RxNet: Rx-refill Graph Neural Network for Overprescribing Detection

Jianfei Zhang
Case Western Reserve University, USA
University of Alberta, Canada
jianfei.zhang@live.ca

Ai-Te Kuo
Auburn University
Auburn, Alabama, USA
aitekuo@auburn.edu

Jianan Zhao, Qianlong Wen
Case Western Reserve University
Cleveland, Ohio, USA
{jxz1244,qxw294}@case.edu

Erin Winstanley
West Virginia University
Morgantown, West Virginia, USA
erin.winstanley@hsc.wvu.edu

Chuxu Zhang*
Brandeis University
Waltham, Massachusetts, USA
chuxuzhang@brandeis.edu

Yanfang Ye*
Case Western Reserve University
University of Notre Dame, USA
yanfang.ye@case.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. Despite a few machine-learning-based methods that have been proposed for detecting overprescribing, they usually ignore the patient prescribing behavior and their performances are not satisfying. In light of this, we propose a novel model RxNet for overprescribing detection in PDMP. RxNet builds a dynamic heterogeneous graph to model Rx refills that are essentially prescribing and dispensing (P&D) relationships among various Rx entries (e.g., patients) whose representations are encoded by graph neural network. In addition, to explore the dynamic Rx-refill behavior and medical condition variation of patients, an RxLSTM network is designed to update representations of patients. Based on the output of RxLSTM, a dosing-adaptive network is leveraged 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 1-year Ohio PDMP data demonstrate that RxNet consistently outperforms state-of-the-art methods in predicting patients at high risk of opioid overdose and drug abuse, with an average of 5.7% and 7.3% improvement on F1 score respectively.

CCS CONCEPTS

• **Applied computing → Health care information systems**; • **Computing methodologies → Neural networks**.

CIKM '21, November 1–5, 2021, Virtual Event, OLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3482465

KEYWORDS

Graph neural network; LSTM; Overprescribing detection; PDMP; Opioid overdose; Drug abuse

ACM Reference Format:
Jianfei Zhang, Ai-Te Kuo, Jianan Zhao, Qianlong Wen, Erin Winstanley, Chuxu Zhang*, and Yanfang Ye*. 2021. RxNet: Rx-refill Graph Neural Network for Overprescribing Detection. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, OLD, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3482465

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. The prescription drugs are easy to be abused. For example, opioid painkillers can bring intense pleasure and well-being feelings in the treatment of chronic diseases while make people increase the dose, which may lead to overdose deaths. The United States is amid opioid overdose and drug abuse epidemic [38]. In particular, opioid-involved overdose deaths rise from 21,088 in 2010 to 46,802 in 2018 [21]. Increasing rates of opioid overdose has become a prominent topic in public health. Early identification of overprescribing drugs may prevent the problem from turning into a drug addiction [31].

In the United States, 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 [17]. The 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 [13, 44]. For example, Grecu *et al.* [16] and Meara *et al.* [32] 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. To model PDMP data, there are several key challenges:

- C1: When physicians consult the PDMP, they have access to various types of information about prescriptions. Every prescription includes information of physicians, patients, medication, and

* Corresponding authors

dispensing instructions. Therefore, the first challenge is how to model prescribing and dispensing (P&D) relationships (e.g., a patient visits physicians, a physician prescribes drugs) 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 [44] and various medication days' supply [48] 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.
- C3: Rx drugs are usually prescribed repeatedly and therefore patients would have multiple prescriptions which have an immediate impact on medication safety [41]. The prescriptions refilled at different times have various significance on the risk of overprescribing [11]. Hence, the third challenge is how to capture the dosing patterns associated with the overprescribing and capture informative patterns behind the prescriptions.

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 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. To summarize, the major contributions of this paper include:

- We propose the problem of overprescribing detection for patients, which is important and meaningful.
- 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.
- 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 RELATED WORK

This paper is closely related to two research lines: overprescribing detection and representation learning on heterogeneous graphs.

2.1 Overprescribing Detection

Rx drugs are a very common medication treatment and strong evidence [37] indicates that they are prominently overprescribed. Medical studies on overprescribing generally involve opioid overdose and drug abuse [14]. Besides PDMP, there are some machine learning-based models for overprescribing detection. For example, a regression approach [27] is proposed for predicting opioid overdose at the state-level. Håkansson *et al.* [18] studied the associations between drug use patterns and criminal behaviors to forecast opioid overdose from crime data. Ertugrul *et al.* [8] develop a multi-head attentional network to learn different representation subspaces of features for opioid overdose forecast. Additionally, there have been some studies on PDMP for analyzing overprescribing [4, 44]. The work [1] provides an overview of the computational challenges and advances in drug response prediction. The deep neural network and gradient boosting models are used in [30] for predicting opioid overdose. Hastings *et al.* [20] analyzes PDMP data and government administrative data to predict the risk of drug abuse. Despite the progress of current studies, few of them focus on modeling and exploring PDMP data which contains dynamic heterogeneous relationships among Rx entries, medical condition variation and dosing patterns of patients. The method developed in [12] for predicting health risk reveals the effectiveness of utilizing an LSTM network followed by a convolutional network, which can be a candidate approach for addressing the issues in detecting overprescribing.

2.2 Representation Learning on Heterogeneous Graphs

Since the PDMP data can be modeled as a dynamic heterogeneous graph, our proposed model is based on representation learning on heterogeneous graphs, which learns a low-dimensional representation for each node and effectively preserves the heterogeneous graph structure. Early works [6, 9] mostly focus on preserving the metapath-based proximity. Recently, graph neural networks (GNNs) [19, 28, 47] are widely used and achieved state-of-the-art performance. These heterogeneous GNNs learn node embedding for each node by aggregating information from its original neighbors [51] or metapath-based neighbors [43, 50]. In addition, many dynamic graph embedding methods have been proposed to model the dynamic interactions in real-world graphs, including matrix factorization based [52], Skip-Gram based [7], auto-encoder based [15, 49], and neural-network based [2, 34, 36, 45, 46] models. While some of the above-mentioned methods focus on addressing the heterogeneity or dynamics of graphs, they can not fully address the inherent heterogeneity and dynamics inside the P&D graph while simultaneously incorporating the medicine knowledge.

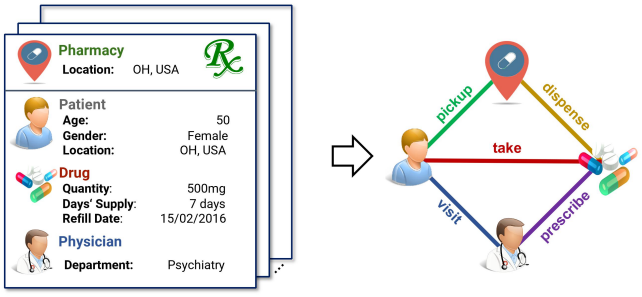
3 PRELIMINARIES

In this section, we first introduce the P&D graph, then formally define the problem of overprescribing detection.

