Predicting sequences of clinical events is often based on the concept of a low-dimensional latent embedding of the entities and events of interest in combination with neural network models, this approach showed good predictive performance in the respective fields. Similar approaches have been highly successful in statistical models for recommendation systems, language models, and knowledge graphs. Graphs inherently capture relationships between entities, which can be helpful in a field of medicine.

The graph model of time can be thought of as a discrete model of the dynamic system that generates the original time series, and the state space of which can be restored according to Takens theorem.

In [1] transformation of time series into complex network representation is proposed to be distinguished into three classes, that are based on:

1. mutual proximity of different segments of a time series (proximity networks),
2. convexity of successive observations (visibility graphs), and
3. transition probabilities between discrete states (transition networks).

Approaches used in personal medicine, predictions and classification use Electronic Health Records (EHR), where data of the patients, their treatments and diagnosis are stored. The main idea is to convert temporal data of EHR into the graph in order to conduct further research, that can exploit graph structure and peculiarities.

EHRs contain heterogeneous data such as medications, diagnoses, clinical notes, sometimes lab results, etc. Thus, diverse modeling techniques are needed to meet the heterogeneity of EHR data.

Nevertheless, most of the articles, that were found to contain words «medical records» and «graph» didn’t seem to investigate graphs in the sense of graph theory or use graphs to represent individual patient data [2]. But there are a few, that actually exploit graph structure to represent time series data.

In [3] Khademi et al. constructed a probabilistic graphical model using clinical data and applied deep belief network is applied to microarray data. The structure of the clinical model is extracted automatically by applying a structure learning algorithm, so the data is represented by a Bayesian network. Then, these two models are integrated using softmax nodes. The model is used for breast cancer prognosis.

In [4] by Esteban et al. latent embedding models are applied to clinical data which is represented by temporal sequences of high-dimensional sparse events. The EHR are processed and represented as Markov embeddings to be used in MLP network to predict the sequence of events recorded in the electronic medical record of each patient. The results turned out to be 0.574 for AUPRC with the time that it took to train for each model with the best set of hyperparameters in the first random split is 6.11 hours.

Zhang et al. [5, 6] proposed an integrative medical temporal graph-based prediction approach. The temporal profile graph is constructed for each patient EHRs. The nodes in the proposed graph are medical events (i.e., diseases, medications, lab tests, etc.). Directed edges in the graph capture the temporal relationships between the vertices, which represent the correlation between events. The weight in the profile graph reflects the average time interval between two clinical events in sequences, and the thicker edges indicate the shorter intervals. In [6] authors propose a homogeneous CNN. But results are not too promising, precision for the model, that identifies the two diseases of a patient with the highest probabilities, turned out to be 0.335 and recall is about 0.343. Thought it is better, than logistic regression or a standard CNN.

A graph-based, semi-supervised learning algorithm for risk predictions with a large quantity of unlabeled data is proposed by Chen et al. [7]. The task of risk prediction is stated as a multi-class classification problem using the Cause of Death information as labels, regarding the health-related death as the «highest risk». In the graph developed in this paper different nodes represent examination items in different categories and edges are their temporal relationships. The weight of the edge is calculated based on the assumption that the newer a record, the more important it is in terms of risk prediction.

Based on the temporal graph representation Liu et al. [8] developed an approach for temporal phenotyping to identify the most significant and interpretable graph basis. Just like in papers above, the nodes in the graph are events in the EHR and the directed edges encode the temporal relationships between pairwise events.

Methods of graph construction mentioned above do not exploit any inter-patient connections, which can be important in making prognosis as similar patients might have similar disease flow.

Tong et al. [9] propose LSTM-GNN model for patient outcome prediction tasks, combining Long Short-Term Memory networks (LSTMs) for extracting temporal features and Graph Neural Networks (GNNs) for extracting the patient neighborhood information. The patient graph is constructed, where the nodes are patients and the edges express relatedness in diagnoses. This information is used to predict in-hospital mortality and length of ICU stay using data from the first 24 hours of the ICU stay. The patient graph is put to the LSTM-GNN, which produces three types of embeddings: the LSTM, GNN and static embeddings, which are concatenated and passed to a fully-connected layer to obtain per-node prediction.

In [10] graph methods to cluster similar patients were developed. Patients are linked when the same or similar patterns can be observed for these patients, this concept enables the construction of a network-like structure.

There are papers, that describe methods to store medical records in a form of a graph like structure just in order to maintain them as a database, thus there is no description of any analysis of this data or usage in ML models.

