*MEDINFO 2023 — The Future Is Accessible*
*J. Bichel-Findlay et al. (Eds.)*


*doi:10.3233/SHTI231060*

# Graph Neural Network Based Multi-Label Hierarchical Classification for Disease Predictions in General Practice

Shengqiang Chi[a], Yuqing Wang[a], Ying Zhang[a], Weiwei Zhu[a], and Jingsong Li[a,b,1]
[a] *Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China*
[b] *Engineering Research Center of EMR and Intelligent Expert Systems, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*

**Abstract.** General practitioners are supposed to be better diagnostics to detect patients with serious diseases earlier, and conduct early interventions and appropriate referrals of patients. However, in the current general practice, primary general practitioners lack sufficient clinical experiences, and the correct rate of general disease diagnosis is low. To assist general practitioners in diagnosis, this paper proposes a multi-label hierarchical classification method based on graph neural network, which integrates medical knowledge and electronic health record (EHR) data to build a disease prediction model. The experimental results based on data consist of 231,783 visits from EHR show that the proposed model outperforms all baseline models in the general disease prediction task with a top-3 recall of 0.865. The interpretable results of the model can effectively help clinicians understand the basis of the model's decision-making.

**Keywords.** General practice, diagnosis prediction, graph neural network, multi-label classification, deep learning

## 1. Introduction

The public and experts hope that general practitioners should detect patients with serious diseases earlier, and intervene as early as possible. However, the current status of general practice (GP) care is far from ideal. In the diagnosis of common diseases (dysentery and angina pectoris) among rural doctors in western China, the completion rate of necessary consultations and examinations was 36%, and the correct rate of diagnosis was 26%[1]. Diagnosis in GP is difficult. The prevalence of severe diseases was low among patients treated by general practitioners. The disease is usually at an early stage, lacking specific symptoms. In addition, primary general practitioners lack sufficient clinical experiences, which may affect decision-making in diagnosis.

In recent years, there have been more and more studies on disease prediction models using electronic health record (EHR) data. GNDP[2] is a spatio-temporal graph neural network (GNN) based disease prediction model, which learns robust representations for patients and simultaneously make disease predictions. Sun[3] et al introduced a GNN-based disease prediction model that utilized external knowledge

[1] Corresponding Author: Jingsong Li, e-mail: ljs@zju.edu.cn.

bases to augment EHR data to learn highly representative embeddings for patients, thereby enabling diseases to be accurately predicted. These methods exploit the knowledge in the feature space and ignore the knowledge in the label space.

In order to solve the difficulties in disease predictions in GP, this study introduces medical knowledge to construct a disease prediction model. The main contributions of this study are as follows: First, introduce knowledge information into the data feature space for node representations by constructing a patient graph using knowledge graph (KG) and EHR data. Then, a GNN-based multi-label hierarchical classification model (MLH-GNN) is proposed to improve prediction performance, which introduces knowledge into the data label space. Finally, extensive experiments on real-world EHR data show that the model outperforms baseline models on general disease predictions.

## 2. Methods

### *2.1. Datasets*

This study was based on the data of Hangzhou Second People's Hospital from 2017 to 2020. The general diseases concerned in this study were determined by clinical experts. We excluded data on patient visits that there were no text data available. We extracted the patient symptoms from the texts through methods such as named entity recognition and entity relationship recognition, for subsequent model constructions and evaluations.

### *2.2. Model construction and evaluation*

Given a disease set $V_D$, a symptom set $V_S$, and a patient set $V_P$, we constructed a patient graph G that includes three node types using medical knowledge and EHR data. The patient graph can be expressed as: $\mathrm{G} = (\mathrm{V}, \mathrm{E})$, where $\mathrm{V} = \{v_i | v_i \in \{V_D \cup V_S \cup V_P\}\}$, $\mathrm{E} = \{(v_i, \mathrm{r}, v_j) | \mathrm{r} \in R, v_i, v_j \in \mathrm{V}\}$, $R$ is the relationship set, which contains the disease-symptom and patient-symptom relationships. Disease-symptom relationships come from a KG constructed by clinical experts. Patient-symptom relationships were obtained from EHR data. The structure of the patient graph is shown in Figure 1.A.

