure 3. Effect of Unstructured Text Proportion on Model Performance

As shown in the experimental results in Figure 3, the proportion of unstructured text in EMRs has a significant impact on model performance. When the proportion is low, the model achieves relatively poor accuracy. For example, with only 10% unstructured text, the accuracy is 84.50%. This suggests that although

structured data is orderly, it lacks rich semantic content, which limits the model's depth of understanding and predictive capability.

As the proportion of unstructured information increases, model accuracy improves. The highest accuracy, 87.30%, is achieved when unstructured text accounts for 50% of the input. At this stage, the text includes detailed descriptions of symptoms, physician assessments, and treatment plans. This helps the model capture latent semantic patterns and clinical pathways. It also maximizes the contextual modeling strengths of Transformer-based architectures. Therefore, introducing a moderate amount of unstructured content can significantly enhance the model's comprehension of complex medical contexts.

However, when the unstructured text proportion further increases to 70% and 90%, model performance declines. Accuracy drops to 86.75% and 85.20%, respectively. This indicates that excessive redundant and non-standard expressions may introduce noise, disrupting the learning process and reducing prediction accuracy. Overall, the results suggest that a balanced combination of structured and unstructured data is essential for effective EMR modeling. Overreliance on either type may weaken the model's overall performance.

Furthermore, this paper gives an analysis of the impact of different risk prediction window lengths on prediction accuracy, and the experimental results are shown in Figure 4.

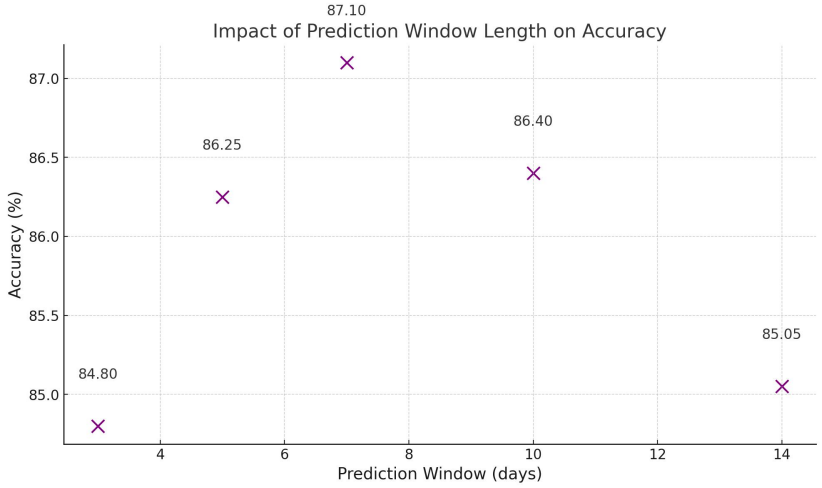


Figure 4. Analysis of the impact of different risk prediction window lengths on prediction accuracy

As shown in the experimental results in Figure 4, the length of the risk prediction window has a significant impact on model accuracy. When the prediction window is short (e.g., 3 days), the model achieves a relatively low accuracy of 84.80%. This may be due to insufficient clinical information within a short time span, which limits the model's ability to identify potential risk trends.

As the prediction window increases, model performance improves. The highest accuracy, 87.10%, is achieved with a 7-day window. This indicates that a medium-length window can capture sufficient clinical changes while avoiding excessive noise. It helps the model form a stable and effective prediction logic. This result confirms the importance of properly setting the prediction window to optimize model performance in risk forecasting.

However, when the window length further increases to 10 and 14 days, the accuracy declines to 86.40% and 85.05%, respectively. This may be due to the inclusion of irrelevant or unstable information over a longer period, which interferes with model learning and judgment. Therefore, in practical applications, the prediction window should be set based on disease characteristics and data cycles to balance accuracy and generalization.

Finally, we give the loss function drop graph, as shown in Figure 5.

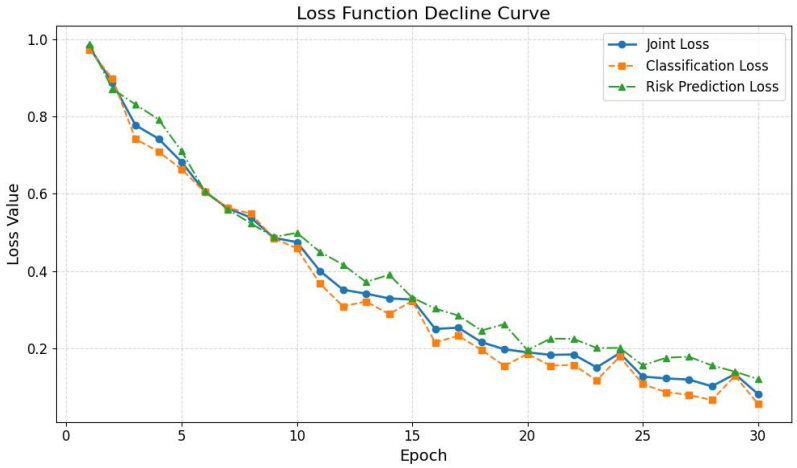


Figure 5. Loss function drop graph

As can be seen from Figure 5, the joint loss, classification loss, and risk prediction loss show a stable downward trend throughout the training process, indicating that the model gradually converges in the multi-task learning process and has good optimization capabilities. In the initial stage, the three loss values are high. As the number of training rounds increases, the model gradually learns effective feature representations and the relationship between tasks, and the loss value decreases significantly.