Hanzlicek et al. [11] described MUDR EHR, a multimedia distributed health record for decision support. The collected data are stored using a directed graph (directed forest). Each tree in the forest describes the data of one patient. And edges describe relations between medical concepts. Kaur et al. [12] described a model, which combines different data stores of patient data. In this architecture, the user creates his request at the interface and the architecture below translates this request into a query to get the data from the most suitable data store for this request. The resulting graph of this paper helps to get the right information from the data storage.

|  |  |  |  |
| --- | --- | --- | --- |
| **Paper reference** | **Nodes representation** | **Edges representation** | **Usage** |
| [3] | Symptoms and diseases | Probabilistic connections | Breast cancer prognosis |
| [8] | Medical events | Temporal relationships | Phenotyping |
| [5, 6] | Risk predictions |
| [7] | Nodes that belong to examination category and individual records nodes (tree structure) | Every attribute type  node is linked to a record type node  that the observation was originally from |
| [9] | Patients | Relatedness in diagnoses | Patient outcome prediction |
| [10] | Predicting and analyzing procedures |

The longitudinal and heterogeneous properties make EHR analysis an inherently difficult challenge. There are other methods used in other fields of interest. But they do not study heterogeneous nature of medical records data. In [13] it is proposed to construct a transition graph out of a time series, by making a delay matrix known as a Hankel matrix that consists of the vectors that are the phase trajectory points in state space. After that, vertices are constructed, by clustering columns in the matrix with the hierarchical clustering algorithm. The weights in assigned to vertices are the number of points in the cluster. The edges of the graph with definite directions are constructed from the sequence of the points in the system trajectory, where the weight of an edge shows how many times the system moves from one point to the other.

# References

1. Donner, Reik & Small, Michael & Donges, Jonathan & Marwan & Norbert & Zou, Yong & Xiang, Ruoxi & Kurths, Juergen. (2010). Recurrence-based time series analysis by means of complex network methods. International Journal of Bifurcation and Chaos - IJBC. 21. 10.1142/S0218127411029021.
2. Schrodt, Jens & Dudchenko, Aleksei & Knaup, Petra & Ganzinger, Matthias. (2020). Graph-Representation of Patient Data: a Systematic Literature Review. Journal of Medical Systems. 44. 10.1007/s10916-020-1538-4.
3. Khademi, Mahmoud & Nedialkov, Nedialko. (2015). Probabilistic Graphical Models and Deep Belief Networks for Prognosis of Breast Cancer. 727-732. 10.1109/ICMLA.2015.196.
4. Cristóbal Esteban, Danilo Schmidt, Denis Krompaß, and Volker Tresp. Predicting sequences of clinical events by using a personalized temporal latent embedding model. Proceedings - 2015 IEEE International Conference on Healthcare Informatics, ICHI 2015, pages 130–139, 2015.
5. Zhang, Shuai & Liu, Lei & Li, Hui & Xiao, Zongshui & Cui, Lizhen. (2016). MTPGraph: A Data-Driven Approach to Predict Medical Risk Based on Temporal Profile Graph. 1174-1181. 10.1109/TrustCom.2016.0191.
6. Zhang, Jinghe & Gong, Jiaqi & Barnes, Laura. (2017). HCNN: Heterogeneous Convolutional Neural Networks for Comorbid Risk Prediction with Electronic Health Records. 214-221. 10.1109/CHASE.2017.80.
7. L. Chen et al., "Mining Health Examination Records—A Graph-Based Approach," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 9, pp. 2423-2437, 1 Sept. 2016, doi: 10.1109/TKDE.2016.2561278.
8. Liu, C., Wang, F., Hu, J., et al. Temporal Phenotyping from longitudinal electronic health records. In: Cao L, Zhang C, Joachims T et al. (eds) the 21th ACM SIGKDD international conference, pp 705–714, 2015
9. Tong, Catherine & Rocheteau, Emma & Veličković, Petar & Lane, Nicholas & Lio, Pietro. (2022). Predicting Patient Outcomes with Graph Representation Learning. 10.1007/978-3-030-93080-6\_20.
10. Diaz Ochoa, Juan G. & Mustafa, Faizan. (2021). Graph Neural Network Modelling as a potentially effective Method for predicting and analyzing Procedures based on Patient Diagnoses. 10.1101/2021.11.25.21266465.
11. Hanzlicek, Petr & Spidlen, Josef & Heroutova, Helena & Nagy, Miroslav. (2005). User interface of MUDR electronic health record. International journal of medical informatics. 74. 221-7. 10.1016/j.ijmedinf.2004.06.003.
12. Kaur, K., and Rani, R., Managing data in healthcare information systems: Many models, one solution. Computer 48(3):52–59, 2015. <https://doi.org/10.1109/MC.2015.77>
13. Kovantsev, Anton & Chunaev, Petr & Bochenina, Klavdiya. (2021). Evaluating Time Series Predictability via Transition Graph Analysis. 1039-1046. 10.1109/ICDMW53433.2021.00135.
14. Using Bayesian Networks for Discovering Temporal-State Transition Patterns in Hemodialysis (PDF). Hawaii International Conference on System Sciences. 6. 153. 10.1109/HICSS.2002.994123.