Journal of Machine Intelligence in Healthcare

E-ISSN: XXXX-XXXX
P-ISSN: XXXX-XXXX

(Research/Review) Article

Lab-Integrated Multi-Aspect Pretraining for Prognostic EHR Modeling

Nova Ariyanto [1,*] and Sherly Nur Ekawati [2,3]

[1] Department of Informatics, Universitas Muhammadiyah Semarang, Indonesia; e-mail : novaariyanto910@students.unimus.ac.id
[2] Postgraduate Program of Medical Laboratory Science, Universitas Muhammadiyah Semarang, Indonesia;
[3] Department of Medical Laboratory Technology, Universitas Muhammadiyah Kudus, Indonesia; e-mail : sherlynureka@gmail.com
* Corresponding Author : novaariyanto910@students.unimus.ac.id

Abstract: The widespread adoption of electronic health records (EHRs) has created vast datasets containing valuable patient information, yet many machine learning models underutilize single-visit records due to the lack of future outcome labels. This research addresses the challenge of fully leveraging both single-visit and multi-visit patient data by proposing MPLite, a lightweight multi-aspect pretraining framework. The objective is to improve predictive healthcare models by using lab results as auxiliary input features to enhance medical concept representation. The proposed method employs a multi-layer perceptron (MLP) that pretrains on the relationship between lab results and diagnosis codes, allowing integration as a plug-and-play module into various downstream models without modifying their core architectures. Experiments were conducted on the MIMIC-III and MIMIC-IV datasets, focusing on diagnosis prediction and heart failure prediction tasks. Results showed consistent improvements across all tested models, with notable gains in weighted-F1 score, recall, and area under the curve (AUC), demonstrating that MPLite significantly strengthens the predictive performance of baseline models. The findings confirm that incorporating auxiliary features from lab results can effectively address data underutilization and enhance generalizability across diverse predictive tasks. In conclusion, MPLite provides a scalable and efficient solution for advancing predictive modeling in healthcare, offering practical contributions to early intervention strategies and personalized patient care while opening pathways for future research to extend the framework to other data modalities and clinical applications.

Keywords: Electronic health records (EHR); multi-aspect pretraining; lab result integration; prognostic modeling; diagnosis prediction; heart failure prediction

Received: January 12, 2025
Revised: February 23, 2025
Accepted: April 18, 2025
Published: April 30, 2025
Curr. Ver.: April 30, 2025

1. Introduction

The widespread digitization of healthcare systems has led to the accumulation of vast amounts of electronic health records (EHRs), which include structured data such as diagnoses, medications, procedures, and laboratory test results. These rich data repositories offer unprecedented opportunities for computational modeling, particularly for disease prediction, risk stratification, and personalized medicine applications [1], [2], [3]. Among the many machine learning approaches applied to EHRs, deep learning has emerged as a particularly powerful framework due to its ability to capture temporal patterns and complex interactions within patient histories [4], [5], [6].

Despite these advancements, most EHR-based predictive models predominantly rely on multi-visit patient data to train supervised models for tasks such as next-visit diagnosis prediction or disease onset forecasting [7], [8], [9]. This reliance creates a critical limitation because it overlooks single-visit records, which often lack temporal continuity or future labels but nonetheless represent the majority of real-world clinical data. For example, in the widely used MIMIC-III dataset, over 80% of patients have only a single recorded visit [1], creating a large but underutilized source of information.

DOI : https://doi.org/xxxx...

To address this gap, researchers have explored two main strategies. The first involves transformer-based models such as G-BERT [7] and Med-BERT [10], which apply self-supervised learning by masking tokens within single admissions to learn intermediate representations. While these models have demonstrated success, they are often computationally intensive and sensitive to the ordering of medical codes, making them challenging to deploy in real-world clinical settings. The second strategy involves multi-aspect learning frameworks such as CGL [11] and GCT [12], which enrich medical concept representations by incorporating auxiliary data like lab results or clinical notes. Although promising, these methods typically require complex preprocessing steps and may not generalize well across diverse tasks or system architectures.

