Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Sister Conferences Best Papers Track

# Rx-refill Graph Neural Network to Reduce Drug Overprescribing Risks (Extended Abstract)*

**Jianfei Zhang**[1,2], **Ai-Te Kuo**[3], **Jianan Zhao**[6], **Qianlong Wen**[6], **Erin Winstanley**[4], **Chuxu Zhang**[5#], **Yanfang Ye**[1,6#]
[1]Case Western Reserve University
[2]Université de Sherbrooke
[3]Auburn University
[4]West Virginia University
[5]Brandeis University
[6]University of Notre Dame
jianfei.zhang@usherbrooke.ca, aitekuo@auburn.edu, {jzhao8, qwen}@nd.edu,
erin.winstanley@hsc.wvu.edu, chuxuzhang@brandeis.edu, yye7@nd.edu

## Abstract

Prescription (aka Rx) drugs can be easily overprescribed and lead to drug abuse or opioid overdose. Accordingly, a state-run prescription drug monitoring program (PDMP) in the United States has been developed to reduce overprescribing. However, PDMP has limited capability in detecting patients' potential overprescribing behaviors, impairing its effectiveness in preventing drug abuse and overdose in patients. In this paper, we propose a novel model RxNet, which builds 1) a dynamic heterogeneous graph to model Rx refills that are essentially prescribing and dispensing (P&D) relationships among various patients, 2) an RxLSTM network to explore the dynamic Rx-refill behavior and medical condition variation of patients, and 3) a dosing-adaptive network to extract and recalibrate dosing patterns and obtain the refined patient representations which are finally utilized for overprescribing detection. The extensive experimental results on a one-year state-wide PDMP data demonstrate that RxNet consistently outperforms state-of-the-art methods in predicting patients at high risk of opioid overdose and drug abuse.

## 1 Introduction

The dispensation of prescription (aka Rx) drugs requires legal medical prescriptions since they can have powerful effects on the human brain and body, some of which are dangerous. Prescription drugs are easy to abuse. For example, opioid painkillers can bring intense pleasure and well-being feelings in the treatment of chronic diseases while making people increase the dose, which may lead to overdose deaths. The United States is amid opioid overdose [Hu *et al.*, 2021] and drug abuse epidemic [Schuchat *et al.*, 2017]. Opioid-involved overdose deaths rise from 21,088 in 2010 to 46,802 in 2018 [Hedegaard *et al.*, 2020]. Increasing rates of opioid overdose has become a prominent topic in public health. Early identification of overprescribing may prevent the problem from turning into a drug addiction [Qian *et al.*, 2021].

The prescription drug monitoring program (PDMP) is a jurisdictionally operated electronic database collected from pharmacies on controlled substances and Rx drugs dispensed to patients in a state. Although PDMP is developed to curb drug overprescribing, its utilization among prescribers is low [Haffajee *et al.*, 2015]. Recent studies have reported significant barriers when using the PDMP database, including difficulty in accessing the database and lack of medical knowledge of its usage [Garcia *et al.*, 2017]. For example, Grecu *et al.* [Grecu *et al.*, 2019] and Meara *et al.* [Meara *et al.*, 2016] found that PDMP has inconsistent and limited effects on detecting drug abuse. Fundamentally, it is due to a lack of effective PDMP data modeling to identify overprescribing behaviors which may cause drug abuse in patients. The challenges for modeling PDMP data are as follows:

- C1: When physicians consult the PDMP, they have access to various types of information about prescriptions. Every prescription includes information about physicians, patients, medication, and dispensing instructions. Therefore, the first challenge is how to model prescribing and dispensing (P&D) relationships (e.g., a patient visits physicians) among different Rx entries (e.g., patients, physicians).
- C2: The prescriptions are filled at different times, making the P&D interactions (relationships) in C1 dynamic and evolutionary. Also, the irregular prescription refill and various medication days' supply [Yang *et al.*, 2015] make the distribution of refills highly non-uniform, and the elapsed time between Rx records vary from days to months. Thus, the second challenge is how to model dynamic interactions and medical condition variation.