Definition 1 (Heterogeneous Graph). A heterogeneous graph [40] $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{T}, \mathcal{R})$ consists of a set of nodes $\mathcal{V}$ connected by a set of edges $\mathcal{E}$. $\mathcal{T}$ and $\mathcal{R}$ represent the sets of node types and edges types. Each node $v \in \mathcal{V}$ and edge $e \in \mathcal{E}$ is associated with a node type mapping function $\psi_v : \mathcal{V} \rightarrow \mathcal{T}$ and an edge type mapping function $\phi_e : \mathcal{E} \rightarrow \mathcal{R}$, respectively. In addition, each node $v \in \mathcal{V}$ could be associated with attribute information $\mathcal{X}_v$.

Based on the above definition, we can construct a P&D graph to model the interactions among four types of P&D entries (i.e., patient, physician, drug, and pharmacy) by using five types of relationships (i.e., patients visit physicians, physicians prescribe drugs, patients take drugs, patients pick up at pharmacies, and

pharmacies dispense drugs). For example, as shown in Figure 1 (a), a 50-year-old psychiatry patient visited the physician who wrote a 7-day supply prescription with 500mg of drugs (e.g., Trazodone) which is refilled on 15/02/2016. The patient picked up the drugs from a pharmacy that is permitted to dispense such Rx drugs. The P&D graph schema is shown in Figure 1 (b).



(a) Prescriptions (b) P&D Graph Schema

Figure 1: An example of a prescription (a) and its corresponding P&D graph schema.

Since Rx drugs have the strictest regulation and prescription requirements, patients have to periodically refill the prescriptions. Thus, given a bunch of refilled prescriptions in different periods, we can build a dynamic P&D graph that can reveal dynamic relationships among P&D entries. P&D graph in each period represents all P&D behaviors during that period. With the dynamic P&D graph, we formally define the problem as follows.

Definition 2 (Overprescribing Prediction). Given the dynamic P&D graph, our task is to learn embeddings $\mathbf{Z}^t$ of P&D entries at each time t. Then the learned patient embeddings $\mathbf{Z}^t_{\text{patient}}$ are further utilized for overprescribing prediction. The prediction problem is formulated as learning a function $f(\mathbf{Z}^t_{\text{patient}}) \rightarrow y^t_{\text{patient}}$ that maps the patient embedding to the binary label of overprescribing incident at given time t, where the overprescribing is referred to high risk of opioid overdose or drug abuse in this work.

4 PROPOSED MODEL

Our model RxNet for modeling a patient's Rx refill records in PDMP is shown in Figure 2. Specifically, at first, RxNet takes a sequence of prescription records represented as a temporal P&D graph and employs a graph neural network with self-attention to learn embeddings of Rx entries. Then, it introduces an Rx-refill LSTM to capture medical condition variation (i.e., irregular refill behavior) and update dynamic embeddings of patients. Finally, a dosing-adaptive convolutional network is designed to capture the dosing patterns and generate refined embeddings of patients which are finally used for overprescribing detection. We elaborate these three components in the following of this section.

4.1 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., $\mathbf{x} \in \mathbb{R}^{d_A}$, onto a new feature space via a transformation matrix: $\mathbf{h}^0 = \mathbf{W}_A^T \mathbf{x}$, where $\mathbf{W}_A \in \mathbb{R}^{d_A \times d_0}$. Then, let $\mathbf{h}^t_v$ 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 $\mathbf{h}^t_v$ as follows:

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

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: $\Phi_v(t) = \exp(-\eta \cdot |t - \tau_v|)$ (τ_v: the time when v appears), where the decay coefficient $\eta > 0$ and $0 < \Phi \leq 1$. Then, we employ a self-attention [42] to perform neighbor information aggregation and update node embeddings as follows:

$$\begin{aligned} \mathbb{H}^t_v &= [\mathbf{h}^t_u \parallel \mathbf{e}^t_{vu} : \forall u \in \mathcal{N}^t_v] \\ \mathbf{h}^t_v &= \text{attention}(\mathbf{h}^t_v \mathbf{W}^t_{\text{q}}, \mathbb{H}^t_v \mathbf{W}^t_{\text{K}}, \mathbb{H}^t_v \mathbf{W}^t_{\text{V}}), \end{aligned} \tag{2}$$

where $\mathcal{N}^t_v$ denotes the neighbor set of node v at t, $\mathbf{e}^t_{vu} \in \mathbb{R}^{d_e}$ is edge feature, $\mathbf{W}^t_{\text{q}} \in \mathbb{R}^{(d_0+1)\times d_h}$, $\mathbf{W}^t_{\text{K}}, \mathbf{W}^t_{\text{V}} \in \mathbb{R}^{(|\mathcal{N}^t_v|\times(d_0+1+d_e))\times d_h}$. The three projection matrices capture the interactions in P&D graph using time information and neighboring node/edge features. The attention weights indicate how a neighboring node contributes to generating target node embedding. Therefore, by using the P&D graph neural network, we can obtain embeddings of patients at each time period. Figure 2 (b) shows an illustration of P&D GNN.

4.2 Rx-refill LSTM

Since P&D graph is dynamic, we can employ the recurrent neural network (e.g., LSTM) to update patient embeddings obtained from Section 4.1. However, the memory cell of a standard LSTM tackles all historical prescription refills without distinction. It thus cannot discern a hierarchy of therapeutic effect using refilled prescriptions between the neurons: how the historical prescription is transferred among the cell states. Differentiating recent and historical cell states can tell us whether the change to cell state is due to historical refill or recent refill. A change due to the recent refills indicates that the underlying medical condition has just changed. For this purpose, we develop an Rx-refill LSTM by which we can derive the medical condition variation of patients through the cell state change, as shown in Figure 2 (c).

4.2.1 RxLSTM. The goal of RxLSTM is to infer a patient's medical condition variation ϵ^t (i.e., the indicator of the an overdose) while updating the current representation $\mathbf{h}^t$ by considering elapsed time δ^t between consecutive refills:

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

where $\mathbf{h}^t \in \mathbb{R}^d$ is computed as the Hadamard product ($\circ$) of the output gate $\mathbf{O}^t$ and the cell state $\mathbf{C}^t$, i.e., $\mathbf{h}^t = \mathbf{O}^t \circ \tanh(\mathbf{C}^t)$. The cell state $\mathbf{C}^t$ is computed using two intermediates: the standard memory cell $\mathbf{C}^t_{\text{standard}}$ and the master memory cell $\mathbf{C}^t_{\text{master}}$. Both