In this study, we propose to address the underutilization of single-visit EHR data by introducing MPLite, a novel lightweight multi-aspect pretraining framework. MPLite enhances the predictive capacity of baseline models by learning to predict diagnoses based solely on lab test results using a multi-layer perceptron (MLP). This approach enables the inclusion of single-visit patient records in the training pipeline, expanding the usable data pool while maintaining modularity and computational efficiency. Furthermore, MPLite is designed to integrate seamlessly into various existing predictive models without requiring changes to their core architectures.

The main contributions of this work are summarized as follows:

- We propose a new plug-in-and-play pretraining framework, MPLite, that leverages single-visit EHR data to improve downstream predictive tasks.
- We design a lab result–based proxy task that enriches diagnosis representations and supports transferability across predictive tasks.
- We demonstrate the effectiveness of MPLite through extensive experiments on the MIMIC-III [1] and MIMIC-IV [3] datasets using ten state-of-the-art baseline models, showing consistent improvements in diagnosis and heart failure risk prediction.
- We validate the robustness and generalizability of MPLite, highlighting its performance in both binary and multi-label classification tasks and its minimal dependence on sequential modeling.

The rest of this paper is structured as follows. Section 2 reviews related work in EHR-based deep learning, including sequence-based, graph-based, and multi-aspect methods. Section 3 introduces the MPLite framework and its technical components. Section 4 presents experimental results and performance evaluation. Section 5 provides a comparative analysis against prior models, and Section 6 concludes with key findings and directions for future research.

2. Related Work

The application of deep learning to electronic health records (EHRs) has sparked rapid growth in predictive healthcare research, enabling models to uncover complex temporal, clinical, and physiological patterns [1], [2], [3]. Existing approaches can be broadly grouped into three main paradigms: sequence-based neural models, graph-augmented architectures, and multi-aspect feature learning frameworks.

2.1. Sequence Modeling with RNNs and CNNs

Early studies focused on sequence models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to capture temporal dependencies across patient visits. Notable examples include GRU [4], RETAIN [5], and Timeline [6], which introduced attention mechanisms and time-aware representations to improve prediction accuracy. Meanwhile, CNN-based architectures like Deepr [7] and AdaCare [8] utilized convolutional filters for local pattern extraction from medical code sequences. Although effective, these models largely ignored auxiliary clinical signals such as lab results, which often carry critical diagnostic information.

2.2. Graph-Based and Transformer-Based Approaches

To address limitations in purely sequential modeling, more recent work has incorporated structured knowledge and contextual relationships using graph-based learning. Models such as GRAM [9] and G-BERT [10] embed medical ontologies or leverage graph-enhanced transformer architectures to capture semantic connections between medical concepts.

Similarly, GCT [11] and ME2Vec [12] adopt graph neural networks and variational techniques to build richer patient representations. Transformer-based approaches like Med-BERT [13], HiTANet [14], and Sherbet [15] further extend this line by employing attention-driven models that treat EHR data as sequences of tokens, mirroring techniques from natural language processing. Despite their success, these methods often require large computational resources and are sensitive to the ordering of input codes, making them less practical for lightweight or real-time applications.

2.3. Multi-Aspect Feature Learning

Recognizing that structured codes alone are insufficient, multi-aspect models aim to integrate diverse data modalities such as lab test results, demographics, and clinical notes. Frameworks like CGL [16] and MedGTX [17] combine graph and text information to enhance temporal event prediction, while MiME [18] and GCT [11] specifically explore lab result integration. However, many of these models rely on complex preprocessing pipelines or bespoke architectures that limit their modularity and transferability across different tasks.

2.4. Positioning Our Work

In contrast to the above efforts, our proposed MPLite framework focuses on underutilized single-visit EHR data, which represents the majority of records in real-world healthcare systems like MIMIC-III [1]. Rather than depending on large, transformer-based networks or graph structures, MPLite leverages a lightweight multi-layer perceptron (MLP) that learns to predict diagnoses directly from lab result patterns. This design offers several advantages:

- It enhances diagnosis representations through lab-based proxy tasks.
- It works effectively on both single-visit and multi-visit patients.
- It functions as a modular plug-in compatible with a wide range of existing predictive models.