*Full version of this paper was originally published at the 2021 ACM International Conference on Information and Knowledge Management (CIKM) and received the Best Paper Award in Full Paper Track ([Zhang *et al.*, 2021]).
#: Corresponding authors.

- C3: Rx drugs are usually prescribed repeatedly and therefore patients would have multiple prescriptions which have an immediate impact on medication safety [Vaduganathan *et al.*, 2020]. The prescriptions refilled at different times have various significance on the risk of overprescribing [Gangal *et al.*, 2020]. Hence, the third challenge is how to capture the dosing patterns associated with overprescribing and capture informative patterns.

In light of the above challenges, we propose a novel model called RxNet to model PDMP data and predict overprescribing that is specifically referred to patients at high risk of drug abuse or opioid overdose. We introduce a P&D heterogeneous graph (P&D graph) to construct heterogeneous relationships (interactions) among different Rx entries. Based on P&D graph, we employ a graph neural network (GNN) [Fan *et al.*, 2022] to learn node embeddings by aggregating P&D information in the graph. To capture dynamics and explore the medical condition variation of patients, we design an Rx-refill LSTM (RxLSTM) that can deal with irregular Rx refill intervals by differentiating recent and historical information in cell state. Furthermore, a dosing-adaptive network (DAN) extracts and recalibrates dosing patterns concealed in prescriptions through convolution operation. The major contributions of this paper include: (1) We propose the problem of overprescribing detection for patients, which is important and meaningful; (2) To handle dynamic heterogeneous relationships among Rx entries, medical condition variation, and dosage patterns of patients, we develop the RxNet model by integrating graph neural network, recurrent neural network, and convolutional operation; (3) We collect a 1-year PDMP dataset and conduct extensive experiments. Promising results demonstrate the effectiveness of our model by a comparison with state-of-the-art methods for predicting overprescribing, i.e., patients at high risk of opioid overdose or drug abuse.

## 2 Methodology

Our model RxNet for modeling a patient's Rx refill records in PDMP is shown in Figure 1, which includes a temporal P&D graph, an Rx-refill LSTM, and a dosing-adaptive convolutional network.

### 2.1 P&D Graph Neural Network

P&D graph in each period $t$ represents the interactions (relationships) of different entries (nodes) during that period. We design a P&D graph neural network to model heterogeneous relationships and learn embeddings of patients. Firstly, let $\mathbf{h}^0$ be the initial node representation at $t = 0$. Nevertheless, nodes of various types have unequal feature dimensions. To address feature heterogeneity, we project the feature vector of the node of type $A$, i.e., $\mathbf{x} \in \mathbb{R}^{d_{\mathrm{A}}}$, onto a new feature space via a transformation matrix: $\mathbf{h}^0 = \mathbf{W}_{\mathrm{A}}^T \mathbf{x}$, where $\mathbf{W}_{\mathrm{A}} \in \mathbb{R}^{d_{\mathrm{A}} \times d_0}$. Then, let $\mathbf{h}_v^t$ denote the embedding of node $v$ at time $t$. Considering different nodes have different appearance time, we further introduce a time decay factor to quantify time influence and reformulate $\mathbf{h}_v^t$ as follows:

$$\mathbf{h}_v^t = \mathbf{h}_v^t \parallel \Phi_v(t), \tag{1}$$

where $\parallel$ is the concatenation operator and $\Phi_v(t)$ is time decay factor. Node appearing long before may have less impact on the current relationships and the node representations. Therefore, we formulate the decay factor as: $\Phi_v(t) = \exp(-\eta \cdot |t - \tau_v|)$ ($\tau_v$: the time when $v$ appears), where the decay coefficient $\eta > 0$ and $0 < \Phi \leq 1$. Then, we employ a self-attention [Vaswani *et al.*, 2017] to perform neighbor information aggregation and update node embeddings:

