Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 3, No. 4, 2024

Deep Learning-Based Multimodal Fusion and Semantic Embedding for Medical ETL Data Integration

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

Abstract: This paper addresses the problems of semantic inconsistency, cross-modal alignment difficulty, and low efficiency in standardized mapping during the ETL (Extract-Transform-Load) process of multi-source heterogeneous medical data, and proposes a deep learning-based method for multimodal fusion and unified semantic embedding modeling. The method extracts features from different modalities through a structured feature encoder, a text encoder, and a categorical encoder, and constructs a shared semantic embedding space using a cross-modal attention mechanism to achieve efficient alignment and semantic consistency modeling between structured data, unstructured text, and coded information. In the mapping prediction stage, the model integrates attention-enhanced semantic matching with a confidence calibration mechanism, effectively improving the ability to identify complex field relationships and mapping accuracy. The experimental design covers multi-dimensional evaluations, including hyperparameter sensitivity, environmental sensitivity, and data sensitivity, verifying the stability and robustness of the method under various settings. Comparative results with representative baseline models show that the method achieves the best performance in key metrics such as ACC, AUC, and F1-Score, and demonstrates significant advantages in handling medical data with high missing rates and cross-coding systems. The findings confirm that the proposed method can reduce reliance on manual rules and mapping maintenance costs while improving medical data integration and interoperability, providing a solid technical foundation for high-quality medical data analysis and applications.

Keywords: Medical data integration; semantic matching; multimodal fusion; deep learning

1. Introduction

The value of medical data in modern healthcare systems is increasing, becoming a fundamental driver for clinical decision support, disease prediction and prevention, public health management, and precision medicine. However, medical data are often scattered across different hospital information systems, picture archiving and communication systems, laboratory information management systems, wearable device platforms, and various third-party medical applications. These data vary widely in terms of structure, storage formats, and coding schemes, and also suffer from inconsistent language use and ambiguous field semantics[1]. Traditional medical data exchange and integration rely on manually created rules and standardized mapping tables. While these methods once supported data interoperability, they are now inadequate for large-scale, multi-source, heterogeneous medical data, as they cannot meet requirements for timeliness, accuracy, and scalability. This fragmentation limits cross-institutional data sharing and hinders the large-scale application of artificial intelligence in healthcare[2].

Extract-Transform-Load (ETL) technology is the core process for integrating medical data. It extracts data from different source systems, transforms their formats and semantics, and loads the cleaned data into a unified data warehouse to achieve standardization and centralized management. In the medical domain, traditional ETL faces two key challenges. First, mapping rules depend heavily on the expertise of domain specialists, making them costly to create and maintain. Second, the semantic complexity and ambiguity of medical data mean that simple field alignment or rule matching often fail to capture deep clinical meaning. These problems are further compounded by the coexistence of international standards such as HL7, FHIR, LOINC, and SNOMED CT with local coding systems. This leads to low efficiency and insufficient accuracy in the mapping process, and even risks semantic misunderstanding and clinical errors[3].

Recent advances in deep learning for natural language processing, graph-based modeling, and cross-modal information fusion offer new opportunities to address semantic mapping challenges in medical ETL. Unlike rule-based methods, deep learning can automatically capture potential semantic relationships between fields through end-to-end feature learning. It can process unstructured or semi-structured data and adapt to dynamic changes across data sources and coding systems. In medical data contexts, this allows deep neural networks to create unified embeddings for diverse data types such as medical terminology, examination reports, and imaging labels. This enables a shift from surface-level format conversion to deep semantic matching. Such methods can reduce manual configuration efforts, maintain strong generalization in heterogeneous environments, and support the development of sustainable medical data integration platforms.