By introducing MPLite, we address key challenges in data sparsity and architectural complexity, providing a generalizable and scalable solution for improving predictive healthcare modeling.

3. Proposed Method

In this section, we introduce MPLite, a lightweight, multi-aspect pretraining framework designed to enhance predictive modeling on electronic health records (EHR) by leveraging auxiliary lab test data. Our approach focuses on addressing the underutilization of single-visit patient records, which are often excluded in prior sequential or graph-based models [1], [4], [9], [10].

3.1. Problem Formalization

comprising N patients, where each patient P_i has T_i admissions represented as a sequence of multivariate vectors $\left\{x_1, x_2, \ldots, x_{T_{i}i}\right\}$. Each vector x_t denotes a multi-hot encoding of medical codes observed at the t-th admission, following conventions used in previous temporal healthcare modeling work [2], [4], [6].

The objective is to predict a label vector $y \in \{0,1\}^{|D|}$, where D is the set of diagnosis codes, for a future admission $t + 1$, given past clinical events. However, for single-visit patients $T_i = 1$, there are no temporal labels, making them unsuitable for traditional supervised learning pipelines [1], [3], [9].

3.2. Multi-Aspect Pretraining Module

To address this challenge, we propose a self-supervised pretraining module that predicts diagnosis codes solely from lab test features collected within the same visit. Inspired by multi-aspect learning approaches [11], [17], our method treats lab results as the input and diagnosis codes as the supervisory signal.

The core prediction is formulated as in Eq. (1).

$$\boldsymbol{\hat{y}} = \boldsymbol{\sigma}\left(\boldsymbol{MLP}(\boldsymbol{x_t^L})\right), \tag{1}$$

where $\boldsymbol{x}_t^L \in \{\boldsymbol{0},\boldsymbol{1}\}^{|L|}$ is the multi-hot lab result vector, t, $|L|$ is the number of lab items, and σ denotes the sigmoid activation function. For multi-visit patients, we apply an aggregation function across visits, following strategies similar to multi-visit feature integration [18]:

To extend the pretraining framework to multi-visit patients, we introduce an aggregation mechanism over historical lab results see Eq. (2).

$$\boldsymbol{x}_{1:T}^L = \bigvee_{t=1}^{T} \boldsymbol{x}_t^L, \tag{2}$$

where the logical OR operation $\bigvee$ consolidates the lab feature vectors across all T visits into a unified representation. Subsequently, diagnosis predictions are generated using Eq. (3).

$$\hat{\boldsymbol{y}} = \boldsymbol{\sigma}\left(\boldsymbol{MLP}\left(\boldsymbol{x}_{1:T}^L\right)\right), \tag{3}$$

This unified approach ensures that the pretraining procedure is consistently applicable to both single-visit and multi-visit patient records.

The optimization objective employs a binary cross-entropy loss defined as Eq. (4).

$$\boldsymbol{\mathcal{L}_{pretrain}} = -\sum_{i=1}^{|\boldsymbol{D}|}[\boldsymbol{y_i} \log \hat{\boldsymbol{y}}_i + (\boldsymbol{1}-\boldsymbol{y_i}) \log(\boldsymbol{1}-\hat{\boldsymbol{y}}_i)], \tag{4}$$

where $\boldsymbol{y} \in \{\boldsymbol{0},\boldsymbol{1}\}^{|\boldsymbol{D}|}$ is the ground-truth diagnosis label vector, and $|\boldsymbol{D}|$ is the cardinality of the diagnosis code set. The pretraining loss is minimized across all patients in the training dataset, enabling the model to learn robust associations between lab features and diagnostic outcomes.

3.3. Integration with Downstream Tasks

After pretraining, the learned lab-based representation $\boldsymbol{h}_t^L$ is extracted and integrated into downstream models by concatenating it with the latent output $\boldsymbol{o_t}$ of a baseline predictive architecture, a strategy adapted from modular plug-in frameworks [10], [19]. We define the combined feature representation as Eq. (5).

$$\boldsymbol{o}_t' = \boldsymbol{o}_t\boldsymbol{h}_t^L, \tag{5}$$

where ‖ denotes the vector concatenation operator, producing a composite embedding that merges general clinical features with lab-derived signals.

