Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 4, No. 7, 2025

Joint Modeling of Medical Images and Clinical Text for Early Diabetes Risk Detection

Yun Zi[1], Xiaoxiao Deng[2]
[1]Georgia Institute of Technology, Atlanta, USA
[2]DePaul University, Chicago, USA
*Corresponding Author: Xiaoxiao Deng; dengxiaoxiao1019@gmail.com

Abstract: This study addresses key challenges in early diabetes prediction, including complex data modalities, heterogeneous semantic distributions, and hidden risk signals. A multimodal data fusion method based on electronic health records is proposed. The method takes structured medical image data (EyePACS) and unstructured clinical text records (MIMIC-III/IV) as core inputs. Constructing a unified temporal alignment mechanism and semantic embedding strategy, enables dynamic association modeling across modalities. In terms of model architecture, a collaborative mechanism between CNN and Transformer is introduced. A channel attention module is integrated to enhance the depth of modality interaction and focus on critical features. In addition, time interval embeddings are employed to strengthen the model's perception of event progression rhythms. A series of comparative experiments, ablation tests, and sensitivity evaluations are conducted. The model is systematically assessed across dimensions such as modality dependence, noise robustness, data distribution, and temporal granularity. The results show that the proposed method demonstrates strong stability and discrimination capability when processing multi-source heterogeneous electronic health record data. It effectively captures early risk patterns associated with diabetes. The method provides accurate data support and model support for the automatic identification of high-risk individuals.

Keywords: Multimodal fusion; diabetes prediction; time embedding; electronic medical record modeling

1. Introduction

Diabetes, as a typical chronic metabolic disease, has shown increasing prevalence, younger onset, and prolonged progression in recent years. It poses a serious threat to human health and places a heavy burden on public health systems. In its early stages, diabetes often presents with subtle or hidden symptoms that are easily overlooked by patients, resulting in missed opportunities for timely intervention and treatment. Numerous studies have shown that potential damage to multiple organ systems begins even at the onset of diabetes. Therefore, achieving accurate early identification is critical for slowing disease progression, reducing the risk of complications, and optimizing resource allocation. Against this backdrop, the extensive clinical data accumulated in electronic health records offers new perspectives and technical foundations for early diabetes prediction through efficient information integration and intelligent decision-making mechanisms[1].

Electronic health record systems, as a key component of healthcare informatization, store a wide range of patient data throughout the medical care process. This includes structured diagnostic and treatment information, unstructured text descriptions, and multi-source data from imaging and laboratory examinations. These diverse data types contain rich latent clues about disease progression and individual health conditions. However, due to the high heterogeneity in semantic expression, temporal structure, and information density across different modalities, traditional approaches often fail to effectively extract and associate cross-modal features. This results in poor generalization and low interpretability in predictive models. Therefore, advancing multimodal data fusion techniques in electronic health records is an important step toward intelligent early disease identification[2].

From a clinical perspective, the onset of diabetes is closely associated with multiple factors, including genetics, metabolism, and lifestyle. Its manifestation involves a complex interplay across various dimensions such as physical signs, laboratory indicators, and drug responses. Single-modality data cannot fully capture an individual's risk profile. In contrast, different modalities offer natural complementarity[3]. For example, medical images reveal visual signs of organ abnormalities. Laboratory tests indicate metabolic changes. Clinical notes contain detailed descriptions of disease evolution and medical reasoning. Multimodal collaborative modeling allows for comprehensive integration of structured and unstructured data. This helps improve prediction accuracy while enhancing model robustness and clinical applicability.

In practical applications, multimodal data fusion faces several challenges. These include inconsistent data quality, redundant information, and difficulty in cross-modal alignment. To address these issues, various fusion strategies have been developed, such as early fusion, late fusion, and joint representation learning. Among them, joint representation learning has become a major focus in multimodal medical modeling due to its ability to preserve modality-specific features while ensuring semantic consistency. In the context of electronic health records, combining temporal modeling with contextual construction strategies offers the potential for deep integration across modalities. This provides a scalable solution for risk prediction in complex chronic diseases such as diabetes[4].