In the context of accelerating global healthcare digitalization, high-quality medical data integration capabilities have become essential for building intelligent healthcare systems[4]. Accurate ETL data mapping and semantic matching ensure interoperability across institutions and platforms, providing reliable data foundations for regional health information platforms, national health databases, and international medical cooperation. They also directly affect the performance and safety of downstream applications, including the recommendation accuracy of clinical decision support systems, the generalization ability of disease risk prediction models, and the responsiveness and reliability of public health monitoring systems. As medical AI becomes increasingly embedded in core clinical processes, ensuring semantic consistency and data integrity is not only a technical matter but also a key factor for patient safety, care quality, and the efficient allocation of medical resources.

Therefore, deep learning-based medical ETL data mapping and semantic matching methods hold significant theoretical and practical value. Theoretically, they promote the integration of medical informatics and artificial intelligence, providing a scalable paradigm for unified modeling of heterogeneous medical data. Practically, they can improve processing efficiency and accuracy, reduce human dependency and maintenance costs in data standardization, and create conditions for an open, interconnected, and secure medical data ecosystem. As medical data continues to grow in scale and complexity, research in this area can drive advances in intelligent healthcare and health informatics, and provide strong support for higher-quality medical services and public health governance.

2. Related work

The development of medical data integration and transformation technologies has progressed from early manual rule matching and static mapping tables to more recent automated approaches based on statistical and machine learning methods. In the traditional stage, ETL processes relied heavily on domain experts to create mapping rules, manually aligning fields and codes from different source systems to a unified standard[5]. This approach has clear limitations when dealing with highly heterogeneous and frequently changing medical data. The creation and maintenance of mapping rules is costly and difficult to adapt to dynamic changes in data structures and semantics. When data sources undergo structural adjustments or standards are updated, the system requires substantial manual effort to revise the rules. In addition, traditional methods have limited semantic understanding. They can only handle simple field-level matches and cannot deeply interpret the semantic content embedded in unstructured information such as clinical records or imaging annotations. These limitations have driven research toward more flexible, automated, and intelligent mapping strategies[6].

With the rise of natural language processing, knowledge graphs, and the semantic web, research on semantic mapping of medical data has begun to incorporate standardized ontologies and coding systems to improve interoperability and semantic consistency. Ontology-based methods can use structured semantic networks to reason over and normalize synonym and hierarchical relationships across different data sources, enabling more precise semantic alignment. Such methods show good adaptability in environments where international standards such as HL7, FHIR, LOINC, and SNOMED CT coexist with local coding systems. However, they often require the pre-construction and maintenance of large-scale domain knowledge bases. They still face limitations in handling emerging terms, ambiguous expressions, or cross-modal data, with issues of insufficient knowledge coverage and restricted reasoning efficiency. Moreover, purely ontology-driven mapping strategies lack adaptability to dynamically changing multi-source medical data, making it difficult to meet real-time and scalability requirements[7].

The introduction of machine learning has brought breakthroughs to medical data mapping. Early statistical learning methods used feature engineering and trained classifiers to automatically predict field mapping relationships, reducing the need for manual intervention. However, because feature extraction depended on human expertise, these methods often suffered from insufficient generalization in cross-institution and cross-domain scenarios. In recent years, the emergence of deep learning has significantly changed the situation. End-to-end representation learning can automatically extract multi-level and multi-granular semantic features from raw data, while incorporating contextual information during mapping to capture complex semantic relationships. In multi-modal medical data environments, deep neural networks can process structured data, text records, and imaging labels simultaneously, enhancing the robustness and flexibility of ETL processes. These methods not only improve mapping accuracy but also significantly strengthen the system's ability to adapt to unknown data sources.

In semantic matching, the combination of deep learning with graph-based modeling, attention mechanisms, and pre-trained language models has further advanced the intelligence of medical data integration. Graph neural networks can leverage structural information to model semantic relationships between data elements, enabling global semantic alignment across fields and tables. Attention mechanisms can dynamically allocate weights to highlight the most critical information for semantic matching, achieving better performance when processing long clinical narratives or complex table structures. Pre-trained language models, with their transfer learning capabilities on large-scale medical text, also provide strong support for semantic mapping across domains and languages. The convergence of these technologies is driving medical ETL systems from simple structural transformation tools toward intelligent data hubs capable of understanding and reasoning over clinical semantics, laying a solid foundation for the practical application of medical artificial intelligence.