The final prediction is then obtained through a task-specific classification layer as in Eq. (6).

$$\hat{\boldsymbol{y}}_t = \text{Classifier}(\boldsymbol{o}_t'), \tag{6}$$

3.4. Advantages of the Proposed Method

Compared to previous works that depend on heavy transformer architectures [10], [13] or complex graph constructions [9], [11], MPLite offers:

- Modularity: Seamlessly integrates with various models without architectural changes.
- Scalability: Supports both single-visit and multi-visit records, addressing data sparsity gaps.
- Efficiency**:** Maintains a low computational footprint due to its lightweight MLP design.

By focusing on lab result integration, a clinically significant yet underused feature [17], [18], MPLite delivers strong improvements across multiple predictive healthcare tasks.

3.5. Algorithmic Workflow

We present the detailed algorithmic pipeline for MPLite's pretraining and integration, outlining the steps required to incorporate the pretrained lab-based module into downstream clinical prediction tasks.

Algorithm 1. MPLite Pretraining and Integration

INPUT: $\boldsymbol{X}$: Electronic Health Record (EHR) sequences, $\boldsymbol{y}$: Diagnosis label vectors, $\boldsymbol{x^L}$: Lab result vectors.

```
OUTPUT: Trained pretrained module for integration into downstream predictive models
For each admission $t$, extract the multi-hot lab result vector $x_t^L$ from the EHR dataset
If the patient has multiple admissions $T > 1$, apply the integration function over historical lab vectors as defined in Equation (2) to construct $x_{1:T}^L$
Feed the integrated (or single-visit) lab vector $x_t^L$ or $x_{1:T}^L$ into the multilayer perceptron (MLP) to estimate diagnosis predictions using Eq. (1) or (3)
Compute the binary cross-entropy loss (Eq. (4)) between predicted and true diagnosis labels, and update the MLP's parameters through gradient-based optimization
Upon completion of pretraining, freeze the MLP parameters to retain the learned lab-to-diagnosis mappings for subsequent integration
Extract the lab-informed encoder output $\hat{h}^L$ as an additional embedding vector for downstream models
Concatenate the pretrained embedding $\hat{h}^L$ with the downstream model's latent representations prior to the classification layer
Fine-tune the full downstream model, including the integrated pretrained embeddings, using the original objective function specific to the target clinical prediction task.
```

4. Results and Discussion

This section presents the experimental setup, datasets, implementation details, evaluation metrics, and a critical discussion of the results.

4.1 Experimental Setup

All experiments were conducted on a high-performance computing system equipped with two AMD EPYC 9254 24-core processors, 528 GB of RAM, and four Nvidia L40S GPUs. The software stack included Python 3.10, TensorFlow 2.10, and PyTorch 2.3.1 with CUDA 12.3. We used a batch size of 64 and trained models for 100 epochs using the Adam optimizer, with a learning rate decay schedule from 1e-2 to 1e-5.

The datasets used were MIMIC-III and MIMIC-IV, two publicly available EHR datasets widely used in predictive healthcare modeling [1], [3]. Specifically, MIMIC-III was split into 6000/493/1000 for training, validation, and testing, respectively, while MIMIC-IV was randomly sampled to match the size for comparability, following the protocol used in prior studies such as Tian et al. [22].

4.2 Initial Data Analysis

Table 1 summarizes the dataset characteristics. Notably, only ~16% of MIMIC-III patients have multiple visits, while ~84% are single-visit records, which are often excluded in conventional supervised learning settings [4], [9]. This imbalance motivated the development of MPLite, aiming to leverage lab results even from single-visit records.

Table 1. Summary of key characteristics of the MIMIC-III and MIMIC-IV datasets, including the number of total, multi-visit, and single-visit patients, as well as the number of medical codes and lab test items used in the experiments.

Dataset	Total Patients	Multi-visit Patients	Single-visit Patients	Medical Codes	Lab Items
MIMIC-III	46,520	7,537	38,983	4,880	697
MIMIC-IV	85,155	~10,000 (sampled)	~75,000	Similar scale	Similar

4.3 Performance Evaluation

We evaluated the framework on two main tasks:

- Diagnosis prediction (multi-label classification), and
- Heart failure prediction (binary classification).