This study focuses on the multimodal nature of healthcare information and the complexity of diabetes progression. It aims to build an effective information fusion framework using large-scale electronic health record data to improve sensitivity and accuracy in early-stage diabetes identification. This research direction has both significant methodological value and strong clinical application potential. By exploring the collaborative relationships among structured, physiological, and linguistic clinical features, the study investigates the feasibility of multimodal intelligent modeling in early diabetes prediction. This contributes to the development of precision medicine and smart healthcare systems and provides data-driven support for chronic disease prevention and control strategies.

2. Related work

In recent years, with the continuous advancement of healthcare informatization, electronic health record (EHR) data has played an increasingly important role in disease prediction and clinical decision support. Many studies have focused on mining clinical knowledge hidden in EHRs[5]. Techniques such as natural language processing, statistical modeling, and machine learning have been used to extract features from both structured data, such as vital signs and laboratory results, and unstructured text, such as progress notes and physician orders. These methods have improved the understanding of medical data to some extent. However, due to strong heterogeneity, poor timeliness, and high missing rates among different data modalities, traditional single-modality modeling frameworks often fail to fully capture cross-modal correlations. This limits model generalization and robustness in real-world applications[6].

To address this issue, multimodal learning approaches have been gradually introduced into healthcare and have shown promising progress. Multimodal learning emphasizes collaborative modeling of medical data from different sources. By designing shared representation spaces or interaction mechanisms, it integrates information from structured data, imaging data, and text data. This enhances the sensitivity and accuracy of prediction models for complex disease representation. In scenarios such as chronic disease prediction, including diabetes, some studies have attempted to combine laboratory results, physiological monitoring data, and clinical text as joint model inputs. These approaches show advantages over traditional methods in modeling depth and expressive power. However, technical challenges remain in practical deployment. These include unreasonable data fusion strategies, imbalanced modality weights, and semantic redundancy[7].

In addition, multimodal fusion strategies in medical applications have shown diverse development patterns. Early fusion methods directly concatenate or encode features from different modalities into a unified framework. These methods are simple to implement and computationally efficient. Traditional fusion pipelines often ignore the semantic relationships and interaction mechanisms across modalities. In late fusion, each modality is processed separately and the outputs are combined only at the decision layer. Although this design can cope with strong modality heterogeneity, it limits feature interaction and weakens model consistency. Recent studies therefore turn to joint representation learning: a shared latent space with cross modal attention or graph based modelling is used to capture higher order dependencies. The YOLO-CBD framework, for example, combines behaviour level feature extraction with targeted aggregation to align visual evidence with contextual cues, which improves model expressiveness and interpretability [8].

Despite the growing potential of multimodal fusion in disease prediction, current research still faces limitations in areas such as data quality control, handling missing modalities, and incorporating clinical knowledge. In real-world clinical applications, EHR data often suffers from class imbalance and incomplete records, which pose challenges to model generalization. At the same time, designing cross-modal fusion mechanisms that align with clinical logic and disease characteristics remains a key issue. Future research needs to explore integration paths that combine structural awareness, semantic guidance, and temporal dependencies. These directions will help bridge the gap between technical validation and clinical implementation.

3. Proposed Methodology

This study designed an early prediction model for diabetes based on multimodal electronic medical record data, aiming to jointly model structured indicators, unstructured text, and time series features to achieve a comprehensive perception of potential risk states. The model architecture is shown in Figure 1.