$$\begin{aligned} \mathbb{H}_v^t &= [\mathbf{h}_u^t \parallel \mathbf{e}_{vu}^t : \forall u \in \mathcal{N}_v^t] \\ \mathbf{h}_v^t &= \mathrm{attention}(\mathbf{h}_v^t \mathbf{W}_{\mathrm{q}}^t, \mathbb{H}_v^t \mathbf{W}_{\mathrm{K}}^t, \mathbb{H}_v^t \mathbf{W}_{\mathrm{V}}^t), \end{aligned} \tag{2}$$

where $\mathcal{N}_v^t$ denotes the neighbor set of node $v$ at $t$, $\mathbf{e}_{vu}^t \in \mathbb{R}^{d_{\mathrm{e}}}$ is edge feature, $\mathbf{W}_{\mathrm{q}}^t \in \mathbb{R}^{(d_0+1) \times d_{\mathrm{h}}}, \mathbf{W}_{\mathrm{K}}^t, \mathbf{W}_{\mathrm{V}}^t \in \mathbb{R}^{(|\mathcal{N}_v^t| \times (d_0+1+d_{\mathrm{e}})) \times d_{\mathrm{h}}}$. The three projection matrices capture the interactions in P&D graph using time information and neighboring node/edge features. Figure 1 (b) shows an illustration of P&D GNN.

### 2.2 Rx-refill LSTM

Since P&D graph is dynamic, we can employ the recurrent neural network (e.g., LSTM) to update patient embeddings. However, the memory cell of a standard LSTM tackles all historical prescription refills without distinction. Hence, we develop an RxLSTM (as shown in Figure 1 (c)) to infer a patient's medical condition variation $\epsilon^t$ (i.e., the indicator of an overdose) while updating the current representation $\mathbf{h}^t$ by considering elapsed time $\delta^t$ between consecutive refills:

$$[\mathbf{h}^t, \epsilon^t] = \mathrm{RxLSTM}(\mathbf{h}^{t-1}, \delta^t), \tag{3}$$

where $\mathbf{h}^t \in \mathbb{R}^d$ is computed as the Hadamard product ($\circ$) of the output gate of an ON-LSTM network (see [Zhang *et al.*, 2021] for more details). The master forget gate $\widetilde{\mathbf{F}}^t$ of the ON-LSTM controls the erasing behavior of the network, indicating where to store the medical condition information. Hence, we define the medical condition variation $\epsilon^t$ as follows:

$$\epsilon^t = \arg\max_i (\widetilde{\mathbf{F}}_i^t - \widetilde{\mathbf{F}}_{i-1}^t). \tag{4}$$

A large $\epsilon^t$ makes most historical information abandoned, indicating that the medical condition has largely changed, that is, a patient may have an overdose. The above definition is given by a non-differentiable function. Thus, we approximate it by the following estimation which approaches the probability that the $i$th entry of $\widetilde{\mathbf{F}}^t$ takes value 1:

$$\epsilon^t = \mathbb{E}[i|\widetilde{\mathbf{F}}_i^t = 1] = \sum_{i=1}^{d_{\text{master}}} \sum_{j \leq i} \Pr(\varrho = i) = d_{\text{master}} - \|\widetilde{\mathbf{F}}^t\|_1,$$

where $\varrho$ is the split point that divides the cell state into two segments: the 0-segment and the 1-segment, indicating a medical condition variation.

### 2.3 Dosing-Adaptive Network

The dosing patterns are similar for a certain medical condition, yet varies across different medical conditions [Jain *et al.*, 2015]. Hence, we develop a dosing-adaptive network to