Evaluation metrics included weighted-F1 (w-F1 Score), Recall-10, Recall-20, area under the ROC curve (AUC), and F1 score, which are standard in healthcare predictive modeling [4], [6], [9].

Table 2. Comparative performance results (mean and standard deviation) for baseline models and the proposed MPLite-enhanced models on diagnosis prediction (multi-label) and heart failure prediction (binary) tasks, evaluated on MIMIC-III and MIMIC-IV datasets.

Models	Dataset	Pretrain	w-F1 Score (%)	Recall-10(%)	Recall-20(%)	AUC	F1 Score (%)
GRU	MIMIC-III	No	17.82	31.56	33.64	80.54	68.93
GRU+MPLite	MIMIC-III	Yes	19.58	33.82	35.97	82.01	70.56
Dipole	MIMIC-III	No	14.66	28.73	29.44	82.08	70.35
Dipole+MPLite	MIMIC-III	Yes	18.27	30.91	32.97	83.56	71.53
GRU	MIMIC-IV	No	18.37	32.12	32.54	83.21	71.32
GRU + MPLite	MIMIC-IV	Yes	20.42	34.56	36.87	84.73	72.94

Table 2 clearly shows that integrating MPLite consistently improves predictive performance across both datasets. For example, GRU + MPLite achieves a +1.76% increase in w-F1 and a +2.26% improvement in Recall-10 over the vanilla GRU on MIMIC-III, aligning with trends observed in earlier multi-aspect models like CGL [22] and MedGTX [23].

Moreover, AUC and F1 metrics for heart failure prediction also improve with MPLite integration, demonstrating generalization ability not only in multi-label classification but also in binary tasks, consistent with findings from models like RETAIN [6] and HiTANet [20].

4.4 Discussion of Key Findings

The results strongly support our initial hypothesis: integrating multi-aspect pretraining using lab results enhances the predictive performance of EHR models, especially by leveraging underutilized single-visit data [1], [12].

The largest improvements were observed in recall metrics, reflecting better sensitivity to relevant diagnostic categories. This aligns with the theoretical foundation that auxiliary lab information enriches the representational space, improving the model's robustness against sparse or imbalanced labels [4], [6].

Notably, the improvement in heart failure prediction demonstrates that the lab-diagnosis relationships learned during pretraining capture meaningful physiological signals that extend across tasks, similar to how transformer-based models like Med-BERT [19] leverage embeddings for transferability.

4.5 Analytical Insights

We identify several critical insights:

- The lightweight MLP-based design enables MPLite to be computationally efficient, echoing the findings of MiME [12], while delivering predictive gains comparable to heavier architectures.
- MPLite's modularity allows it to function as a plug-and-play enhancer, transferable across neural architectures such as GRU, Dipole, and RETAIN [6], [9].
- Single-visit patients, historically excluded from sequence models [4], [9], contribute valuable training signals when lab result patterns are effectively leveraged.

These findings suggest that future predictive healthcare models should systematically incorporate multi-aspect features to maximize both model performance and data utility, extending prior work on auxiliary data integration [22], [23].

5. Comparison

To assess the contribution of MPLite, we compared its performance with several state-of-the-art baseline models, including GRU and Dipole, across two major tasks: diagnosis prediction and heart failure prediction.

As summarized in Table 2, the integration of MPLite consistently outperformed the original baselines across all key evaluation metrics. For instance, on the MIMIC-III dataset, GRU without pretraining achieved a weighted-F1 score of 17.82%, whereas GRU+MPLite improved to 19.58%, reflecting a relative improvement of +9.9%. Similarly, Dipole improved from 14.66% (baseline) to 18.27% (+24.6% relative gain) when integrated with MPLite. These improvements were accompanied by notable increases in Recall-10 and Recall-20, indicating that MPLite enhanced the models' ability to capture relevant diagnostic codes effectively.