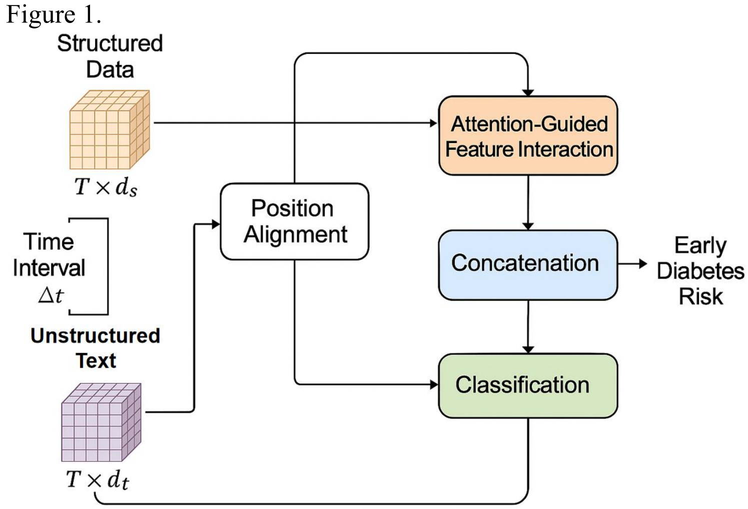

Figure 1. The overall architecture of the multimodal early diabetes risk prediction model based on electronic health records.

First, the structured input is set as a multidimensional numerical variable matrix $X_s \in R^{T \times d}$, which T represents the time step and d_s the structured feature dimension; the unstructured text is represented as a coded vector sequence $X_t \in R^{T\times d}$, and embedded extraction is performed through a pre-trained language model. To ensure the synchronization and correspondence of different modalities in the time dimension, a position alignment mechanism is used to match the temporal indexes X_s and X_t, thereby obtain a unified alignment representation $X_m = [X_s; X_t] \in R^{T\times(d_s+d_t)}$.

In the construction of fusion representation, an attention-guided feature interaction module is introduced to improve the model's ability to model cross-modal semantic associations. Specifically, for each time step t, the model calculates the query-key-value triple:

$$Q_t = X_m^t W_Q, K_t = X_m^t W_K, V_t = X_m^t W_V$$

Where $W_Q, W_K, W_V \in R^{(d_s+d_t)\times d_h}$ is the learnable parameter and d_h is the latent space dimension. Then, the weighted representation is calculated through the self-attention mechanism:

$$Z_t = \text{Softmax}(\frac{Q_t K_t^T}{\sqrt{d_h}})V_t$$

This process enables the model to capture significant interactions between different modal features and improves the ability to model potential risk patterns.

To further enhance the perception of continuous time features, the model introduces a time embedding module to map the original time interval vector $\Delta t \in R^T$ into a composite representation of periodicity and aperiodicity:

$$\phi(\Delta t) = [w\cdot \Delta t, \sin(w\cdot \Delta t)]$$

Where $w \in R$ is a learnable frequency parameter. This mechanism can effectively capture the non-uniform time distribution between medical events and help model potential physiological rhythms and behavioral patterns. The fused representation is concatenated with the original multimodal features as the input of the subsequent classification module.

Finally, the model extracts the joint semantics through the feedforward network and outputs the diabetes risk prediction result $\hat{y} \in [0,1]$. In the training phase, the binary cross entropy loss function is used for optimization, which is defined as:

$$L = -(y\cdot \log(\hat{y}) + (1-y)\cdot \log(1-\hat{y}))$$

Where $y \in [0,1]$ represents the true annotation label. In order to improve the model's adaptability to imbalanced sample distribution, weighted loss terms or adversarial perturbation mechanisms can be further introduced to enhance generalization and robustness. The entire method framework builds an end-to-end trainable prediction system from multimodal information representation and cross-modal alignment to temporal modeling and risk prediction.

4. Dataset

The multimodal dataset used in this study is composed of two publicly available medical data sources: EyePACS and MIMIC-III/IV. EyePACS is an image dataset for diabetic retinopathy screening. It contains a large number of color fundus images that provide visual information related to diabetes, such as hemorrhages, exudates, and microvascular abnormalities in the retina. These visual features are often regarded as key indicators in the progression of diabetes. Therefore, they serve as input for the visual modality to capture potential diabetic risk signs from the structure of the eye.

