

# CS598: Deep Learning for Healthcare Final Project

----Paper selected: Context-Aware Health Event Prediction via Transition Functions on Dynamic Disease Graphs <sup>[1]</sup>

Team #164

Team member: Zexi Yan, Stanley Wen, and Bo Zhang

Corresponding emails: zexiyan2@illinois.edu, keyanw2@illinois.edu, zhang375@illinois.edu

## I. General Problem

The primary challenge addressed by the paper is the limitation of existing health event prediction models which consider diagnoses as independent entities, neglecting the clinical relationships among diseases. This oversight hinders the ability to effectively utilize combinational disease information and understand the dynamic nature of disease development over time. This leads to the two problems the paper is trying to address.

1. Disease combinations in medical practice form a global graph structure that reveals hidden patterns among diseases, with individual patient visits represented as local subgraphs. Despite the potential to predict future health events by analyzing these structures, common deep learning models like GRAM [2], Timeline [3], and G-BERT [4] do not utilize this graph structure for health event predictions.
2. The progression of a disease in a patient is dynamic, as evidenced by changing diagnosis priorities and the emergence of new diagnoses in EHR datasets like MIMIC-III [5]. This dynamic nature, where diseases evolve and impact patients differently over time, suggests the need for a model that can dynamically represent disease development and learn the transition from potential to actual diagnoses.

## II. Specific approach

To overcome these challenges, the paper introduces a novel context-aware learning framework named Chet. This framework employs transition functions on dynamic disease graphs, enabling the model to capture both global and local context from patient visits and model disease transition processes.

1. It utilizes disease combination information by creating a weighted global disease combination graph with various node properties from all patients' historical diagnoses. Dynamic subgraphs are then designed for each patient visit to incorporate both the local context of diseases present in that visit and the global context from the entire EHR dataset.
2. Then it defines multiple diagnosis roles for each visit based on the variation of node properties in dynamic subgraphs to represent disease transition processes. This helps to explore the development scheme of diseases.
3. It designs transition functions for each role to extract historical contexts.

4. It finally integrates all visits of a patient and adopts an attention-based method to predict future health events.

### III. Hypotheses to be tested

1. Chet can outperform existing state-of-the-art health event prediction models by effectively utilizing the dynamic and combinational nature of disease information. Specifically, we will compare Chet with the following models:
  - a. CNN-based model: Deepr [6].
  - b. GRAM [2].
2. The introduction of transition functions and dynamic graph learning will provide a more nuanced understanding of disease progression, leading to more accurate health event predictions. This will be tested by the Ablations described below.

### IV. Ablations planned:

1. Evaluate the model's performance without the transition functions on the prediction accuracy to understand their contribution to the model's performance.
2. Evaluate the model's performance without the dynamic graph learning component to understand their contribution to the model's performance.

### V. Description of how you will access the data:

The project will use the MIMIC-III [5] and MIMIC-IV [7] datasets for model validation and testing. These datasets are publicly available and contain comprehensive patient records, including diagnoses and visit information, which are suitable for developing and evaluating the proposed health event prediction model.

### VI. Discussion of the feasibility of the computation:

Given the complexity of the proposed approach, significant computational resources will be required. The initial experiments were conducted using a machine with specifications such as an Intel i9-9900K CPU, 64GB memory, and a Geforce RTX 2080 Ti GPU, as used in the original study. The actual time of computation is not specified in the original paper.

We intend to leverage Google Collab for replicating the study, though the computational resources of Collab may not match those of the local machine utilized in the original research. It may be necessary to decrease the dataset size for our experiments. We will regularly evaluate the computational feasibility, making necessary modifications to our setup to accommodate the resource constraints.

### VII. Statement on the use of existing code:

The paper's code is accessible on GitHub, and while we plan to reference it for a better grasp of the study, we will not use it directly in our work. Instead, our approach involves redeveloping the code using the

PyHealth Package, aiming to enhance its functionality and potentially contribute our improvements back to the PyHealth community.

## References

- [1] Lu, Chang, Tian Han, and Yue Ning. "Context-aware health event prediction via transition functions on dynamic disease graphs." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 4, pp. 4567-4574. 2022.
- [2] Choi, Edward, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. "GRAM: graph-based attention model for healthcare representation learning." In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 787-795. 2017.
- [3] Bai, Tian, Shanshan Zhang, Brian L. Egleston, and Slobodan Vucetic. "Interpretable representation learning for healthcare via capturing disease progression through time." In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 43-51. 2018.
- [4] Shang, Junyuan, Tengfei Ma, Cao Xiao, and Jimeng Sun. "Pre-training of graph augmented transformers for medication recommendation." arXiv preprint arXiv:1906.00346 (2019).
- [5] Johnson, Alistair EW, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. "MIMIC-III, a freely accessible critical care database." Scientific data 3, no. 1 (2016): 1-9.
- [6] Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe and Svetha Venkatesh, " Deepr : A Convolutional Net for Medical Records," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 22-30, Jan. 2017.
- [7] Johnson, Alistair, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. "Mimic-iv." PhysioNet. Available online at: <https://physionet.org/content/mimiciv/1.0/>(accessed August 23, 2021) (2020): 49-55.