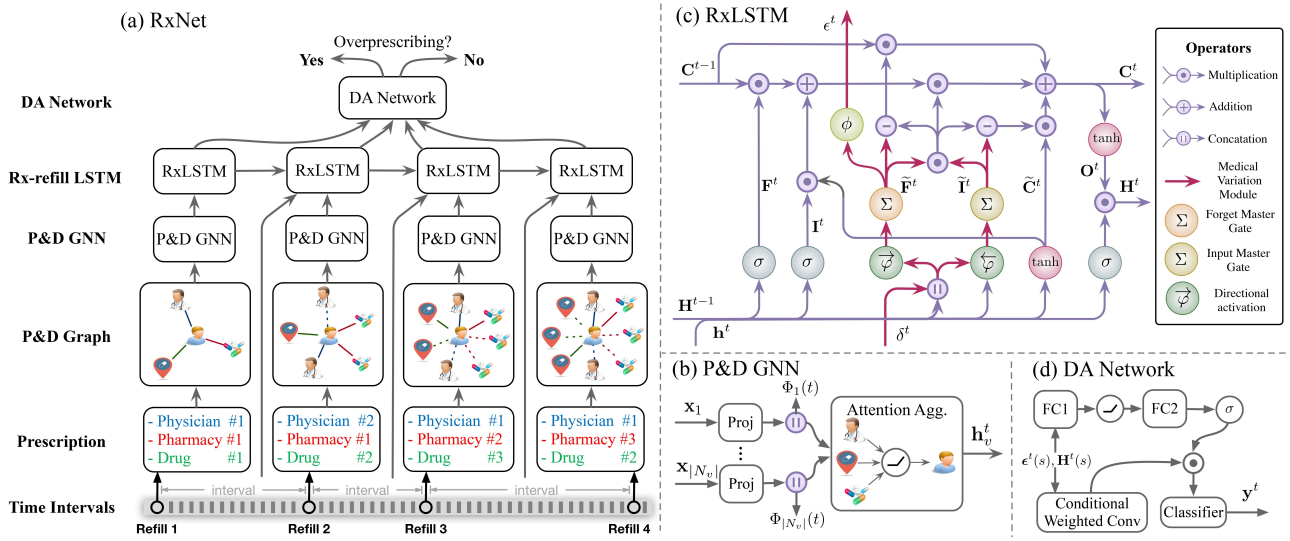


Figure 1: (a) The framework of RxNet; (b) The P&D graph neural network; (c) The Rx-refill LSTM module; (d) The dosing-adaptive network.

extract and recalibrate dosing patterns behind the prescriptions by fusing dynamic embedding of patient through a convolution operation, as shown in Figure 1 (d). Given a sequence of historical hidden states $[\mathbf{h}^{t-s}, \dots, \mathbf{h}^{t-1}, \mathbf{h}^t]$, and medical condition variations $[\epsilon^{t-s}, \dots, \epsilon^{t-1}, \epsilon^t]$ in previous $s$ steps (checkup window) before time $t$, we employ the convolutional network with filter kernels $\mathbf{K}^t = [\boldsymbol{\kappa}_1^t, \boldsymbol{\kappa}_2^t, \dots, \boldsymbol{\kappa}_m^t]$ to generate the dosing pattern:

$$\begin{aligned} \mathbf{z}_i^t &= \boldsymbol{\kappa}_i^t * (\overrightarrow{\varphi}(\boldsymbol{\epsilon}^t(s)) \circ \mathbf{H}^t(s)) \\ \mathbf{H}^t(s) &= [\mathbf{h}^{t-s}, \dots, \mathbf{h}^{t-1}, \mathbf{h}^t] \\ \boldsymbol{\epsilon}^t(s) &= [\epsilon^{t-s}, \dots, \epsilon^{t-1}, \epsilon^t], \end{aligned} \tag{5}$$

where $\boldsymbol{\kappa}_i^t$ is a 1D spatial kernel that acts on the corresponding channel of $\mathbf{H}^t(s) \in \mathbb{R}^{d \times (s+1)}$. Particularly, $\overrightarrow{\varphi}(\boldsymbol{\epsilon}^t(s)) \in [0,1]^{s+1}$ quantifies the probabilistic distances between the current refill record and historical refill records, whose values are monotonically increasing. A large value means the medical condition of $\mathbf{h}^{t-s+i}$ greatly differs from the current medical condition of $\mathbf{h}^t$. The input embeddings $\mathbf{H}^t(s)$ are weighted by the medical condition distance in each convolution. Therefore, the computed $\mathbf{z}_i^t$ of each kernel can extract dosing patterns for representing the whole medical condition. The convolution output $\mathbf{Z}^t = [\mathbf{z}_1^t, \mathbf{z}_2^t, \dots, \mathbf{z}_m^t]$ is then produced by summarizing multiple patterns through all channels, where the channel dependencies are embedded in $\boldsymbol{\kappa}_i^t$.