On the other hand, the MIMIC-III and MIMIC-IV datasets offer rich clinical electronic health record text data. This includes discharge summaries, examination records, medication information, and diagnostic descriptions. These data cover complete medical activities and disease trajectories. As inputs to the textual modality, they provide information about a patient's medical history, physiological state, and clinical reasoning. This helps to supplement symptoms and treatment factors that may not be directly visible in the visual modality. Text features related to diabetes are extracted from these records to construct structured representations.

During data preprocessing, the two datasets are aligned and merged to form paired multimodal samples of images and text. Sample labels are assigned based on confirmed diabetes diagnoses, categorized as either "diabetic" or "non-diabetic." To ensure data quality, image samples are normalized in size and filtered based on quality. Text data undergoes tokenization, stop word removal and medical entity recognition. The resulting multimodal dataset contains aligned cross-modal pairs, providing high-quality support for subsequent fusion and discriminative modeling tasks.

5. Experimental Results

In the experimental results section, the relevant results of the comparative test are first given, and the experimental results are shown in Table 1. This part is designed to evaluate the overall effectiveness of the proposed method by comparing it with several baseline models under consistent conditions. The comparison focuses on core performance metrics across different approaches, aiming to highlight the strengths and weaknesses of each method. Through this analysis, the study establishes a reference framework to better understand the relative advantages of the multimodal fusion strategy in the context of early diabetes prediction.

Table 1: Comparative experimental results

Method	**Accuracy**	**AUC**	**F1-Score**
CNN[9]	0.842	0.873	0.831
Transformer[10]	0.855	0.881	0.845
GNN[11]	0.836	0.866	0.820
CNN-Transformer[12]	0.869	0.892	0.856
LSTM-Transformer[13]	0.861	0.888	0.849
Ours	0.889	0.913	0.875

The results in the table show that traditional single-modality models, such as CNN and Transformer, achieve a certain level of performance in early diabetes prediction. CNN reaches an accuracy of 0.842 and an AUC of 0.873, while the Transformer reaches 0.855 and 0.881, respectively. These results suggest that the visual and textual modalities each have some discriminative capability. However, both models are limited in terms of information coverage and risk feature modeling. They fail to capture the semantic complementarity and structural coordination across modalities, leading to performance bottlenecks in the F1-score, especially under imbalanced data conditions.

In contrast, fusion models such as CNN-Transformer and LSTM-Transformer outperform the baseline models on all three metrics. Their F1 scores reach 0.856 and 0.849, respectively. These results indicate that by introducing sequential modeling and joint visual-textual perception mechanisms, the models can better handle temporal dependencies and modality complementarity present in early diabetes symptoms. Such fusion structures offer clear advantages in representing potential pathological signals and capturing intrinsic cross-modal associations. This validates the effectiveness of multimodal collaborative modeling strategies.

It is worth noting that although GNN models have strengths in certain structural tasks, their overall performance in this study is lower than other methods. This may be due to their limited ability to model dynamic semantic paths between visual and textual modalities. GNNs are less effective when dealing with non-graph-structured input data. This further highlights the strong dependence of early multimodal prediction tasks on semantic aggregation strategies and fusion mechanisms.

The method proposed in this study achieves the highest values across all three key metrics. In particular, it reaches an AUC of 0.913, indicating strong classification stability and risk identification ability. This performance advantage is attributed to the joint modeling of structured image features, unstructured textual semantics, and temporal embedding features. The use of channel attention mechanisms enhances the complementary expression of different modalities. As a result, the model significantly improves sensitivity and accuracy in detecting early signs of diabetes. The experimental results confirm the adaptability and broad application potential of the proposed method in multimodal electronic health record environments.

This paper then gives an experiment on the interference of medical text noise injection on the discrimination performance, and the experimental results are shown in Figure 2. The purpose of this experiment is to evaluate how varying levels of random noise in clinical text affect the model's ability to distinguish between different risk categories. By introducing controlled disturbances into the unstructured textual input, the study explores the robustness of the semantic processing component and examines the extent to which text integrity influences overall predictive stability in the multimodal setting.