Based on disease-symptom triplets in the KG, the initial embeddings of disease and symptom nodes are obtained using transE[4]. The initial embedding of a patient node is obtained by averaging of the initial embeddings of all its symptom nodes.

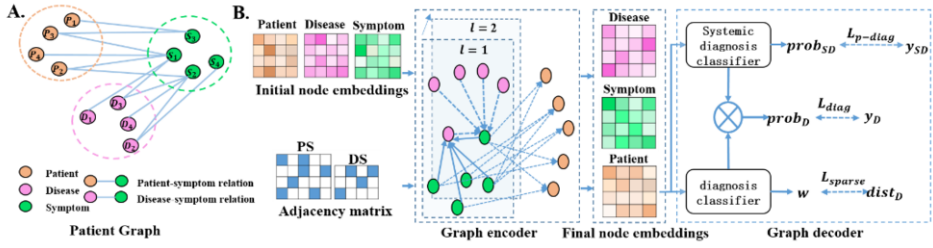


**Figure 1.** The structure of the patient graph and the proposed model.

Figure 1.B shows the overall structure of the proposed model. A graph attention-based encoder was used to generate embeddings for nodes. First, $L$ layers of graph attentions were performed on the patient graph. For node $i \in V$, the attention coefficient $\alpha_{ij}^{l}$ between its neighbor node $j \in N(i)$ and itself in the layer $l$ is calculated:

$$\alpha_{ij}^{l} = \frac{\exp(LeakyReLU(a([Wh_i^l \| Wh_j^l])))}{\sum_{k\in N(i)}\exp(LeakyReLU(a([Wh_i^l \| Wh_k^l])))} \quad (1)$$

where $W$ is a weight matrix of a shared linear transformation applied to every node, $a$ is a shared attentional mechanism. $[\cdot\|\cdot]$ is an operator for vector concatenations, $h_i^l$ is the embedding of node $i$ in the layer $l$, which is calculated as:

$$h_i^l = \sigma(\sum_{j\in N(i)} \alpha_{ij}^l W h_j^{l-1}) \quad (2)$$

where $\sigma(\cdot)$ is the activation function. After passing through the $L$ layer graph attention network, the final node embeddings are obtained. Secondly, each graph attention layer is followed by a normalization layer, a dropout layer and an activation layer.

Figure 2 shows a simple example of the disease hierarchy. Diseases in the $L_D$ layer are those needed to be predicted. Diseases in the $L_{SD}$ layer are system classifications of diseases based on KG, which noted as $SD_i, i = 1,2,\ldots,N^{L_{SD}}$. $N^{L_{SD}}$ is the number of disease system classifications in the $L_{SD}$ layer.

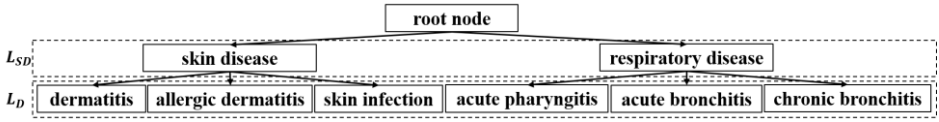


**Figure 2.** A simple example of disease hierarchy.

Next, we constructed $N^{clf}$ binary classifiers, where $N^{clf} = N^{L_{SD}} + N^D$. $N^{clf}$ predicted probabilities are obtained for patient $P_i$ by inputting the node embeddings into classifiers. The probability of the patient $P_i$ with the disease $D_j$ is calculated as:

$$p_i^{D_j} = prob_{SD_c} \times prob_{D_j} \quad (3)$$

where $prob_{SD_c}$ is the probability of $P_i$ has $SD_c$ predicted by the classifier $clf_c$, $SD_c$ is the disease system classification of $D_j$, $prob_{D_j}$ is the probability of $P_i$ has $D_j$ predicted by the classifier $clf_j$. The labels corresponding to classifiers are disease diagnosis and their system classifications. The loss function of the proposed model is defined as:

$$L = L_{diag} + L_{p-diag} \quad (4)$$

$$L_{diag} = -\frac{1}{N^P}\frac{1}{N^D}\sum_{i=1}^{N^P}\sum_{j=1}^{N^D}(y_i^{D_j}\log(p_i^{D_j}) + (1-y_i^{D_j})\log(1-p_i^{D_j})) \quad (5)$$

$$L_{p-diag} = -\frac{1}{N^P}\frac{1}{N^{L_{SD}}}\sum_{i=1}^{N^P}\sum_{c=1}^{N^{L_{SD}}}(y_i^{SD_c}\log(prob_{SD_c})) + (1-y_i^{SD_c})\log(1-prob_{SD_c})) \quad (6)$$

where $N^D$ is the number of diseases, $N^P$ is the number of patients, $y_i^{D_j}$ is the ground truth of $P_i$ with $D_j$, $y_i^{SD_c}$ is the ground truth of disease system classifications.

The performance of MLH-GNN was compared with logistic regression (LR), random forest (RF), GCN[5], GraphSAGE[6] and GAT[7]. The following is a brief description of the training: (1) LR and RF: The symptoms in each visit were one-hot encoded, and finally 1,469 features were obtained. (2) GCN, GraphSAGE and GAT: These GNN-based models took the patient graph as input. Multi-label classifiers were trained based on the Binary Relevance strategy[8]. We split the data into training, validation, and test sets in a ratio of 6:2:2. The MLH-GNN model is trained by the Adam optimizer, an activation function of ReLU, and a learning rate of 1e-4. The number of graph attention layers is 2. The model performance was evaluated from two aspects of multi-label learning and recommendation. Evaluation metrics widely-used in multi-label learning were employed: hamming loss, one error, coverage, ranking loss, average precision, and macro-averaging AUC. The recommendation performance was evaluated by precision (P), recall (R) and F1 score (F1) of the top-K diseases.

## 3. Results

### *3.1. Data characteristics*

A total of 70 general diseases and 182,384 patients whose 231,783 visits contained at least one general diseases were identified. On average, there were 1.10 general disease diagnoses per visit. A total of 1,469 symptoms were extracted from the EHR data, with an average of 2.98 symptoms in each visit. The number of visits varied widely by disease, with a median of 3105.5 visits per disease (range: 450.25-24369.1).

### *3.2. Model performance and interpretation*

The multi-label learning and recommendation performance of 6 models are shown in Tables 1 and 2. The results show that GNN-based models can outperform LR and RF. The MLH-GCN model outperforms all benchmark models in general disease predictions.

**Table 1.** The multi-label learning performance of 6 models. ↓ indicates "the smaller the better," whereas ↑ indicates "the larger the better.".

| Metrics | Hamming loss ↓ | One error ↓ | Coverage ↓ | Ranking loss ↓ | Average precision ↑ | macro-averaging AUC ↑ |
|---|---|---|---|---|---|---|
| LR | 0.012 | 0.528 | 3.056 | 0.025 | 0.775 | 0.958 |
| RF | 0.010 | 0.432 | 6.104 | 0.062 | 0.780 | 0.914 |
| GCN | 0.010 | 0.437 | 2.814 | 0.022 | 0.798 | 0.963 |
| GraphSAGE | 0.010 | 0.426 | 2.749 | 0.021 | 0.801 | 0.967 |
| GAT | 0.010 | 0.427 | 2.692 | **0.020** | 0.800 | 0.969 |
| MLH-GNN | **0.010** | **0.417** | **2.632** | **0.020** | **0.802** | **0.970** |