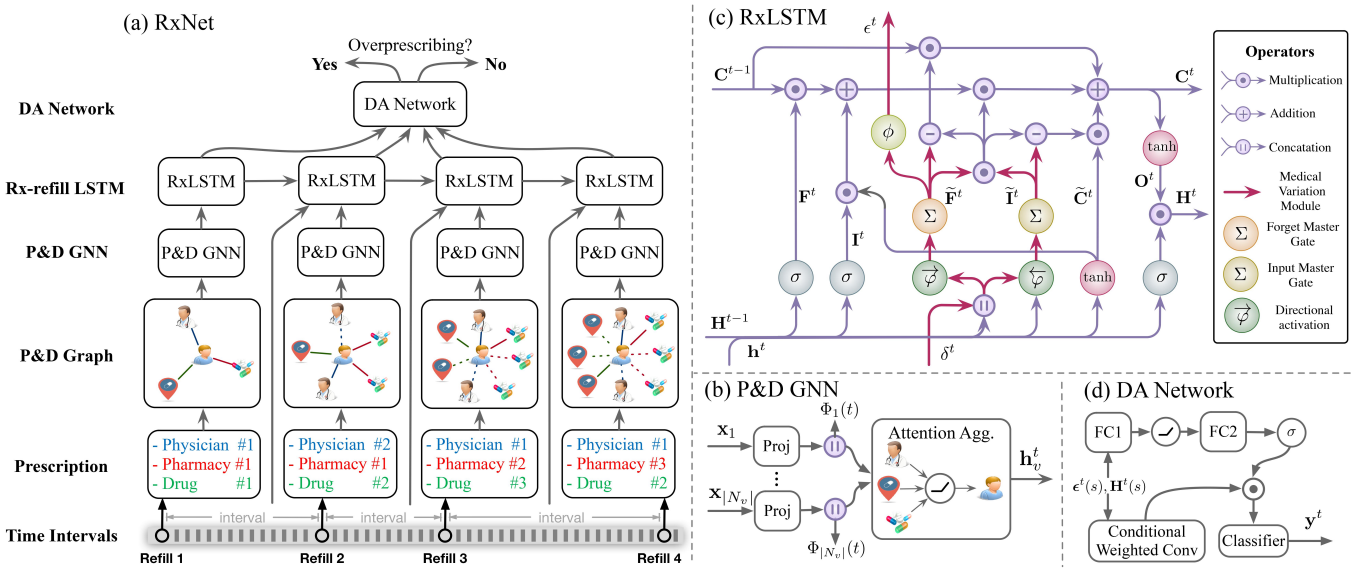


Figure 2: (a) The overall architecture of RxNet. It first takes the P&D data across different times to construct temporal P&D graph. Then, RxNet aggregates the heterogeneous neighborhood information by P&D GNN and further encodes the dynamic node embeddings by RxLSTM. Finally, DA network is employed to capture the dosing patterns and generate refined embeddings of patients for overprescribing detection; (b) The P&D graph neural network with core operators summarized in the right part; (c) The Rx-refill LSTM module; (d) The dosing-adaptive network.

of the two cells are updated by partially forgetting the previous memory $\mathbf{C}^{t-1}$ and adding a new (candidate) memory content $\widetilde{\mathbf{C}}^t$:

$$\begin{aligned} \mathbf{C}^t &= \Omega^t \circ \mathbf{C}^t_{\text{standard}} + \mathbf{C}^t_{\text{master}} \\ \mathbf{C}^t_{\text{standard}} &= \mathbf{F}^t \circ \mathbf{C}^{t-1} + \mathbf{I}^t \circ \widetilde{\mathbf{C}}^t \\ \mathbf{C}^t_{\text{master}} &= (\widetilde{\mathbf{F}}^t - \Omega^t) \circ \mathbf{C}^{t-1} + (\widetilde{\mathbf{I}}^t - \Omega^t) \circ \widetilde{\mathbf{C}}^t \\ \Omega^t &= \widetilde{\mathbf{F}}^t \circ \widetilde{\mathbf{I}}^t, \end{aligned} \tag{4}$$

where $\mathbf{F}^t, \mathbf{I}^t, \mathbf{O}^t, \widetilde{\mathbf{C}}^t \in \mathbb{R}^{d_g}$ are forget gate, input gate, output gate, and candidate cell used to control the erasing and writing operation on cell state $\mathbf{C}^t$. We modify the master forget gate $\widetilde{\mathbf{F}}$ and master input gate $\widetilde{\mathbf{I}}$ by integrating time interval information into the current input variables $\widetilde{\mathbf{h}}^t$ and the previous hidden state $\mathbf{h}^{t-1}$ [39]:

$$\begin{pmatrix} \mathbf{F}^t \\ \mathbf{I}^t \\ \mathbf{O}^t \\ \widetilde{\mathbf{C}}^t \\ \widetilde{\mathbf{F}}^t \\ \widetilde{\mathbf{I}}^t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \\ \overrightarrow{\varphi} \\ \overleftarrow{\varphi} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{\mathsf{F}} \,\|\, \boldsymbol{\omega}_{\mathsf{F}} \\ \mathbf{w}_{\mathsf{I}} \,\|\, \boldsymbol{\omega}_{\mathsf{I}} \\ \mathbf{w}_{\mathsf{O}} \,\|\, \boldsymbol{\omega}_{\mathsf{O}} \\ \mathbf{w}_{\mathsf{C}} \,\|\, \boldsymbol{\omega}_{\mathsf{C}} \\ \mathbf{w}_{\widetilde{\mathsf{F}}} \,\|\, \boldsymbol{\omega}_{\widetilde{\mathsf{F}}} \\ \mathbf{w}_{\widetilde{\mathsf{I}}} \,\|\, \boldsymbol{\omega}_{\widetilde{\mathsf{I}}} \end{pmatrix} \begin{bmatrix} \widetilde{\mathbf{h}}^t \| \delta^t \\ \mathbf{h}^{t-1} \| \delta^t \end{bmatrix} + \begin{pmatrix} \mathbf{b}_{\mathsf{F}} \\ \mathbf{b}_{\mathsf{I}} \\ \mathbf{b}_{\mathsf{O}} \\ \mathbf{b}_{\mathsf{C}} \\ \mathbf{b}_{\widetilde{\mathsf{F}}} \\ \mathbf{b}_{\widetilde{\mathsf{I}}} \end{pmatrix}, \tag{5}$$

where $\mathbf{w}_{\mathsf{F}}, \mathbf{w}_{\mathsf{I}}, \mathbf{w}_{\mathsf{O}}, \mathbf{w}_{\mathsf{C}} \in \mathbb{R}^{M \times d}$, $\boldsymbol{\omega}_{\mathsf{F}}, \boldsymbol{\omega}_{\mathsf{I}}, \boldsymbol{\omega}_{\mathsf{O}}, \boldsymbol{\omega}_{\mathsf{C}} \in \mathbb{R}^{M \times M}$ are weight matrices, $\mathbf{b}_{\mathsf{F}}, \mathbf{b}_{\mathsf{I}}, \mathbf{b}_{\mathsf{O}}, \mathbf{b}_{\mathsf{C}} \in \mathbb{R}^{M}$ are bias terms of standard cell. Besides, $\mathbf{w}_{\widetilde{\mathsf{F}}}, \mathbf{w}_{\widetilde{\mathsf{I}}} \in \mathbb{R}^{M \times (d + d_{\mathrm{h}} + 2)}$, $\boldsymbol{\omega}_{\widetilde{\mathsf{F}}}, \boldsymbol{\omega}_{\widetilde{\mathsf{I}}} \in \mathbb{R}^{M \times M}$ are weight matrices, $\mathbf{b}_{\widetilde{\mathsf{F}}}, \mathbf{b}_{\widetilde{\mathsf{I}}} \in \mathbb{R}^{M}$ are bias terms of the master cell. The activation function φ is a cumulative sum of the softmax, i.e., $\varphi(\cdot) = \text{cumsum}(\text{softmax}(\cdot))$. The arrow indicates the direction of cumulative sum. The values of $\widetilde{\mathbf{F}}^t$ monotonically increase from 0 to 1 and the values of $\widetilde{\mathbf{I}}^t$ monotonically decrease from 1 to 0. The master gates control how prescription information is stored in different neurons. $\widetilde{\mathbf{F}}^t$ decides which dimensions of $\mathbf{C}^t$ to store long-term Rx-refill information (i.e., $\mathbf{C}^{t-1}$), and $\widetilde{\mathbf{I}}^t$ decides which dimensions to store short-term Rx-refill information (i.e., $\widetilde{\mathbf{C}}^t$). The product of the two master gates, i.e., Ω^t, determines which dimensions to store the overlap between $\mathbf{C}^{t-1}$ and $\widetilde{\mathbf{C}}^t$. The independent information of $\mathbf{C}^{t-1}$ and $\widetilde{\mathbf{C}}^t$ stored in $\mathbf{C}^t$ are computed as $\widetilde{\mathbf{F}}^t - \Omega^t$ and $\widetilde{\mathbf{I}}^t - \Omega^t$, respectively. After formulating $\mathbf{h}^t$, we now describe how to compute the medical condition variation ϵ^t.