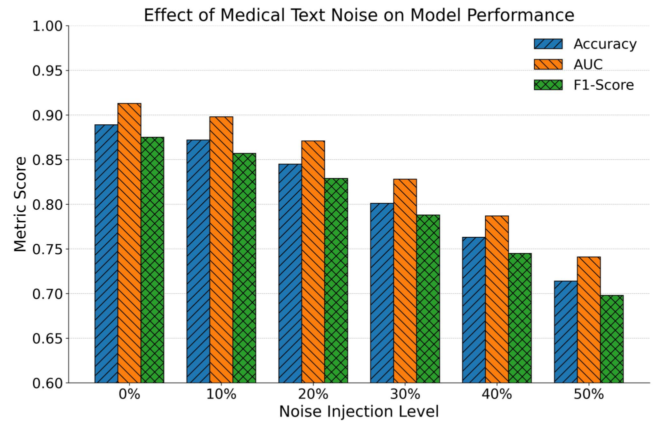

Figure 2. Experiment on the interference of medical text noise injection on discrimination performance

From the figure, it can be observed that as the proportion of noise injection into medical text increases, the model's performance on all evaluation metrics shows a clear downward trend. The decline becomes more pronounced when the noise level exceeds 30 percent. This indicates that the textual modality plays a critical role in early diabetes identification. The semantic integrity and structural clarity of text are essential for the model's comprehension and decision-making. When a large portion of the original text is disrupted by random noise, the model struggles to extract relevant risk cues, leading to degraded prediction results.

Further analysis shows that the AUC remains relatively stable across different levels of noise. At 10 percent or 20 percent noise, the AUC stays at a high level, suggesting that the model retains some discrimination ability even when partial semantic content is damaged. This may be attributed to the complementary effect of visual information in the multimodal structure. It provides redundancy when the textual input is incomplete or erroneous. However, when the noise level exceeds 40 percent, the AUC also drops rapidly. This suggests that the model can no longer effectively integrate information across modalities, and its overall perception ability significantly weakens.

At the same time, the F1-score is the most sensitive to noise injection. Under high-noise scenarios, it drops sharply. This change indicates that the model's ability to distinguish between positive and negative samples is impaired when dealing with complex textual interference. As a result, it becomes difficult to maintain a balance between recall and precision. Since early symptoms of diabetes are often implicitly expressed in text, noise directly reduces the model's capacity to detect key semantics. This leads to a simultaneous increase in both false positives and false negatives.

In conclusion, the experiment further confirms the importance of high-quality text data in multimodal diabetes prediction tasks. It also demonstrates that the model's performance strongly depends on the semantic completeness of textual input. Future model designs should pay greater attention to the robustness of the text modality and its resistance to noise.

This is essential for ensuring model stability and practical value when dealing with incomplete or corrupted real-world data.

This paper also gives the influence of a single mode on the experimental results, and the experimental results are shown in Figure 3.