The dosing patterns have various significances associated with medical conditions [Gangal *et al.*, 2020; Islam, 2019]. The network is required to capture the importance of patterns and increase the sensitivity of informative patterns. For this purpose, we squeeze global spatial information into a channel descriptor $\boldsymbol{\gamma}^t$ and then recalibrate $\boldsymbol{\gamma}^t$ to obtain:

$$\begin{aligned} \mathbf{G}^t &= \sigma(\mathbf{W}_{g2} \cdot \mathrm{ReLU}(\mathbf{W}_{g1} \cdot \boldsymbol{\gamma}^t + \mathbf{b}_{g1}) + \mathbf{b}_{g2}) \\ \boldsymbol{\gamma}^t &= \frac{1}{s} \sum\nolimits_{i=0}^{s} \overrightarrow{\varphi}(\boldsymbol{\epsilon}^t(s))_i \cdot \mathbf{h}^{t-s+i}, \end{aligned} \tag{6}$$

where $\mathbf{G}_i^t$ measures the importance of pattern $\mathbf{z}_i^t$, $\mathbf{W}_{g1} \in \mathbb{R}^{d_g \times d}$, $\mathbf{W}_{g2} \in \mathbb{R}^{m \times d_g}$, $\mathbf{b} \in \mathbb{R}^{d_{g1}}$, and $\mathbf{b} \in \mathbb{R}^{d_{g2}}$. The descriptor $\boldsymbol{\gamma}^t \in \mathbb{R}^d$ at current refill is computed as the average of hidden states within the checkup window, and therefore can be regarded as the *dosing regimen theme*. As shown in Figure 1 (d), the recalibration is achieved by two fully connected layers: i) the dimensionality-reduction layer with the activation ReLU to compress the representation, and ii) the dimensionality-increasing layer returning to the channel dimension of patterns $\mathbf{Z}^t$. Finally, we rescale $\mathbf{Z}^t$ through a channel-wise gating mechanism [Hu *et al.*, 2018] and obtain $\widetilde{\mathbf{Z}}^t = \mathbf{Z}^t \circ \mathbf{G}^t$.

### 2.4 Objective Function

After obtaining $\widetilde{\mathbf{Z}}^t$, we employ a binary classifier to predict medical condition of the patient. The objective function is to minimize the cross-entropy loss over $T$ checkup windows:

$$\begin{aligned} L &= -\sum_{t=1}^{T} \hat{\mathbf{y}}^t \log \mathbf{y}^t + (1 - \hat{\mathbf{y}}^t) \log(1 - \mathbf{y}^t) \\ \mathbf{y}^t &= \sigma(\mathbf{W}_y \widetilde{\mathbf{Z}}^t + \mathbf{b}_y), \end{aligned} \tag{7}$$

where $\mathbf{W}_y \in \mathbb{R}^m$, $\mathbf{y}^t$ denotes the prediction score of overprescribing, and $\hat{\mathbf{y}}^t$ indicates the ground-truth label.