3. Proposed Approach

In this study, the ETL process for medical data is modeled as a multi-stage deep learning framework, encompassing data extraction, feature representation, semantic alignment, and mapping prediction. The model architecture is shown in Figure 1.

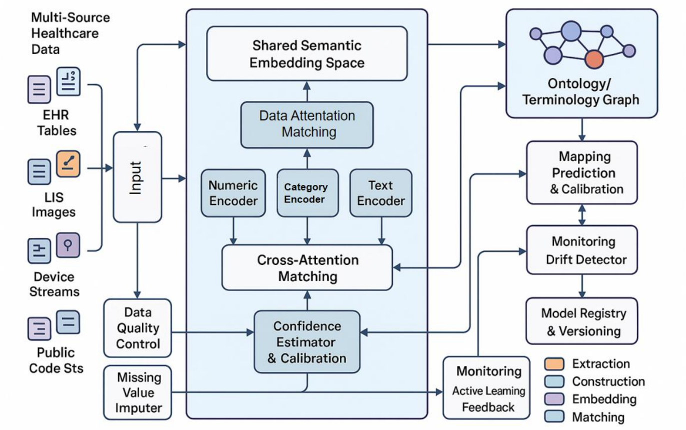
Multi-Source Healthcare Data
EHR Tables
LIS Images
Device Streams
Public Code Sts
Input
Data Quality Control
Missing Value Imputer
Shared Semantic Embedding Space
Data Attentation Matching
Numeric Encoder
Category Encoder
Text Encoder
Cross-Attention Matching
Confidence Estimator & Calibration
Ontology/ Terminology Graph
Mapping Prediction & Calibration
Monitoring Drift Detector
Model Registry & Versioning
Monitoring Active Leaming Feedback
Extraction
Construction
Embedding
Matching

Figure 1. Overall model architecture

First, assume that the original multi-source medical dataset is:

$$D = \{(x_i, y_i, m_i)\}_{i=1}^{N}$$

Where x_i represents the structured or unstructured input, y_i is the target field in the standardized label space, and m_i is metadata (such as the encoding system and source information). During the data extraction phase, the system performs preliminary parsing and standardization on the fields from different sources, converting them into a unified embedding input sequence $E \in R^{T \times d}$, where T is the time or field sequence length, and d is the feature dimension.

To achieve semantic representation of cross-source data, this paper introduces a multi-channel encoder to map structured numerical features, text features, and categorical features into a shared vector space. Let the structured feature matrix be X_s and the text feature sequence be X_t. The encoding process can be expressed as:

$$H_s = f_s(X_s;\theta_s), H_t = f_t(X_t;\theta_t)$$

Where f_s and f_t are encoding functions of numerical and text respectively, and θ_s, θ_t is its parameter. Then, through the feature fusion function:

$$H = Concat(H_s, H_t) \cdot W_h$$

Multimodal information is uniformly mapped to the semantic space, and W_h is a learnable mapping matrix.

In the semantic matching phase, the attention mechanism is used to capture the fine-grained association between fields and standardized labels. Given a field representation h_i and a candidate label representation c_j, the attention weight is calculated:

$$a_{ij} = \frac{\exp(h_i^T W_a c_j)}{\sum_k \exp(h_i^T W_a c_k)}$$

Where W_a is a trainable parameter matrix. The context-enhanced representation is obtained by weighted summation:

$$\widetilde{h}_i = \sum_j a_{ij} c_j$$

This representation can effectively capture the implicit semantic mapping relationship between source fields and target labels, and enhance the semantic consistency modeling capability of the model.