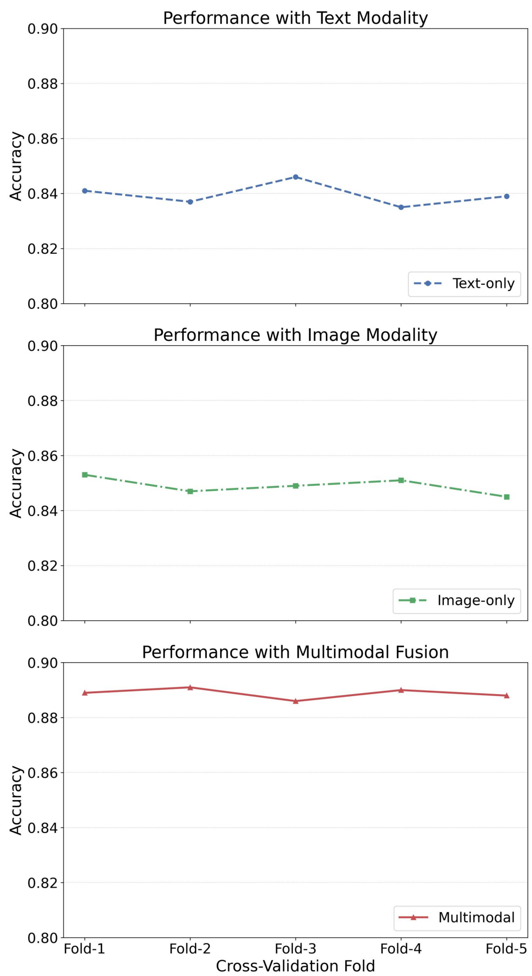

Figure 3. The influence of a single mode on the experimental results

From the results shown in the figure, it can be seen that under the condition of using only the text modality, the model maintains a relatively stable accuracy across five-fold cross-validation. The overall fluctuation is minimal, indicating a certain degree of semantic stability. This suggests that clinical text records do contain important clues related to diabetes risk, especially within patients' medical histories and clinical notes. However, due to the abstract nature of text and variability in expression, the model still struggles to capture comprehensive individual risk features through a single semantic channel.

Further examination of the visual modality results reveals that the model performs slightly better than the text-only input. The accuracy is more concentrated across folds. This is mainly because diabetic retinopathy exhibits more direct and observable features at the visual level. Image inputs help the model identify potential pathological patterns, such as microvascular damage, from structural characteristics. However, relying solely on the visual modality leaves semantic gaps. It cannot reflect critical patient information such as metabolic indicators or medication history, which limits the model's overall capacity.

It is worth emphasizing that the multimodal fusion approach achieves higher and more stable accuracy across all folds. This indicates a strong complementary relationship between structured image features and unstructured textual semantics. Through joint modeling, the model can capture various dimensions of diabetes manifestation, enhancing both predictive accuracy and robustness. The fusion mechanism effectively overcomes the limitations of single-modality expression and provides stronger support for early risk identification.

This paper also gives a sensitivity analysis of the impact of changes in data imbalance on model performance, and the experimental results are shown in Figure 4.

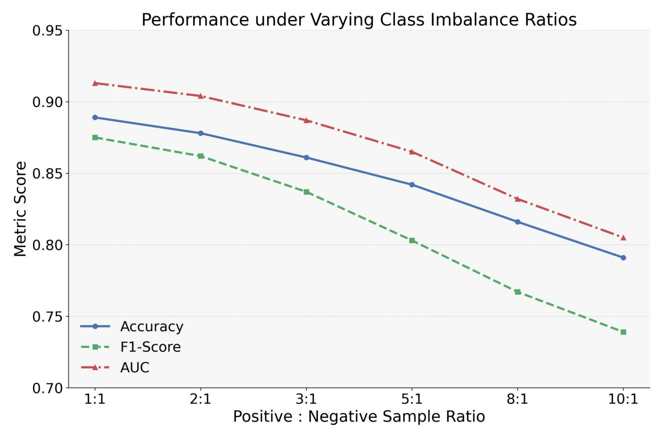

Figure 4. Sensitivity analysis of model performance to changes in data imbalance ratio

The results in the figure show that as the ratio of positive to negative samples becomes increasingly imbalanced, the overall model performance on all evaluation metrics declines. The performance drop becomes more pronounced when the ratio reaches 8:1 and 10:1. This trend indicates that the proposed multimodal prediction model is sensitive to data distribution. In particular, when positive samples are significantly scarce, the model's ability to detect risk signals is substantially disrupted, leading to reduced classification accuracy.

A closer examination reveals that the F1-score declines more sharply than Accuracy and AUC. This suggests a shift in the model's predictive balance under imbalanced conditions. The model becomes biased toward the majority class, resulting in reduced recall and poor overall balance. This issue is especially critical in medical risk identification tasks, where missing high-risk cases can directly compromise early intervention and affect subsequent clinical decisions.