4.2.2 Medical Condition Variation. Essentially, the master forget gate $\widetilde{\mathbf{F}}^t$ controls the erasing behavior of the network, from which we know where to store the medical condition information. Hence, we define the medical condition variation ϵ^t as follows:

$$\epsilon^t = \arg\max_i (\widetilde{\mathbf{F}}^t_i - \widetilde{\mathbf{F}}^t_{i-1}). \tag{6}$$

The value of ϵ^t decides how much historical information is used to calculate $\mathbf{C}^t$. A large ϵ^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 ith entry of $\widetilde{\mathbf{F}}^t$ takes value 1:

$$\epsilon^t = \mathbb{E}[i | \widetilde{\mathbf{F}}^t_i = 1] = \sum_{i=1}^{d_{\text{master}}} \sum_{j \le i} \Pr(\varrho = i) = d_{\text{master}} - \|\widetilde{\mathbf{F}}^t\|_1, \tag{7}$$

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

4.3 Dosing-Adaptive Network

Medical studies [35, 41] have demonstrated that multiple prescriptions have an immediate impact on drug use safety. The dosing patterns are similar for a certain medical condition, yet varies across different medical conditions [26]. Hence, we develop a dosing-adaptive network to extract and recalibrate dosing patterns behind the prescriptions by fusing dynamic embedding of patient through a convolution operation, as shown in Figure 2 (d).

4.3.1 Learning Dosing Patterns. Given a sequence of historical hidden states $[\mathbf{h}^{t-s}, \ldots, \mathbf{h}^{t-1}, \mathbf{h}^{t}]$, and medical condition variations $[\epsilon^{t-s}, \ldots, \epsilon^{t-1}, \epsilon^{t}]$ (obtained from Section 4.2) in previous s steps (checkup window) before time t, we employ the convolutional network with filter kernels $\mathbf{K}^t = [\boldsymbol{\kappa}_1^t, \boldsymbol{\kappa}_2^t, \ldots, \boldsymbol{\kappa}_m^t]$ to generate the dosing pattern:

$$\begin{aligned} \mathbf{z}_i^t &= \boldsymbol{\kappa}_i^t * (\overrightarrow{\varphi}(\boldsymbol{\epsilon}^t(s)) \circ \mathbf{H}^t(s)) \\ \mathbf{H}^t(s) &= [\mathbf{h}^{t-s}, \ldots, \mathbf{h}^{t-1}, \mathbf{h}^{t}] \\ \boldsymbol{\epsilon}^t(s) &= [\epsilon^{t-s}, \ldots, \epsilon^{t-1}, \epsilon^{t}], \end{aligned} \tag{8}$$

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 $\mathbf{Z}^t = [\mathbf{z}_1^t, \mathbf{z}_2^t, \ldots, \mathbf{z}_m^t]$ is then produced by summarizing multiple patterns through all channels, where the channel dependencies are implicitly embedded in $\boldsymbol{\kappa}_i^t$.

4.3.2 Recalibration of Patterns. The dosing patterns have various significances associated with medical conditions [11, 25]. 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:

$$\begin{aligned} \mathbf{G}^t &= \sigma(\mathbf{W}_{g2} \cdot \mathrm{ReLU}(\mathbf{W}_{g1} \cdot \boldsymbol{\gamma}^t + \mathbf{b}_{g1}) + \mathbf{b}_{g2}) \\ \boldsymbol{\gamma}^t &= \frac{1}{s} \sum\nolimits_{i=0}^{s} \overrightarrow{\varphi}(\boldsymbol{\epsilon}^t(s))_i \cdot \mathbf{h}^{t-s+i}, \end{aligned} \tag{9}$$

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 2 (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 [24] and obtain:

$$\widetilde{\mathbf{Z}}^t = \mathbf{Z}^t \circ \mathbf{G}^t. \tag{10}$$

The channel-wise attention weights of patterns are calculated by the dosing regimen theme at the current refill instead of using global average pooling to generate channel-wise statistics or calculating alignment between historical representations.

4.4 Objective Function

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

$$\begin{aligned} L &= -\sum_{t=1}^{T} \widehat{\mathbf{y}}^t \log \mathbf{y}^t + (1 - \widehat{\mathbf{y}}^t) \log(1 - \mathbf{y}^t) \\ \mathbf{y}^t &= \sigma(\mathbf{W}_y \widetilde{\mathbf{Z}}^t + \mathbf{b}_y), \end{aligned} \tag{11}$$

where $\mathbf{W}_y \in \mathbb{R}^m$, $\mathbf{y}^t$ denotes the prediction score of overprescribing, and $\widehat{\mathbf{y}}^t$ indicates the ground truth, e.g., high risk of opioid overdose or drug abuse.

5 EXPERIMENTS

We conduct extensive experiments on the PDMP data and evaluate the performance of RxNet.

5.1 PDMP Data

The PDMP data contain 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. Figure 3 lists the 20 mostly prescribed drugs, of which 8 are opioids and 12 are non-opioid. We can see that Oxycodone and Hydrocodone are the two most popular opioids with over 500k and 300k refilled prescriptions, respectively. Table 1 shows the statistics of Rx entries and P&D relations in this PDMP data. (N.B.: Most patients have 1 or 2 Rx refills every month.)

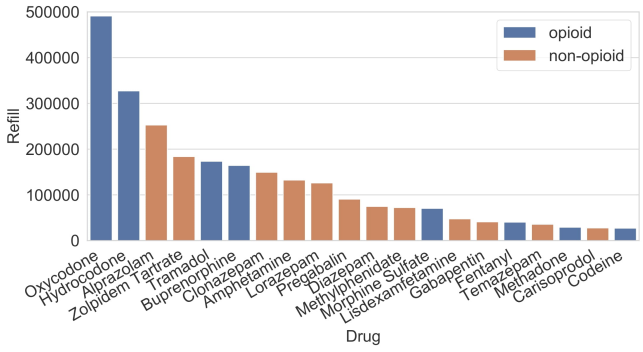


Figure 3: The 20 mostly prescribed drugs (8 opioids in cyan).