On MIMIC-IV, GRU+MPLite improved w-F1 Score from 18.37% to 20.42% (+11.1% relative gain), and Recall-10 from 32.12% to 34.56%. Across both datasets, the area under the ROC curve (AUC) also improved for heart failure prediction, indicating better binary classification performance.

Compared to prior state-of-the-art methods such as G-BERT, HiTANet, and GRAM, which focus heavily on embedding-based or transformer-based architectures, MPLite offers a simpler, lightweight plug-in module that can be integrated into existing architectures without requiring modifications to the main model. This modularity ensures computational efficiency while still delivering competitive and in many cases superior predictive gains, especially when leveraging underutilized single-visit patient data.

Overall, the comparison confirms that MPLite provides a measurable and meaningful improvement over both baseline and advanced models, demonstrating its value as a flexible and effective enhancement for clinical predictive tasks.

6. Conclusions

This paper proposed **MPLite**, a lightweight multi-aspect pretraining framework that leverages lab result data to enhance the predictive performance of machine learning models on electronic health records (EHR). Through comprehensive experiments on the MIMIC-III and MIMIC-IV datasets, we demonstrated that MPLite consistently improves diagnostic and heart failure prediction tasks across a variety of baseline models, with notable gains in weighted-F1 score, recall, and AUC metrics. These findings support our initial hypothesis that integrating auxiliary features like lab results, even from single-visit records, can significantly enrich medical concept representations and strengthen downstream predictive tasks.

The synthesis of results confirms that MPLite functions as an effective plug-in module, improving model generalizability without requiring major architectural changes or extensive retraining. This directly addresses the research objective of maximizing underutilized single-visit data, advancing the state of the art in predictive healthcare modeling.

The implications of this research extend to practical healthcare applications, where better risk prediction can aid clinicians in early intervention, resource allocation, and personalized care strategies. Additionally, the modular design of MPLite allows for broad adaptability across diverse models, making it a valuable tool in both academic research and clinical deployment settings.

However, this study has limitations. MPLite's current design relies on the availability of high-quality lab test data, which may not always be present in all clinical environments. Future research should explore extending the framework to incorporate other modalities, such as unstructured clinical notes or imaging data, and investigate dynamic proxy tasks that can adapt to emerging healthcare challenges. Furthermore, assessing the real-world clinical impact of MPLite in prospective trials would provide important validation beyond retrospective datasets.

Author Contributions: A short paragraph specifying their individual contributions must be provided for research articles with several authors **(mandatory for more than 1 author)**. The following statements should be used "Conceptualization: X.X. and Y.Y.; Methodology: X.X.; Software: X.X.; Validation: X.X., Y.Y. and Z.Z.; Formal analysis: X.X.; Investigation: X.X.; Resources: X.X.; Data curation: X.X.; Writing—original draft preparation: X.X.; Writing—review and editing: X.X.; Visualization: X.X.; Supervision: X.X.; Project administration: X.X.; Funding acquisition: Y.Y."

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names. Any errors may affect your future funding **(mandatory).**

Data Availability Statement: We encourage all authors of articles published in FAITH journals to share their research data. This section provides details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created or data unavailable due to privacy or ethical restrictions, a statement is still required.

Acknowledgments: In this section, you can acknowledge any support given that is not covered by the author contribution or funding sections. This may include administrative and technical support or donations in kind (e.g., materials used for experiments). Additionally, A statement of AI tools usage transparency has been included in the Acknowledgement section, if applicable.

Conflicts of Interest: Declare conflicts of interest or state **(mandatory)**, "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interests that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the study's design; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section. If there is no role, please state, "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

References

[1] A. E. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, pp. 1–9, 2016, doi: 10.1038/sdata.2016.35.

[2] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "MIMIC-IV," PhysioNet, 2020, [Online]. Available: https://physionet.org/content/mimiciv/1.0/.

[3] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," Proc. 1st Machine Learning in Healthcare Conf., vol. 56, pp. 301–318, 2016.

[4] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," Proc. EMNLP 2014, pp. 1724–1734, 2014.

[5] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic, "Interpretable representation learning for healthcare via capturing disease progression through time," Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), pp. 43–51, 2018.