At the same time, AUC, which measures the overall ranking ability of the model, shows relatively better resistance to imbalance, although it also exhibits a clear downward trend. This may be because the model can still maintain some level of boundary structure through multimodal information fusion.

However, it remains inadequate in delineating the decision boundaries for minority classes. This highlights a limitation in the current modeling strategy when handling severely skewed sample distributions.

This paper also gives the impact of time interval embedding granularity setting on the model prediction effect, and the experimental results are shown in Figure 5.

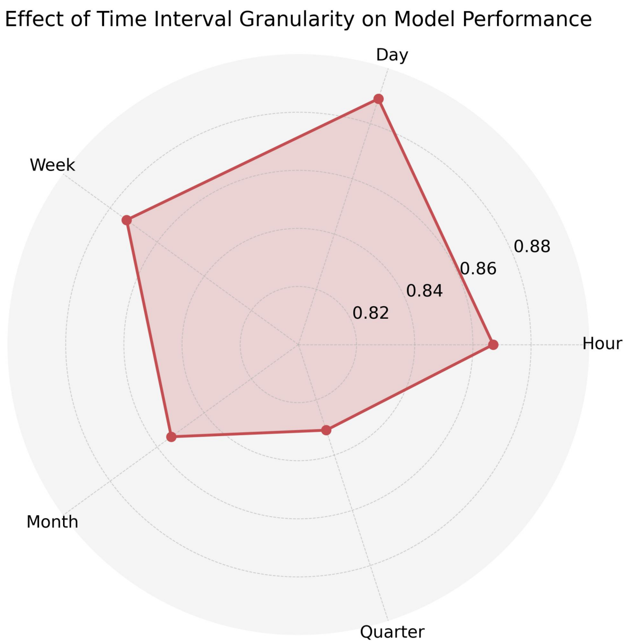

Figure 5. The impact of time interval embedding granularity setting on model prediction performance

The figure shows that the granularity of time interval embedding has a clear impact on the model's predictive performance. The best result is achieved at the "Day" level, where the model reaches a performance close to 0.89. This indicates that medium-granularity time embeddings are more effective in capturing key temporal relationships between events in medical scenarios. They avoid the noise introduced by overly fine granularity while preserving enough temporal signals to model the progression of risk.

In comparison, the "Hour" and "Week" granularities result in slightly lower performance, though still at a relatively high level. This suggests that the model has some adaptability to local temporal fluctuations. However, "Hour-level" embeddings may introduce excessive detail, which interferes with the modeling of long-term trends. On the other hand, "Week" granularity may weaken the model's sensitivity to closely linked events, leading to performance instability.

When the granularity is further expanded to "Month" and "Quarter," the model performance drops significantly. The "Quarter" level, in particular, shows the lowest result. This suggests that overly coarse time encoding dilutes the sequential nature of critical clinical events. As a result, the model struggles to capture the logic of disease progression. This issue is especially evident in chronic conditions like diabetes, where stage-specific patterns are important.

These findings demonstrate that the choice of time embedding granularity plays a key role in multimodal modeling. It affects not only the fidelity of temporal information but also the alignment between semantic and structural features. In practical applications, the time granularity should be adjusted based on task requirements and data characteristics to balance detail retention with generalization.

6. Conclusion

This study addresses key challenges in the early identification of diabetes and proposes a multimodal prediction method based on electronic health records. The method fully leverages the complementary characteristics of structured medical images and unstructured clinical texts. By introducing attention mechanisms and time interval embeddings, it enables joint modeling and semantic perception of patients' multidimensional health states. The experimental design includes model structure comparison, modality ablation analysis, and various sensitivity tests. The results validate the robustness and adaptability of the proposed method under different data conditions, providing a practical solution for disease prediction in complex medical scenarios.