In addition, Table 2 provides an overview of the demographic characteristics (attributes) of patients. We crawl the adverse drug reactions (e.g., constipation caused by opioids) from VigiAccess (http://www.vigiaccess.org), and select the 26 most common reactions as the attributes of a drug. The pharmacy feature is a one-hot vector of 20 retail pharmacies such as Walgreens, CVS Health, Walmart Stores, Rite Aid Corp, etc. The specialties, e.g., Internal, Psychiatry, Oncology, are used as the attributes of physicians.

Table 1: Statistics of the PDMP data.

Rx entry	Count	P&D relation	Count
Patient	297,361	Patient – Drug	713,471
Physician	41,303	Patient – Physician	830,433
Drug (Opioid)	90 (16)	Patient – Pharmacy	553,893
Pharmacy	2,862	Physician – Drug	221,653
		Pharmacy – Drug	72,318

Table 2: Demographics of patients.

Observation	Number of patients (%)
Gender: female / male	128,026 (43%) / 169,235 (57%)
Age Group: <18 / 18-29	18,703 (6%) / 14,319 (5%)
30-44 / 45-64	55,417 (19%) / 143,545 (48%)
65-79 / ≥80	51,789 (17%) / 13,587 (5%)
Residence: urban / rural	142,733 (48%) / 154,628 (52%)

5.2 Baseline Methods

We compare our model with ten baseline methods as follows.

- LSTM [22] takes patients' features as input and models the prescription dynamics to generate the prediction results.
- ON-LSTM [39] separates hidden state dimensions with long and short-term information. It takes the patient features as input and differentiates the long-term and short-term information in hidden state dimensions.
- T-LSTM [2] incorporates the time decay to weaken the impact of historical P&D information when learning the embeddings on the base of a standard LSTM architecture.
- StageNet [12] uses patients' time-specific information and time interval information to learn the temporal representations and stage variation, which are then used as the input of a stage-adaptive convolutional network.
- EvoNet [45] updates the node representations in P&D graph via a GRU recurrent network. It employs a generative model which predicts the topology of the graph at the next time step and constructs a graph instance that corresponds to that topology.
- GraphSAGE [19] aggregates information from sampled neighbors and the temporal neighborhood is aggregated with equal attention coefficients.
- CTDNE [33] learns dynamic node embeddings directly from the P&D graph by temporal random walk based proximity.
- JODIE [29] couples recurrent model to jointly learn the dynamic embeddings of patients from a sequence of P&D interactions. The patient embedding is updated by two mutually recursive recurrent neural networks.
- TGAT [46] adopts temporal graph attention layer to aggregate temporal neighborhood features with time-feature interactions. It makes use of the edge/relation information (i.e., dosage) and employs self-attention to learn node embeddings.
- TGN [36] constructs temporal graph network and learns the node embeddings using both node and edge information through self-attention.

5.3 Experimental Settings

Data Settings. We study two overprescribing detection tasks, i.e., predicting high risk opioid overdose and high risk drug abuse. According to the medical settings in [23], we consider patients at high risk when their daily dose decreases in 4 consecutive refill prescriptions. Among the total 297,361 patients, 227,520 (77%) have taken at least an opioid during the 1-year follow-up period, 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 ti me period. Training and testing ratios are 70% and 30%, respectively.

MME Calculation. The patients are considered at high risk of overdose if they have a maximum average daily dose greater than the CDC-recommended cutoff of 90 Morphine Milligram Equivalents (MME) [10]. The opioid overdose potential is often evaluated by the MME per day, which is formally defined as follows:

$$\begin{aligned}\text{MME/day} =&(\text{Number of Units or Days' Supply})\times\\ &\text{Strength per Unit}\times\text{MME Conversion Factor},\end{aligned}$$

The opioid dose for each prescription can be computed according to the total MME. Higher dosages of opioids are associated with a higher risk of overdose and death. Calculating the total daily dosage of opioids helps identify patients who may be at high risk of overdose. Concretely, we provide an example to illustrate the calculation of MME.

Example 1. A patient who is suffering from chronic lower back pain takes ER oxycodone 30mg BID. The total daily amount of each prescription (i.e., Oxycodone/day) is 30 mg/unit × 2 units = 60mg. The patient is at high risk of an overdose because (s)he is prescribed 60 mg × 1.5 = 90 MME/day in total.

Parameter Settings. The representation dimension for all models is set as 128. For RxNet, the checkup window length s determines the timescale of the extracted dosing patterns is set to 1 month. The time decay coefficient η in Eq. 1 is set to 1.0. The feature dimension of edge $\mathbf{e}$ in Eq. 2 is set to 2, including the day's supply and quantity. We set the kernel size as the length of the checkup window, which allows each kernel to extract different patterns that represent the whole regimen. Furthermore, in order to reduce the number of parameters in RxNet, we configure $\widetilde{\mathbf{F}}^t, \widetilde{\mathbf{I}}^t \in [0,1]^{d_{\text{master}}}$, where their dimension is reduced to $d_{\text{master}} = d/c$ by a downsize factor c. By doing so, every neuron within each c-sized chunk shares the same master gates. A smaller c can make the model describe the severity variation in more detail.

ON-LSTM is implemented with chunk size 10. We use LSTM aggregator for GraphSAGE. For StageNet, the patient's visit information in the original work is replaced by the P&D representation. For other baseline methods, we use the default setting adopted in the original papers. All the models are implemented by Tensorflow with the Adam optimizer. Code will be available upon publication.

Evaluation Metrics. To evaluate the overprescribing detection (binary classification) performance, we utilize two popular metrics: precision-recall AUC (PRAUC) and F1 score.

5.4 Performance Comparison

To evaluate model performances on various observational data with different periods, we construct four datasets spanning 3 months, 6 months, 9 months, and the entire period (12 months), respectively. All results are reported in Table 3. Overall, 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 are 4.85% and 6.35% for PRAUC and F1 score, demonstrating its strong capability in learning patient embeddings for overprescribing detection. From the longitudinal study perspective, the superiority of RxNet compared to others lies in its ability of exploring structured and unstructured information, as well as medical condition variation at different time. RxNet has better performance than dynamic graph neural network models (i.e., TGAT, TGN), showing the benefit of incorporating medical condition variation into dynamic models. It achieves a higher PRAUC and F1 than StageNet mainly because the patients' P&D representations fed into LSTM part in RxNet are more informative than the original patients' prescription information provided to ON-LSTM in StageNet. Additionally, we can find ON-LSTM obtains higher prediction scores than LSTM since the exploration of neuron dependencies in LSTM can determine the short-term and long-term P&D information between neurons. Moreover, TGN and TGAT have better performances than static graph neural network model (i.e., GraphSAGE), indicating the strength of considering temporal information in P&D graph.