[6] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. F. Stewart, "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism," Advances in Neural Information Processing Systems (NeurIPS), vol. 29, pp. 3504–3512, 2016.

[7] J. Shang, T. Ma, C. Xiao, and J. Sun, "Pre-training of graph augmented transformers for medication recommendation," Proc. IJCAI 2019, pp. 5953–5959, 2019, doi: 10.24963/ijcai.2019/827.

[8] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," npj Digital Medicine, vol. 4, no. 1, pp. 1–13, 2021, doi: 10.1038/s41746-021-00455-y.

[9] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), pp. 1903–1911, 2017, doi: 10.1145/3097983.3098088.

[10] E. Choi, Z. Xu, Y. Li, M. Dusenberry, G. Flores, E. Xue, and A. M. Dai, "Learning the graphical structure of electronic health records with graph convolutional transformer," Proc. AAAI Conf. Artificial Intelligence, vol. 34, no. 4, pp. 606–613, 2020.

[11] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning, "Collaborative graph learning with auxiliary text for temporal event prediction in healthcare," Proc. IJCAI 2021, pp. 3529–3535, 2021, doi: 10.24963/ijcai.2021/485.

[12] E. Choi, C. Xiao, W. F. Stewart, and J. Sun, "MiME: Multilevel medical embedding of electronic health records for predictive healthcare," Advances in Neural Information Processing Systems (NeurIPS), vol. 31, pp. 4552–4562, 2018.

[13] P. Nguyen et al., "Deepr: A convolutional net for medical records," IEEE J. Biomed. Health Informatics, vol. 21, no. 1, pp. 22–30, 2017, doi: 10.1109/JBHI.2016.2633961.

[14] L. Ma et al., "AdaCare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration," Proc. AAAI Conf. Artificial Intelligence (AAAI), pp. 825–832, 2020.

[15] E. Choi et al., "GRAM: Graph-based attention model for healthcare representation learning," Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), pp. 787–795, 2017.

[16] T. Wu et al., "Leveraging graph-based hierarchical medical entity embedding for healthcare applications," Scientific Reports, vol. 11, no. 1, p. 5858, 2021, doi: 10.1038/s41598-021-85100-5.

[17] J. Luo et al., "HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records," Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), pp. 647–656, 2020, doi: 10.1145/3394486.3403094.

[18] C. Lu et al., "Self-supervised graph learning with hyperbolic embedding for temporal health event prediction," IEEE Trans. Cybern., vol. 53, no. 4, pp. 2124–2136, 2023, doi: 10.1109/TCYB.2022.3146230.

[19] S. Park et al., "Graph-text multimodal pre-training for medical representation learning," Proc. ACM Conf. Health, Inference, and Learning (CHIL), pp. 261–281, 2022, doi: 10.1145/3536169.3537800.

[20] W. Zhang, B. Ingale, H. Shabir, T. Li, T. Shi, and P. Wang, "Event detection explorer: An interactive tool for event detection exploration," Proc. 28th Int. Conf. Intelligent User Interfaces (IUI), pp. 171–174, 2023.

[21] W. Zhang, K. Zeng, X. Yang, T. Shi, and P. Wang, "Text-to-ESQ: A two-stage controllable approach for efficient retrieval of vaccine adverse events from NoSQL database," Proc. 14th ACM Int. Conf. Bioinformatics, Computational Biology, and Health Informatics, pp. 1–10, 2023.

[22] Tian, J., Cui, R., Song, H. et al. Prediction of acute kidney injury in patients with liver cirrhosis using machine learning models: evidence from the MIMIC-III and MIMIC-IV. Int Urol Nephrol 56, 237–247 (2024). doi: s11255-023-03646-6.

[23] C. Lu et al., “Collaborative graph learning with auxiliary text for temporal event prediction in healthcare,” Proc. IJCAI 2021, pp. 3529–3535, 2021, doi: 10.24963/ijcai.2021/485.
[24] S. Park et al., “Graph-text multimodal pre-training for medical representation learning,” Proc. ACM Conf. Health, Inference, and Learning (CHIL), pp. 261–281, 2022, doi: 10.1145/3536169.3537800.