By incorporating cross-modal alignment mechanisms and fine-grained temporal modeling strategies, the proposed multimodal fusion framework significantly enhances the model's ability to capture early features of diabetes. It overcomes the representational limitations of single modalities. The results show that in the presence of complex structures and dispersed semantics in electronic health records, the fusion of multi-source information offers stronger generalization and discrimination capabilities. This enables more accurate identification of individual risk states from multiple perspectives. The method holds practical significance for improving chronic disease screening efficiency and optimizing healthcare resource allocation.

In addition, this study systematically explores the model's stability when faced with common medical application challenges, such as data imbalance, noise interference, and variations in temporal granularity. These analyses reveal key factors affecting the deployment of multimodal models in real-world settings. They not only provide empirical support for model usability but also lay a foundation for future improvements in multimodal medical AI systems. Overall, the proposed method demonstrates generalizability in understanding complex medical data and performing intelligent classification. It shows potential for transfer across different diseases and healthcare systems.

7. Future work

Future research may extend to multi-center data integration, dynamic modeling of heterogeneous modalities, and self-supervised learning guided by clinical knowledge. These directions aim to improve the model's ability to perceive unknown pathological structures and enhance task adaptability. Combined with lightweight deployment strategies and user interaction feedback mechanisms, this approach could support the development of efficient and interpretable intelligent screening systems for real clinical environments. It may further promote the practical implementation of AI-assisted healthcare systems in chronic disease prevention and control.

References

[1] Refat M A R, Al Amin M, Kaushal C, et al. A comparative analysis of early stage diabetes prediction using machine learning and deep learning approach[C]//2021 6th International Conference on Signal Processing, Computing and Control (ISPCC). IEEE, 2021: 654-659.

[2] Afsaneh E, Sharifdini A, Ghazzaghi H, et al. Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review[J]. Diabetology & Metabolic Syndrome, 2022, 14(1): 196.

[3] Gupta H, Varshney H, Sharma T K, et al. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction[J]. Complex & Intelligent Systems, 2022, 8(4): 3073-3087.

[4] Aslan M F, Sabanci K. A novel proposal for deep learning-based diabetes prediction: converting clinical data to image data[J]. Diagnostics, 2023, 13(4): 796.

[5] Naseem A, Habib R, Naz T, et al. Novel Internet of Things based approach toward diabetes prediction using deep learning models[J]. Frontiers in Public Health, 2022, 10: 914106.

[6] Zhu T, Li K, Herrero P, et al. Deep learning for diabetes: a systematic review[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(7): 2744-2757.

[7] Pal S, Mishra N, Bhushan M, et al. Deep learning techniques for prediction and diagnosis of diabetes mellitus[C]//2022 International mobile and embedded technology conference (MECON). IEEE, 2022: 588-593.

[8] Peng, S., Zhang, X., Zhou, L., & Wang, P. “YOLO-CBD: Classroom Behavior Detection Method Based on Behavior Feature Extraction and Aggregation.” Sensors, 25 (10), 3073 (2025).

[9] Madan P, Singh V, Chaudhari V, et al. An optimization-based diabetes prediction model using CNN and Bi-directional LSTM in real-time environment[J]. Applied Sciences, 2022, 12(8): 3989.

[10] Zhu T, Kuang L, Piao C, et al. Population-specific glucose prediction in diabetes care with transformer-based deep learning on the edge[J]. IEEE transactions on biomedical circuits and systems, 2024, 18(2): 236-246.

[11] Lei Y, Lin S, Li Z, et al. GNN-fused CapsNet with multi-head prediction for diabetic retinopathy grading[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 107994.

[12] Rezaee K, Farnami F. Innovative Approach for Diabetic Retinopathy Severity Classification: An AI-Powered Tool using CNN-Transformer Fusion[J]. Journal of Biomedical Physics & Engineering, 2025, 15(2): 137.

[13] Xiong X, Yang X L, Cai Y, et al. Exploring the potential of deep learning models integrating transformer and LSTM in predicting blood glucose levels for T1D patients[J]. Digital Health, 2025, 11: 20552076251328980.