5.5 Ablation Study

Since RxNet integrates several essential modules (e.g., P&D GNN, RxLSTM, and DAN), we conduct extensive ablation studies to analyze the contributions of different modules by considering the following five model variants:

- **RxNet-Edge**: The edge features in P&D GNN is discarded when performing neighborhood information aggregation in Eq. 2.
- **RxNet-Time**: The time-aware decay is not considered for computing the node representations in Eq. 1.
- **RxLSTM**: The outputs of RxLSTM $[\mathbf{h}^t, \ldots, \mathbf{h}^T]$ are directly used for prediction. That is, removing the DAN in RxNet.
- **RxNet-Con**: The condition-weighted convolution $\mathbf{z}^t$ in Eq. 8 is estimated by the average of $\mathbf{H}^t$ within the checkup window.
- **RxNet-Rec**: The convolution output $\mathbf{Z}^t$ is used as the final representations without recalibration of patterns in Eq. 9.

The results of all variant models are reported in Table 4. From this table, we can find RxNet consistently outperforms all model variants. Specifically, the superior performance of RxNet relative to RxNet-Edge and RxNet-Time reveals that using P&D relation information and time decay can strengthen P&D GNN. In addition, the improvement of RxNet over RxLSTM demonstrates the benefit of DAN. Moreover, the comparison between RxNet and RxNet-Con reveals the importance of considering the medical condition variation. The superior performance of RxNet relative to RxNet-Rec demonstrates the effectiveness of recalibrating dosing patterns.

5.6 Performance over Different Drugs

To further show the performance of our model over different drugs, we report the predicted results in terms of F1 on every drug in Figure 4. As can be seen from the figure, for the most commonly used opioids such as Tramadol, Oxycodone, and Morphine Sulfate, RxNet can yield a 0.75 or higher F1 score. This demonstrates that our proposed model can effectively predict overprescribing for popular opioids. Similarly, RxNet achieves as high as a 0.8 F1 score when it predicts drug abuse of Tramadol, Fentanyl, and Zolpidem Tartrate which are the commonly prescribed drugs in PDMP. Both the opioid overdose and drug abuse for opioids Tramadol, Morphine Sulfate, and Hydrocodone can be easily predicted by our model. Furthermore, the overdose on Fentanyl is hard to predict in comparison with abuse in patients. The effective predictions on the commonly used opioids and non-opioid Rx drugs indicate that RxNet can forecast the drug crisis, i.e., the potential risk for a population when the drugs are dispensed.

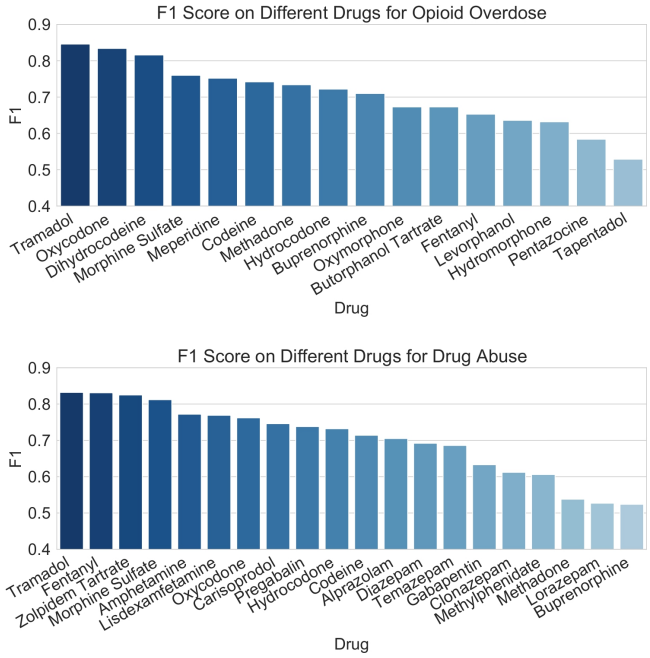


Figure 4: The performance (F1 score) of RxNet over different drugs for opioid overdose (top) and drug abuse (bottom).

5.7 Case Study

It is vital for physicians to evaluate the risk of overprescribing for each patient timely. To demonstrate the effectiveness of RxNet in capturing the time-sensitive patterns, we show a real example of RxNet application. Figure 5 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

Table 3: Comparison of all models' prediction performances, in terms of PRAUC and F1, in various observation periods (3, 6, 9, 12 months) of follow-up study. The numbers in parentheses are the corresponding standard deviations. The best results are highlighted in bold and the runner-up is underlined.

	Opioid Overdose		Drug Abuse		Opioid Overdose		Drug Abuse	
	PRAUC	F1	PRAUC	F1	PRAUC	F1	PRAUC	F1
	3 months				**6 months**			
LSTM	61.44 (.42)	60.82 (.24)	57.96 (.27)	58.14 (.53)	65.49 (.21)	66.24 (.27)	68.23 (.20)	64.37 (.38)
ON-LSTM	70.56 (.30)	71.04 (.25)	72.67 (.39)	71.85 (.18)	67.52 (.16)	64.57 (.23)	65.84 (.31)	67.99 (.23)
T-LSTM	65.93 (.16)	61.63 (.37)	64.76 (.48)	66.57 (.46)	67.65 (.29)	69.41 (.38)	66.49 (.37)	65.82 (.40)
StageNet	71.44 (.33)	71.92 (.28)	73.27 (.21)	70.39 (.16)	74.13 (.35)	72.39 (.22)	74.95 (.20)	75.63 (.32)
GraphSAGE	65.29 (.34)	63.72 (.15)	67.41 (.29)	69.04 (.35)	68.90 (.23)	66.42 (.38)	71.17 (.39)	72.05 (.26)
EvoNet	62.35 (.30)	59.57 (.33)	64.60 (.21)	66.07 (.19)	63.94 (.27)	65.12 (.44)	60.53 (.19)	60.75 (.46)
CTDNE	70.34 (.37)	68.29 (.22)	66.81 (.18)	69.86 (.30)	73.48 (.36)	71.55 (.13)	72.67 (.25)	71.59 (.26)
JODIE	73.25 (.41)	73.18 (.11)	67.03 (.17)	69.62 (.14)	73.91 (.36)	70.07 (.35)	68.41 (.15)	66.43 (.34)
TGAT	68.48 (.32)	70.76 (.23)	73.34 (.18)	72.88 (.42)	75.81 (.22)	72.59 (.43)	74.55 (.28)	73.66 (.26)
TGN	73.90 (.33)	74.06 (.46)	74.43 (.24)	74.14 (.28)	77.59 (.36)	76.28 (.26)	75.84 (.23)	74.13 (.39)
RxNet	**78.46** (.28)	**77.37** (.31)	**77.14** (.27)	**77.63** (.36)	**80.03** (.34)	**78.29** (.55)	**77.98** (.25)	**79.07** (.40)
	9 months				**12 months**			
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 4: Comparison of model variants' performances, in terms of PRAUC and F1, for various observation periods (3, 6, 9, 12 months). The numbers in parentheses are the corresponding standard deviations. The best results are highlighted in bold.