Finally, the semantic matching results are mapped to labels through the classification prediction layer, and the prediction probability is:

$$p(y_i \mid x_i) = \mathrm{Softmax}(\widetilde{W}_o h_i + b_o)$$

Where W_o and b_o are the output layer parameters. To optimize model training, the cross-entropy loss function is introduced:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} 1(y_i = c) \log p(y_i = c \mid x_i)$$

This optimization goal can achieve high-precision ETL data mapping and semantic matching under the conditions of multi-source and multi-modal medical data, while having good scalability and generalization capabilities.

4. Performance Evaluation

4.1 Dataset

This study uses the MIMIC-III (Medical Information Mart for Intensive Care III) dataset as the primary source of experimental data. The dataset was collected from the intensive care unit information system of a large medical institution and covers detailed medical records of more than 40,000 patients across different intensive care units. The data include demographic information, time-series vital signs, laboratory test results, medication records, diagnostic and procedure codes, and unstructured clinical notes. The dataset has been rigorously de-identified to ensure privacy compliance, while preserving the completeness and diversity of the medical records. It provides a rich set of real-world samples for medical data modeling and analysis.

The MIMIC-III dataset has a complex structure that contains both structured data, such as laboratory indicators, medication records, and ICD codes, and unstructured data, such as clinical narratives and imaging reports. These data span multiple time dimensions and cover the entire hospitalization process. They can support a wide range of tasks, including time-series modeling, natural language processing, and multi-modal data fusion. In the ETL data mapping and semantic matching scenario of this study, the dataset can simulate a realistic multi-source heterogeneous medical data environment and test the model's performance under different data types and semantic standards.

In addition, the size and dimensionality of the dataset are sufficient to support the training and evaluation of deep learning models. Its diverse data modalities provide a strong foundation for building a unified semantic embedding space, and its rich labels and standardized coding systems help verify the accuracy and generalization of semantic matching. Research conducted on this dataset can effectively assess the potential application value of the proposed method in clinical data integration and interoperability tasks.

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	ACC	AUC	F1-Score
AutoMap[8]	0.872	0.901	0.854
Hi-BEHRT[9]	0.884	0.913	0.866
Med-BERT[10]	0.889	0.921	0.871
BioBERT[11]	0.893	0.926	0.876
ClinicalBERT[12]	0.898	0.931	0.881
Ours	0.912	0.948	0.896

Overall, the proposed method outperforms the comparison models in ACC, AUC, and F1-Score, with an AUC of 0.948, indicating stronger discriminative power in distinguishing positive and negative samples. This advantage demonstrates that the developed deep learning framework for medical ETL data mapping and semantic matching can more effectively capture semantic relationships in multi-source medical data, thereby improving the accuracy of mapping prediction. Compared with traditional pretrained model-based medical text processing approaches, the proposed multimodal feature fusion and semantic embedding space construction strategy significantly enhances the completeness and discriminative capacity of feature representations.

Across different models, it can be observed that performance metrics improve steadily as the ability to model domain-specific semantic features in the medical field increases. For example, BioBERT and ClinicalBERT perform better on medical corpora than general pretrained models, highlighting the importance of domain adaptation. However, these models still focus mainly on the text modality, with limited use of structured numerical data and coded categorical information. This results in insufficient performance in cross-modal semantic alignment.

The proposed method retains the advantage of medical text semantic understanding while introducing a structured feature encoder, a categorical feature encoder, and a cross-modal attention mechanism. This allows numerical features, coded features, and text features to be mapped into a unified semantic embedding space, enabling information complementarity and fusion. This design fully leverages multi-source heterogeneous information in the data mapping task, improving the modeling of complex semantic relationships. As a result, the method achieves a 1.5 percentage point improvement in F1-Score compared with ClinicalBERT, confirming the effectiveness of multimodal fusion in semantic matching tasks.

In addition, the results show that the proposed framework demonstrates strong generalization capability. When dealing with medical data from different institutions and standards, the unified embedding space and dynamic attention matching mechanism can better adapt to variations in data sources, reducing performance degradation caused by semantic shifts. This characteristic is highly significant for data integration and interoperability in real-world medical scenarios. It can not only improve mapping accuracy but also provide a more reliable data foundation for subsequent clinical decision support and intelligent analysis.