## 3 Experiments

### 3.1 Experimental Setup

**Dataset.** The collected PDMP data contains 2,751,137 prescriptions written by 41,303 physicians for 297,361 unique patients from the Ohio state in 2016. These prescriptions involve 90 different Rx drugs (including 16 opioids) and are (re)filled at total 2,862 pharmacies. We consider patients at high risk when their daily dose decreases in 4 consecutive refills. Among all the patients, 227,520 (77%) have taken at

| | 9 months | | | | 12 months | | | |
|---|---|---|---|---|---|---|---|---|
| | Opioid Overdose | | Drug Abuse | | Opioid Overdose | | Drug Abuse | |
| | PRAUC | F1 | PRAUC | F1 | PRAUC | F1 | PRAUC | F1 |
| LSTM | 64.65 (.32) | 65.56 (.42) | 68.58 (.27) | 67.86 (.19) | 68.18 (.22) | 68.93 (.20) | 62.48 (.27) | 65.32 (.32) |
| ON-LSTM | 69.28 (.47) | 73.85 (.23) | 72.34 (.29) | 70.44 (.37) | 74.54 (.26) | 72.71 (.24) | 71.42 (.23) | 70.29 (.19) |
| T-LSTM | 63.48 (.21) | 67.11 (.33) | 65.71 (.32) | 65.59 (.25) | 71.57 (.34) | 73.35 (.16) | 69.75 (.18) | 71.88 (.22) |
| StageNet | 74.89 (.23) | 74.91 (.29) | 73.22 (.28) | 77.93 (.27) | 74.41 (.25) | 71.93 (.20) | 70.80 (.24) | 73.92 (.21) |
| GraphSAGE | 73.11 (.20) | 72.18 (.15) | 73.26 (.17) | 70.42 (.26) | 70.25 (.38) | 69.82 (.44) | 72.51 (.27) | 72.40 (.22) |
| EvoNet | 66.82 (.26) | 67.68 (.28) | 67.84 (.47) | 64.36 (.24) | 69.74 (.35) | 69.43 (.23) | 62.05 (.12) | 65.72 (.17) |
| CTDNE | 65.25 (.34) | 66.91 (.26) | 69.48 (.23) | 71.30 (.28) | 73.72 (.15) | 68.57 (.33) | 75.28 (.43) | 74.89 (.29) |
| JODIE | 75.18 (.37) | 77.41 (.38) | 72.19 (.42) | 73.87 (.33) | 75.07 (.39) | 73.57 (.25) | 71.28 (.17) | 74.18 (.30) |
| TGAT | 75.71 (.19) | 73.36 (.34) | 76.20 (.22) | 78.97 (.28) | 73.41 (.27) | 75.24 (.20) | 73.12 (.24) | 71.52 (.33) |
| TGN | 77.30 (.14) | 76.28 (.42) | 74.43 (.33) | 73.44 (.20) | 77.56 (.32) | 78.41 (.21) | 75.19 (.45) | 75.47 (.23) |
| **RxNet** | **81.69** (.34) | **82.45** (.33) | **78.46** (.24) | **82.92** (.20) | **83.65** (.32) | **82.29** (.22) | **78.47** (.18) | **79.22** (.14) |

Table 1: Comparison of all models' prediction performances, in terms of PRAUC and F1, for two observation periods (9, 12 months) of follow-up study. The best results are highlighted in bold and the runner-up is underlined.

least an opioid in a year, 3,495 were at high risk of opioid overdose and 4,438 were at high risk of Rx drug abuse. The labels over time indicate whether a patient is at high risk of opioid overdose or drug abuse at each time period. Training and testing ratios are 70% and 30%, respectively.

**Baseline Methods and Evaluation Metrics.** We compare our model to ten baseline models, including LSTM [Hochreiter and Schmidhuber, 1997], ON-LSTM [Shen *et al.*, 2019], T-LSTM [Baytas *et al.*, 2017], StageNet [Gao *et al.*, 2020], EvoNet [Wu *et al.*, 2020], GraphSAGE [Hamilton *et al.*, 2017], CTDNE [Nguyen *et al.*, 2018], JODIE [Kumar *et al.*, 2019], TGAT [Xu *et al.*, 2020], and TGN [Rossi *et al.*, 2020]. To evaluate the model performance, we utilize two metrics: precision-recall AUC (PRAUC) and F1-score.