The classification loss decreases relatively quickly and tends to stabilize after the midterm (about the 10th round), indicating that the model achieves better performance in the disease automatic typing task earlier. The decline of risk prediction loss is a gradual process, reflecting that time-aware modeling has greater complexity in fitting the patient's disease course development trend, and the model requires more rounds to capture long-term dependent features. The joint loss is always between the three and steadily decreases with the improvement of the performance of the two subtasks, verifying the effectiveness of the joint optimization strategy in maintaining the performance balance of each task. Finally, the three curves maintain stable fluctuations in the low loss value range, indicating that the model not only has good convergence but also exhibits stable training effects and strong generalization capabilities in multi-task scenarios.

5. Conclusion

This study proposes an improved Transformer-based model for automatic classification and risk prediction using electronic medical records (EMRs). To address the limitations of traditional methods in modeling capacity and semantic understanding, the model integrates medical prior knowledge, context enhancement, and a multi-task learning mechanism. Experiments on the MIMIC-III dataset show that the proposed method outperforms mainstream pre-trained models in both classification accuracy and risk prediction. The results indicate strong clinical adaptability and generalization ability. In feature modeling, the study emphasizes the role of unstructured text in EMRs. By controlling its proportion and exploring different embedding strategies, the impact on model performance is systematically analyzed. Additionally, the effect of prediction window length on risk prediction is evaluated. This provides theoretical support for window setting in real clinical applications. The results show that a moderate amount of unstructured text and a medium-length prediction window help improve performance. This further validates the effectiveness of the proposed approach.

This research improves the efficiency of EMR understanding and risk modeling at the algorithmic level. It also provides theoretical support for decision-making in intelligent healthcare systems. The model's structure and optimization process are highly scalable. They can adapt to diverse clinical data formats and lay the groundwork for integrating multimodal information in future predictive systems. Moreover, the use of multi-task learning expands EMR modeling toward deeper semantic analysis and comprehensive evaluation. Future work will focus on enhancing model interpretability and clinical deployment capability. New techniques such

as graph neural networks and prompt learning will be explored to improve modeling of latent relationships across EMRs. The study will also be extended to multi-center and multi-language EMR scenarios to assess the model's adaptability and robustness in heterogeneous data environments. These efforts aim to advance the practical application of medical artificial intelligence in broader clinical settings.

References

[1] Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ digital medicine, 4(1), 86.

[2] Li, I., Sun, C., Liu, H., & Wang, F. (2020). BEHRT: Transformer for Electronic Health Records. Scientific Reports, 10(1), 7155.

[3] Huang, K., Altosaar, J., & Ranganath, R. (2019). ClinicalBERT: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342.

[4] Sahu R, Marriott E, Siegel E, et al. Introducing the Large Medical Model: State of the art healthcare cost and risk prediction with transformers trained on patient event sequences[J]. arXiv preprint arXiv:2409.13000, 2024.

[5] Lee, C. H., Schmidt, M., Murtha, A., Bistritz, A., Sander, J., & Greiner, R. (2005, October). Segmenting brain tumors with conditional random fields and support vector machines. In International Workshop on Computer Vision for Biomedical Image Applications (pp. 469-478). Berlin, Heidelberg: Springer Berlin Heidelberg.

[6] Xia, H., Yang, Y., Pan, X., Zhang, Z., & An, W. (2020). Sentiment analysis for online reviews using conditional random fields and support vector machines. Electronic Commerce Research, 20, 343-360.

[7] Moser, G., & Serpico, S. B. (2012). Combining support vector machines and Markov random fields in an integrated framework for contextual image classification. IEEE Transactions on Geoscience and Remote Sensing, 51(5), 2734-2752.

[8] Song, H., Rajan, D., Thiagarajan, J., & Spanias, A. (2018, April). Attend and diagnose: Clinical time series analysis using attention models. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

[9] Li, Y., Wehbe, R. M., Ahmad, F. S., Wang, H., & Luo, Y. (2023). A comparative study of pretrained language models for long clinical text. Journal of the American Medical Informatics Association, 30(2), 340-347.

[10] Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342.

[11] Chen, Huan-Yu, et al. "Lung cancer prediction using electronic claims records: a transformer-based approach." IEEE Journal of Biomedical and Health Informatics 27.12 (2023): 6062-6073.

[12] Chen, Peng, et al. "Named entity recognition of Chinese electronic medical records based on a hybrid neural network and medical MC-BERT." BMC medical informatics and decision making 22.1 (2022): 315.

[13] Cui, Xiaohui, et al. "Fusion of softlexicon and roberta for purpose-driven electronic medical record named entity recognition." Applied Sciences 13.24 (2023): 13296.

[14] Sharaf, Shyni, and V. S. Anoop. "An analysis on large language models in healthcare: a case study of BioBERT." arXiv preprint arXiv:2310.07282 (2023).