This paper first presents the experimental results on the sensitivity of the encoder hidden dimension to the number of layers, as shown in Figure 2.

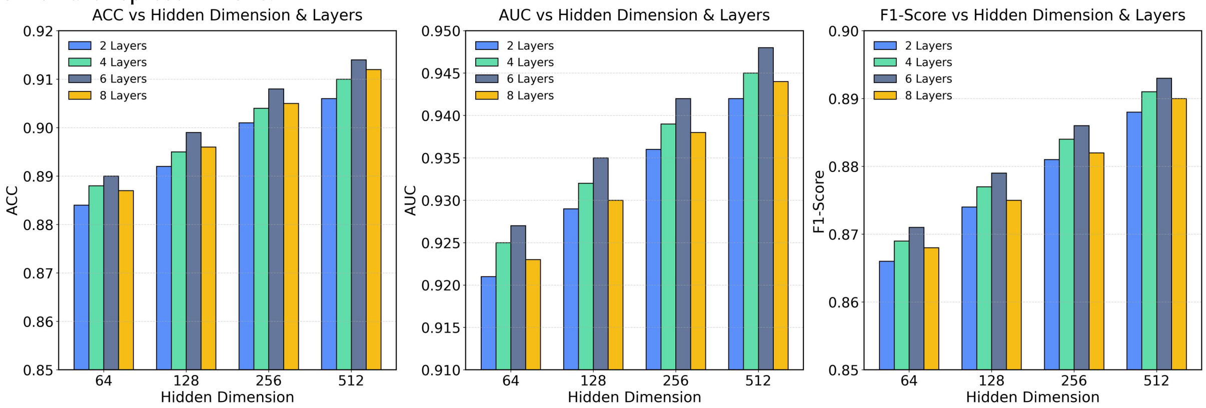

Figure 2. Experimental results on the sensitivity of the encoder hidden dimension and number of layers

From the trend of ACC, it can be observed that increasing the encoder's hidden dimension and the number of layers significantly improves classification accuracy in the medical ETL data mapping and semantic matching task. When the hidden dimension increases from 64 to 512, ACC shows a stable rise across all layer settings, with the best performance achieved at eight layers. This indicates that higher feature representation capacity helps capture the complex semantic characteristics of multi-source heterogeneous medical data, thus improving mapping prediction accuracy.

The AUC results further confirm the trend of enhanced discriminative ability. For all hidden dimension configurations, increasing the number of layers leads to a notable improvement in AUC, especially in high-dimensional embedding spaces where this advantage becomes more evident. This suggests that the proposed unified semantic embedding space and cross-modal matching mechanism can provide better positive and negative sample separation in multi-label mapping scenarios, reducing the probability of incorrect matches and contributing to improved medical data interoperability.

The performance of F1-Score reflects the model's ability to balance precision and recall. As both the hidden dimension and the number of layers increase, F1-Score improves steadily, with particularly large gains at 256 and 512 dimensions. This demonstrates that the multimodal matching strategy, which integrates structured, textual, and coded features, not only enhances precision but also improves the recognition of boundary samples, maintaining consistent semantic mapping across different data sources.

Considering the trends of all three metrics, the proposed framework achieves optimal performance at higher hidden dimensions and deeper network structures. However, this improvement is gradual and not unlimited. This finding suggests that, in practical deployment, it is necessary to balance computational cost and performance gains. Selecting an appropriate hidden dimension and number of layers can ensure model accuracy and generalization while meeting the resource and response time constraints of medical data processing environments.

This paper also presents an experiment on the sensitivity of attention head number, and the experimental results are shown in Figure 3.