### 3.2 Performance Comparison

Table 1 shows the results on the two datasets spanning 9 and 12 months. It can be seen that all models perform better in the context of a long period of follow-up because they are highly dependent on data distribution. RxNet significantly outperforms all baseline methods in all settings. The average improvement over the best baseline method is 4.85% and 6.35% for PRAUC and F1-score, demonstrating its strong capability for overprescribing detection. From the longitudinal study perspective, the superiority of RxNet compared to others lies in its ability to explore structured and unstructured information and medical condition variation at different times.

### 3.3 Case Study

Figure 2 illustrates the predicted risk of Tramadol overdose for a 48-year-old female patient who has an overdose (marked as red point) in June 2016. The predicted risk remains relatively low by June 2016 and the medical condition variation has a small value. At the end of June, the variation reaches a maximum value and the predicted risk rises rapidly, indicating that the patient's medical condition becomes highly risky. Here, the medical condition of the patient is at quite high risk. In this case, physicians can take intervention in advance. To conclude, with the help of RxNet in providing time-sensitive risk prediction, physicians are able to forecast the overprescribing and prevent it in advance.

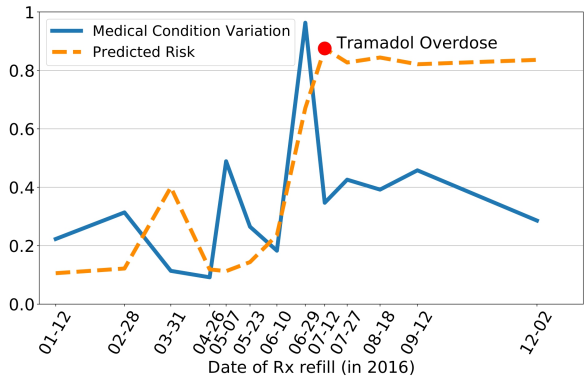


Figure 2: The medical condition variation and predicted risk of overprescribing Tramadol for a patient.

## 4 Conclusions

In this paper, we propose the RxNet model for detecting the Rx-refill caused overprescribing, i.e., predicting patients at high risk of opioid overdose or drug abuse. We construct a P&D graph and employ a self-attention based GNN to learn patient embeddings. To capture medical condition variations, we incorporate a newly designed RxLSTM to update dynamic patient embeddings. Moreover, we introduce a dosing-adaptive network to explore and recalibrate the dosing patterns of patients. The empirical results on a 1-year PDMP data demonstrate the effectiveness of RxNet in PDMP.

## Acknowledgments

This work is partially supported by the NSF under grants IIS-2209814, IIS-2203262, IIS-2214376, CNS-2203261, CMMI-2146076, CNS-1814825, IIS-2217239, IIS-2027127, and OAC-2218762, the NIJ 2018-75-CX-0032.

## References

[Baytas *et al.*, 2017] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. Patient subtyping via time-aware LSTM networks. In *KDD*, pages 65–74, 2017.

[Fan *et al.*, 2022] Yujie Fan, Mingxuan Ju, Chuxu Zhang, and Yanfang Ye. Heterogeneous temporal graph neural network. In *SDM*, pages 657–665, 2022.

[Gangal *et al.*, 2020] Neha S Gangal, Ana L Hincapie, and Roman Jandarov et al. Association between a state law allowing pharmacists to dispense naloxone without a prescription and naloxone dispensing rates. *JAMA Network Open*, 3(1):e1920310–e1920310, 2020.

[Gao *et al.*, 2020] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, and Jimeng Sun. StageNet: Stage-aware neural networks for health risk prediction. In *WWW*, pages 530–540, 2020.

[Garcia *et al.*, 2017] Ashley A Garcia, Kristen D Rosen, Erin Finley, and Jennifer Sharpe Potter. A systematic review of barriers and facilitators to implementing a prescription drug monitoring program. *Drug and Alcohol Dependence*, 100(171):e69, 2017.

[Grecu *et al.*, 2019] Anca M Grecu, Dhaval M Dave, and Henry Saffer. Mandatory access prescription drug monitoring programs and prescription drug abuse. *J. Policy. Anal. Manage.*, 38(1):181–209, 2019.