	Opioid Overdose		Drug Abuse		Opioid Overdose		Drug Abuse	
	PRAUC	F1	PRAUC	F1	PRAUC	F1	PRAUC	F1
	3 months				**6 months**			
RxNet-Edge	74.38 (.26)	76.73 (.17)	74.23 (.35)	74.36 (.28)	72.91 (.33)	74.33 (.27)	72.68 (.18)	74.18 (.41)
RxNet-Time	73.58 (.42)	75.39 (.33)	71.61 (.17)	73.15 (.31)	77.69 (.43)	77.52 (.27)	76.85 (.21)	77.93 (.24)
RxLSTM	69.46 (.47)	65.82 (.29)	66.14 (.42)	63.08 (.37)	72.68 (.19)	69.41 (.39)	68.53 (.35)	67.66 (.39)
RxNet-Con	75.74 (.34)	74.29 (.22)	73.11 (.39)	71.39 (.28)	78.39 (.17)	77.54 (.34)	74.14 (.26)	74.55 (.30)
RxNet-Rec	74.65 (.17)	73.29 (.29)	72.42 (.29)	73.34 (.25)	74.69 (.28)	75.87 (.36)	72.30 (.41)	71.57 (.24)
RxNet	**78.46** (.28)	**77.37** (.31)	**77.14** (.27)	**77.63** (.36)	**80.03** (.34)	**78.29** (.55)	**77.98** (.25)	**79.07** (.40)
	9 months				**12 months**			
RxNet-Edge	70.24 (.44)	68.94 (.43)	71.35 (.32)	72.18 (.23)	75.05 (.31)	74.81 (.15)	73.98 (.21)	75.28 (.15)
RxNet-Time	75.82 (.39)	74.01 (.28)	72.12 (.42)	73.55 (.35)	78.29 (.26)	77.29 (.18)	74.33 (.26)	73.15 (.27)
RxLSTM	67.87 (.38)	68.26 (.19)	70.25 (.21)	69.34 (.20)	70.46 (.19)	68.53 (.25)	67.68 (.21)	69.34 (.20)
RxNet-Con	78.17 (.23)	77.62 (.34)	75.70 (.34)	71.09 (.36)	80.52 (.27)	78.18 (.24)	73.85 (.24)	75.69 (.16)
RxNet-Rec	76.73 (.15)	75.31 (.23)	75.57 (.36)	75.07 (.13)	74.65 (.17)	73.29 (.29)	72.42 (.29)	73.34 (.25)
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)

(given by Eq. 6) 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. It clearly shows that the medical condition of the patient is at quite a 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.

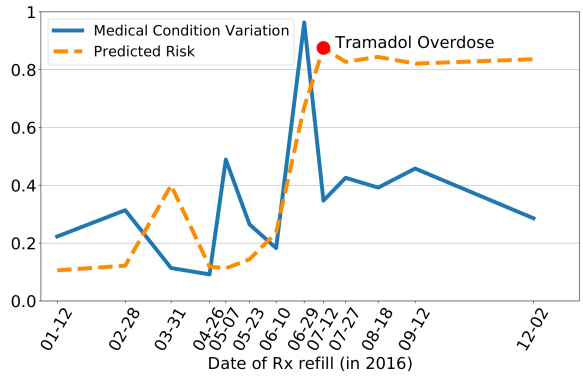


Figure 5: The medical condition variation and predicted risk of overprescribing Tramadol for a patient.

5.8 Medical Condition Variation Analysis

To understand the impact of different medical conditions on drug use, we analyze the medical condition stability of patients on every single Rx drug. RxNet outputs ϵ (in Eq. 6) that indicates a change of medical condition at a certain time. A large ϵ indicates the patient's current condition has changed a lot, compared to historical status. A patient with a stable condition will have a low ϵ. We compute the average of ϵ over all patients for each drug. Figure 6 presents these average values for some commonly prescribed drugs, from which we can see that Tramadol, Tapentadol, and Butorphanol Tartrate are more likely to be overprescribed than other drugs.

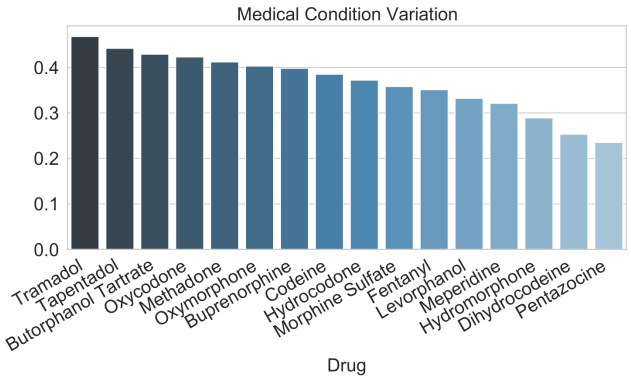


Figure 6: The average medical condition variation values over all patients for different drugs.

5.9 Overdose Hazard Analysis

To assess the model's ability for predicting overdose risk of patients, we apply the survival function [3] to transform the outcomes of RxNet to a series of monotonically increasing probabilities, which indicate the cumulative overdose hazard [5]. Figure 7 shows the comparison of the cumulative overdose hazard between a high-risk patient and a low-risk patient.

- The high-risk patient is a 62-year-old male patient, who fills his prescriptions for Morphine & Comb (an opioid) monthly. He has overdosed on Morphine & Comb in the 4th month.
- The low-risk patient is a 48-year-old female patient, who fills her prescription for Fentanyl (an opioid) monthly. She remained overdose-free throughout the year.

It can be seen from Figure 7 that RxNet clearly distinguishes between the two patients as early as at the beginning of the third month, after which the high-risk patient gets worse and worse while the low-risk patient has a low probability of overdose for a long time. The high-risk patient could thus be issued a warning, indicating that overdose would soon occur, and offered advice on early treatment or lifestyle change. Therefore, RxNet is a good aid to PDMP to help the early prognosis of overprescribing.

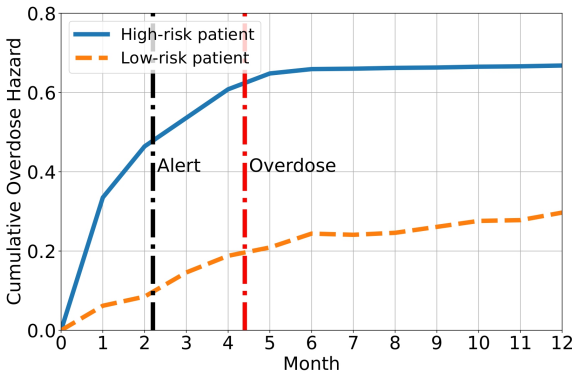


Figure 7: The 1-year cumulative overdose hazard predicted by RxNet for a high-risk (blue line) versus a low-risk (orange line) patient. The red dashed line is drawn at month 5, indicating the opioid overdose time of the high-risk patient. The black dash indicates the time when an alert can be issued.

6 CONCLUSIONS

In this paper, we proposed to detect the Rx-refill caused overprescribing, i.e., predicting patients at high risk of opioid overdose or drug abuse. To solve the problem, we developed a novel model called RxNet. We constructed a P&D graph and employed an self-attention based GNN to learn patient embeddings. To capture medical condition variations, we incorporated a newly designed RxLSTM to update dynamic patient embeddings. Moreover, we introduced a dosing-adaptive network to explore and recalibrate the dosing patterns of patients. The empirical results on a 1-year PDMP data demonstrated the effectiveness of RxNet by comparing with state-of-the-art methods, and revealed the promise of RxNet in PDMP.