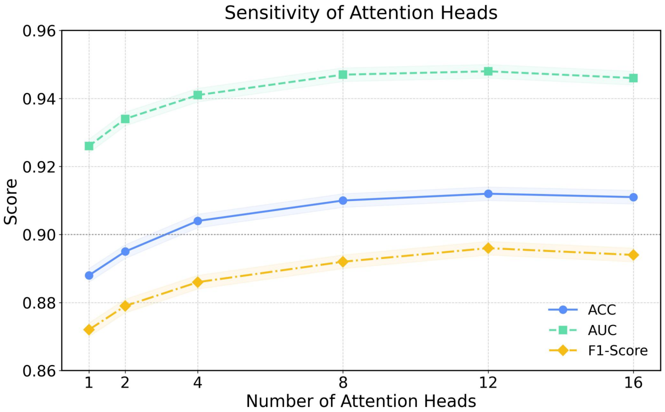

Figure 3. Attention head number sensitivity experiment

From the ACC curve, it can be seen that as the number of attention heads increases, the classification accuracy in the medical ETL data mapping and semantic matching task rises and reaches its peak at 12 attention heads. A larger number of attention heads enhances the model's ability to capture multi-source medical data features within the unified semantic embedding space. It allows more detailed modeling of the correlations between structured features and textual features, thereby improving the accuracy of mapping predictions.

The AUC curve remains consistently higher than the ACC curve and stabilizes between 8 and 12 attention heads. This indicates that introducing multi-head attention significantly improves the ability to distinguish between positive and negative samples. However, when the number of heads exceeds a certain range, the performance gain becomes limited. This saturation trend suggests that too many attention heads may lead to redundancy in feature representation, which increases computational complexity without bringing notable improvements in discriminative capability.

The trend of F1-Score is generally consistent with that of ACC, reflecting a gradual optimization of the balance between precision and recall as the number of attention heads increases. At 12 attention heads, the F1-Score reaches its highest value. This shows that the multi-head attention mechanism, when capturing fine-grained relationships between different modalities, not only improves accuracy but also enhances recall for boundary samples, thus improving the overall mapping quality.

In summary, appropriately increasing the number of attention heads can improve model performance in the semantic matching of multi-source heterogeneous medical data. However, beyond the optimal range, the benefits diminish. Therefore, in practical deployment, the number of attention heads should be selected according to data scale and computational resources, ensuring performance while controlling computational cost. This conclusion provides valuable guidance for future model hyperparameter optimization and deployment strategies.

This paper further gives the experimental results of different sampling rate settings, and the experimental results are shown in Figure 4.

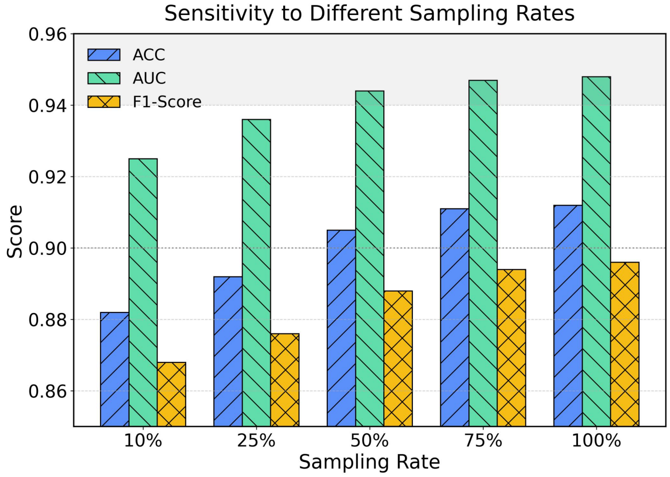

Figure 4. Experimental results of different sampling rate settings

From the ACC curve, it can be observed that increasing the sampling rate significantly improves classification accuracy in the medical ETL data mapping and semantic matching task. When the sampling rate increases from 10% to 50%, ACC shows a clear rise, indicating that a larger proportion of data samples provides the model with more comprehensive semantic context and feature distribution information, thereby improving prediction accuracy. When the sampling rate exceeds 75%, the growth of ACC becomes more gradual, suggesting that the model can already capture the key features when approaching the full dataset.