**Table 2.** The recommendation performance of 6 models.

| Metrics | K=1 | | | K=2 | | | K=3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| LR | 0.675 | 0.614 | 0.643 | 0.427 | 0.777 | 0.551 | 0.309 | 0.843 | 0.452 |
| RF | 0.692 | 0.631 | 0.660 | 0.111 | 0.787 | 0.195 | 0.064 | 0.853 | 0.119 |
| GCN | 0.709 | 0.645 | 0.675 | 0.439 | 0.798 | 0.566 | 0.314 | 0.858 | 0.460 |
| GraphSAGE | 0.711 | 0.647 | 0.678 | 0.441 | 0.801 | 0.569 | 0.316 | 0.861 | 0.462 |
| GAT | 0.707 | 0.643 | 0.674 | 0.440 | 0.801 | 0.568 | 0.316 | 0.861 | 0.462 |
| MLH-GNN | **0.714** | **0.649** | **0.680** | **0.445** | **0.810** | **0.574** | **0.317** | **0.865** | **0.464** |

We utilized PGExplainer[9] to explain the model's decision. PGExplainer is able to provide model-level explanations for each instance with a global view of the GNN model. Influenza, acute nasopharyngitis and urticarial with probabilities of 0.783, 0.215 and 0.002 were recommended for the patient diagnosed of "influenza" with symptoms of nausea, fever, runny nose, fatigue, sore throat, nasal congestion, and cough. The supported symptoms for "influenza" are nausea, sore throat, cough, fever, runny nose, nasal congestion and fatigue with corresponding weights of 0.203, 0.166, 0.147, 0.138, 0.130, 0.124 and 0.093, which is thought to be reasonable by clinicians.

## 4. Discussion

In this study, we validated that the proposed model had good performance in general disease predictions with EHR data. GNN models perform better compared with LR and

RF, which may be due to the introduction of medical knowledge. The information of relationships between symptoms and patients can be better integrated into the model, through the graph topology of the disease-symptom-patient relationships. Based on the GNN, the proposed model introduces the hierarchical relationships of diseases into the label space, to further improve the prediction performance.

When predicting a patient's disease, several possible diseases can be recommended based on the patient's symptoms to assist doctors in diagnosis. The interpretability results of the model can help doctors understand the basis of the model's decision-making. Future works will include validating the performance of the model on external datasets and evaluating the model's utility in real clinical scenarios.

## 5. Conclusions

This study proposed a GNN-based multi-label hierarchical classification model for general disease predictions. Based on the GNN, the model effectively combines medical knowledge and clinical data, and makes full use of medical knowledge as the internal information of EHR data to improve the prediction accuracy. Experimental results on real-world EHR datasets show that the proposed model outperforms other baseline models. The interpretable results of the model can help clinicians understand the basis of the model's decision-making.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 82001925), the Zhejiang Provincial Natural Science Foundation of China (No. LQ21H180002), the Key Research Project of Zhejiang Lab (No. 2022ND0AC01).

## References

[1] S. Sylvia, Y. Shi, H. Xue, X. Tian, H. Wang, Q. Liu, A. Medina, and S. Rozelle, Survey using incognito standardized patients shows poor quality care in China's rural clinics, Health policy planning 30 (2015), 322-333, doi: 10.1093/heapol/czu014.

[2] B.Q. Y Li, X Zhang, H Liu, Graph Neural Network-Based Diagnosis Prediction, Big Data 8 (2020), 379-390, doi: 10.1089/big.2020.0070.

[3] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, and F. Yang, Disease Prediction via Graph Neural Networks, Ieee Journal of Biomedical and Health Informatics 25 (2021), 818-826, doi: 10.1109/JBHI.2020.3004143.

[4] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, Translating embeddings for modeling multi-relational data, Advances in Neural Information Processing Systems 26 (2013).

[5] T.N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:.02907 (2016).

[6] W.L. Hamilton, R. Ying, and J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025-1035.

[7] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, Graph attention networks, arXiv preprint arXiv:.10903 (2017).

[8] M.-L. Zhang and Z.-H. Zhou, A review on multi-label learning algorithms, IEEE transactions on knowledge data engineering 26 (2014), 1819-1837, doi: 10.1109/TKDE.2013.39.

[9] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, Parameterized explainer for graph neural network, arXiv preprint arXiv:.04573 (2020).