[Haffajee *et al.*, 2015] Rebecca L Haffajee, Anupam B Jena, and Scott G Weiner. Mandatory use of prescription drug monitoring programs. *JAMA*, 313(9):891–892, 2015.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, pages 1024–1034, 2017.

[Hedegaard *et al.*, 2020] Holly Hedegaard, Arialdi M. Miniño, and Margaret Warner. Drug overdose deaths in the United States, 1999–2018. *National Center for Health Statistics (NCHS)*, 2020.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[Hu *et al.*, 2018] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018.

[Hu *et al.*, 2021] Chuanbo Hu, Minglei Yin, Bin Liu, Xin Li, and Yanfang Ye. Detection of illicit drug trafficking events on instagram: A deep multimodal multilabel learning approach. In *CIKM*, pages 3838–3846, 2021.

[Islam, 2019] M Mofizul Islam. Pattern and probability of dispensing of prescription opioids and benzodiazepines among the new users in australia: a retrospective cohort study. *BMJ Open*, 9(12), 2019.

[Jain *et al.*, 2015] Shipra Jain, Prerna Upadhyaya, and Jaswant Goyal et al. A systematic review of prescription pattern monitoring studies and their effectiveness in promoting rational use of medicines. *Perspect. Clin. Res.*, 6(2):86, 2015.

[Kumar *et al.*, 2019] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *KDD*, pages 1269–1278, 2019.

[Meara *et al.*, 2016] Ellen Meara, Jill R Horwitz, Wilson Powell, Lynn McClelland, Weiping Zhou, A James O'malley, and Nancy E Morden. State legal restrictions and prescription-opioid use among disabled adults. *NEJM*, 375(1):44–53, 2016.

[Nguyen *et al.*, 2018] Giang Hoang Nguyen, John Boaz Lee, Ryan A Rossi, Nesreen K Ahmed, Eunyee Koh, and Sungchul Kim. Continuous-time dynamic network embeddings. In *WWW*, pages 969–976, 2018.

[Qian *et al.*, 2021] Yiyue Qian, Yiming Zhang, Yanfang Ye, and Chuxu Zhang. Distilling meta knowledge on heterogeneous graph for illicit drug trafficker detection on social media. 34:26911–26923, 2021.

[Rossi *et al.*, 2020] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael M. Bronstein. Temporal graph networks for deep learning on dynamic graphs. In *ICML*, 2020.

[Schuchat *et al.*, 2017] Anne Schuchat, Debra Houry, and Gery P Guy. New data on opioid use and prescribing in the United States. *JAMA*, 318(5):425–426, 2017.

[Shen *et al.*, 2019] Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron C. Courville. Ordered neurons: Integrating tree structures into recurrent neural networks. In *ICLR*, 2019.

[Vaduganathan *et al.*, 2020] Muthiah Vaduganathan, Jeroen van Meijgaard, Mandeep R Mehra, Jacob Joseph, Christopher J O'Donnell, and Haider J Warraich. Prescription fill patterns for commonly used drugs during the COVID-19 pandemic in the United States. *JAMA*, 2020.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017.

[Wu *et al.*, 2020] Changmin Wu, Giannis Nikolentzos, and Michalis Vazirgiannis. EvoNet: A neural network for predicting the evolution of dynamic graphs. In *ICANN*, pages 594–606, 2020.

[Xu *et al.*, 2020] Da Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. In *ICLR*, 2020.

[Yang *et al.*, 2015] Zhuo Yang, Barth Wilsey, and Michele Bohm et al. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *The Journal of Pain*, 16(5):445–453, 2015.

[Zhang *et al.*, 2021] Jianfei Zhang, Ai-Te Kuo, Jianan Zhao, Qianlong Wen, Erin L. Winstanley, Chuxu Zhang, and Yanfang Ye. RxNet: Rx-refill graph neural network for overprescribing detection. In *CIKM*, pages 2537–2546, 2021.