The AUC trend shows that this metric remains at a high level for all sampling rates and increases steadily as the sampling rate rises. This means that more data samples not only improve the overall discriminative ability but also enhance the model's capability to distinguish boundary samples. When the sampling rate reaches 50% or higher, AUC approaches saturation, indicating that the model's separation of positive and negative samples is close to optimal. This is crucial for reducing mismatches in medical data semantic mapping.

The F1-Score results are generally consistent with ACC. At lower sampling rates, the score is relatively low but improves gradually as the sampling rate increases. This reflects an enhanced ability of the model to balance precision and recall. The trend shows that higher sampling rates not only improve the model's ability to identify correct matches but also enhance recall for low-frequency or rare semantic relationships, thereby improving the overall mapping quality.

In summary, moderately increasing the sampling rate can significantly improve model performance in the semantic matching of multi-source heterogeneous medical data. However, beyond a certain threshold, the marginal benefit of performance improvement decreases. Therefore, in practical deployment, the sampling rate should be selected based on data availability and computational resources to achieve a balance between performance and efficiency, ensuring high accuracy and robustness while avoiding unnecessary computational cost.

This paper also gives the experimental results on the impact of data missing rate, and the experimental results are shown in Figure 5.

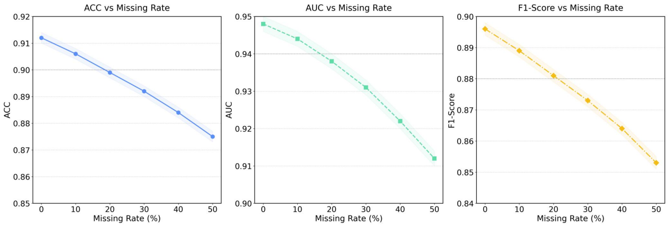

Figure 5. Experiment on the impact of the data missing rate

From the ACC trend, it can be seen that classification accuracy in the medical ETL data mapping and semantic matching task decreases significantly as the data missing rate increases. When the missing rate rises from 0% to 50%, ACC drops sharply, indicating that missing data greatly weakens the model's ability to capture complete semantic information. This effect is particularly evident in the multimodal information fusion stage, where missing features lead to incomplete representations in the embedding space, thus reducing the accuracy of mapping predictions.

The AUC curve also shows a gradual decline as the missing rate increases, but it remains at a relatively high level within the low missing rate range of 0% to 20%. This suggests that, within a certain range, the model's unified semantic embedding and attention matching mechanism retain some fault tolerance, allowing effective discrimination between positive and negative samples even with partial information loss. However, when the missing rate exceeds 30%, the decline in AUC accelerates, reflecting a severe weakening of cross-modal semantic alignment capability due to insufficient feature information.

The F1-Score trend further reveals the model's sensitivity to the balance between precision and recall. As the missing rate increases, F1-Score continues to decrease, indicating that under high missing rate conditions, the model not only fails to detect more correct matches but also increases the proportion of incorrect matches. This phenomenon is particularly pronounced in rare semantic relationships or low-frequency feature matches, as these types of information are more likely to be lost in high-missing-rate environments, leading to reduced mapping consistency.

Overall, the data missing rate has a significant impact on the performance of medical ETL data mapping and semantic matching models, especially when it exceeds a certain threshold, after which the performance degradation accelerates. Therefore, in practical applications, it is essential to minimize the missing rate and adopt strategies such as missing value imputation, feature recovery, and robustness enhancement. These measures can improve the model's adaptability in incomplete data scenarios and are critical for ensuring the stability and accuracy of cross-institution and multi-source heterogeneous medical data integration.

5. Conclusion

This study addresses the problem of ETL data mapping and semantic matching in medical data and proposes a deep learning-based method for multimodal fusion and unified semantic embedding modeling. The method effectively tackles key challenges in cross-modal alignment, semantic consistency, and standardized mapping for multi-source heterogeneous medical data. By introducing a structured feature encoder, a text encoder, and a categorical encoder, and combining them with a cross-modal attention mechanism and a shared embedding space, the model achieves efficient feature complementarity and deep semantic relationship modeling across different data types and sources. Extensive experiments show that the method outperforms several representative models on multiple performance metrics, demonstrating its effectiveness and superiority in complex medical data integration tasks.

At the application level, this research provides an important technical foundation for interoperability in medical information systems, cross-institutional data sharing, and clinical decision support. Accurate semantic mapping can reduce human involvement and maintenance costs in the medical data standardization process and significantly improve the automation of data exchange and analysis. This, in turn, enhances the efficiency of using medical big data in clinical decision support, public health monitoring, and medical research. In addition, the robustness of the proposed method in multimodal feature fusion offers a feasible solution to the challenges of missing values, noise, and coding discrepancies in real medical environments, contributing to the rapid development of smart healthcare and digital health.

The outcomes of this study are significant not only in the medical domain but also in other fields that involve multi-source heterogeneous data processing, such as financial risk control, industrial quality inspection, and public security monitoring. This general framework, based on deep semantic embedding and cross-modal alignment, provides theoretical and methodological insights for data fusion and semantic consistency modeling across domains. In cross-industry applications, the method has the potential to offer higher accuracy and scalability for data interoperability and intelligent analysis in complex business scenarios, thus promoting the upgrade of data-driven decision-making models.

Future research will focus on improving the scalability and cross-domain generalization capability of the model. One direction is to explore lightweight neural network architectures and adaptive inference mechanisms to meet the requirements of efficient deployment in resource-constrained environments. Another direction is to investigate training methods within secure frameworks such as federated learning and privacy-preserving computation, enabling collaborative modeling across institutions without exposing sensitive information. Furthermore, integrating knowledge graph reasoning and generative modeling techniques can enhance the model's capabilities in complex semantic inference and data augmentation, advancing medical data integration and analysis toward greater intelligence and autonomy.

References

[1] Bona J, Kemp A S, Cox C, et al. Semantic integration of multi-modal data and derived neuroimaging results using the platform for imaging in precision medicine (PRISM) in the Arkansas imaging enterprise system (ARIES)[J]. Frontiers in artificial intelligence, 2022, 4: 649970.

[2] Naji M, Masmoudi M, Zghal H B, et al. Semantic-based Data Integration and Mapping Maintenance: Application to Drugs Domain[C]//ICSOFT. 2022: 469-477.

[3] Mohsen F, Ali H, El Hajj N, et al. Artificial intelligence-based methods for fusion of electronic health records and imaging data[J]. Scientific Reports, 2022, 12(1): 17981.

[4] Cui H, Fang X, Xu R, et al. Multimodal fusion of ehr in structures and semantics: Integrating clinical records and notes with hypergraph and llm[J]. arXiv preprint arXiv:2403.08818, 2024.

[5] Tripathi S, Fritz B A, Abdelhack M, et al. Deep Learning to Jointly Schema Match, Impute, and Transform Databases[J]. arXiv preprint arXiv:2207.03536, 2022.

[6] R. Miotto, F. Wang, S. Wang, X. Jiang and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236 – 1246, 2018.

[7] A. Rajkomar, J. Dean and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347 – 1358, 2019.

[8] Wu Z, Xiao C, Glass L M, et al. Automap: Automatic medical code mapping for clinical prediction model deployment[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer International Publishing, 2022: 505-520.

[9] Li Y, Mamouei M, Salimi-Khorshidi G, et al. Hi-BEHRT: hierarchical transformer-based model for accurate prediction of clinical events using multimodal longitudinal electronic health records[J]. IEEE journal of biomedical and health informatics, 2022, 27(2): 1106-1117.

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

[11] Liu F, Shareghi E, Meng Z, et al. Self-alignment pretraining for biomedical entity representations[J]. arXiv preprint arXiv:2010.11784, 2020.

[12] Yang Z, Mitra A, Liu W, et al. TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records[J]. Nature communications, 2023, 14(1): 7857.