7 ACKNOWLEDGMENTS

This work is partially supported by the National Science Foundation (NSF) under grants IIS-2107172, IIS-2140785, IIS-2027127, IIS-2040144, IIS-1951504, CNS-1940859, CNS-1814825, and OAC-1940855, the National Institute of Justice (NIJ) 2018-75-CX-0032.

REFERENCES

[1] George Adam, Ladislav Rampášek, Zhaleh Safikhani, Petr Smirnov, Benjamin Haibe-Kains, and Anna Goldenberg. 2020. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ precision oncology* 4, 1 (2020), 1–10.

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

[3] Alexis Bellot and Mihaela van der Schaar. 2018. Multitask Boosting for Survival Analysis with Competing Risks. In *NIPS*. 1397–1406.

[4] Cary J Blum, Lewis S Nelson, and Robert S Hoffman. 2016. A survey of physicians' perspectives on the New York state mandatory prescription monitoring program (ISTOP). *Journal of Substance Abuse Treatment* 70 (2016), 35–43.

[5] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In *KDD*. 1721–1730.

[6] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*. 135–144.

[7] Lun Du, Yun Wang, Guojie Song, Zhicong Lu, and Junshan Wang. 2018. Dynamic Network Embedding : An Extended Approach for Skip-gram based Network Embedding. In *IJCAI*. 2086–2092.

[8] Ali Mert Ertugrul, Yu-Ru Lin, and Tugba Taskaya-Temizel. 2019. CASTNet: community-attentive spatio-temporal networks for opioid overdose forecasting. In *ECML-PKDD*. Springer, 432–448.

[9] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *CIKM*. 1797–1806.

[10] Jeffrey Fudin, Mena Raouf, Erica L Wegrzyn, and Michael E Schatman. 2018. Safety concerns with the Centers for Disease Control opioid calculator. *Journal of pain research* 11 (2018), 1.

[11] Neha S Gangal, Ana L Hincapie, Roman Jandarov, Stacey M Frede, Jill M Boone, Neil J MacKinnon, Kathleen Koechlin, Jolene DeFiore-Hyrmer, Amy Holthusen, and Pamela C Heaton. 2020. Association Between a State Law Allowing Pharmacists to Dispense Naloxone Without a Prescription and Naloxone Dispensing Rates. *JAMA Network Open* 3, 1 (2020), e1920310–e1920310.

[12] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, and Jimeng Sun. 2020. StageNet: Stage-Aware Neural Networks for Health Risk Prediction. In *WWW*. 530–540.

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

[14] Haifan Gong, Chaoqin Qian, Yue Wang, Jianfeng Yang, Sheng Yi, and Zichen Xu. 2019. Opioid Abuse Prediction Based on Multi-Output Support Vector Regression. In *ICMLT*. 36–41.

[15] Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. 2018. DynGEM: Deep Embedding Method for Dynamic Graphs. *arXiv preprint arXiv:1805.11273* (2018).

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

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

[18] Anders Håkansson and Virginia Jesionowska. 2018. Associations between substance use and type of crime in prisoners with substance use problems–a focus on violence and fatal violence. *Substance abuse and rehabilitation* 9 (2018), 1.

[19] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS*. 1024–1034.

[20] Justine S Hastings, Mark Howison, and Sarah E Inman. 2020. Predicting high-risk opioid prescriptions before they are given. *Proceedings of the National Academy of Sciences* 117, 4 (2020), 1917–1923.

[21] Holly Hedegaard, Arialdi M. Miniño, and Margaret Warner. 2020. Drug Overdose Deaths in the United States, 1999–2018. *National Center for Health Statistics* (2020).

[22] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comp.* 9, 8 (1997), 1735–1780.

[23] Han Hu, NhatHai Phan, James Geller, Huy T. Vo, Manasi Bhole, Xueqi Huang, Sophie Di Lorio, Thang Dinh, and Soon Ae Chun. 2018. Deep Self-Taught Learning for Detecting Drug Abuse Risk Behavior in Tweets. In *CSoNet*, Vol. 11280. 330–342.

[24] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-Excitation Networks. In *CVPR*. 7132–7141.

[25] M Mofizul Islam. 2019. 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).

[26] Shipra Jain, Prerna Upadhyaya, Jaswant Goyal, Abhijit Kumar, Pushpawati Jain, Vikas Seth, and Vijay V Moghe. 2015. A systematic review of prescription pattern monitoring studies and their effectiveness in promoting rational use of medicines. *Perspectives in Clinical Research* 6, 2 (2015), 86.

[27] Alene Kennedy-Hendricks, Matthew Richey, Emma E McGinty, Elizabeth A Stuart, Colleen L Barry, and Daniel W Webster. 2016. Opioid overdose deaths and Florida's crackdown on pill mills. *American journal of public health* 106, 2 (2016), 291–297.

[28] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

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

[30] Wei-Hsuan Lo-Ciganic, James L Huang, Hao H Zhang, Jeremy C Weiss, Yonghui Wu, C Kent Kwoh, Julie M Donohue, Gerald Cochran, Adam J Gordon, Daniel C Malone, et al. 2019. Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Network Open* 2, 3 (03 2019), e190968–e190968.

[31] John Lu, Sumati Sridhar, Ritika Pandey, Mohammad Al Hasan, and George Mohler. 2019. Investigate Transitions into Drug Addiction through Text Mining of Reddit Data. In *KDD*. 2367–2375.

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

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

[34] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B Schardl, and Charles E Leiserson. 2020. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs.. In *AAAI*. 5363–5370.

[35] Leonard J Paulozzi, Gail K Strickler, Peter W Kreiner, and Caitlin M Koris. 2015. Controlled substance prescribing patterns – prescription behavior surveillance system, eight states, 2013. *Morbidity and Mortality Weekly Report: Surveillance Summaries* 64, 9 (2015), 1–14.

[36] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael M. Bronstein. 2020. Temporal Graph Networks for Deep Learning on Dynamic Graphs. In *ICML*.

[37] Daniel J Safer. 2019. Overprescribed medications for US adults: four major examples. *Journal of Clinical Medicine Research* 11, 9 (2019), 617.

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

[39] Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron C. Courville. 2019. Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks. In *ICLR*.

[40] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB Endowment* 4, 11 (2011), 992–1003.

[41] Muthiah Vaduganathan, Jeroen van Meijgaard, Mandeep R Mehra, Jacob Joseph, Christopher J O'Donnell, and Haider J Warraich. 2020. Prescription Fill Patterns for Commonly Used Drugs During the COVID-19 Pandemic in the United States. *JAMA* (2020).

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

[43] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S. Yu. 2019. Heterogeneous Graph Attention Network. In *WWW*. 2022–2032.

[44] Matthew J Witry, Barbara J St Marie, Brahmendra Reddy Viyyuri, and Paul D Windschitl. 2020. Factors Influencing Judgments to Consult Prescription Monitoring Programs: A Factorial Survey Experiment. *Pain Management Nursing* 21, 1 (2020), 48–56.

[45] Changmin Wu, Giannis Nikolentzos, and Michalis Vazirgiannis. 2020. EvoNet: A Neural Network for Predicting the Evolution of Dynamic Graphs. *arXiv* (2020).

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

[47] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *ICLR*.

[48] Zhuo Yang, Barth Wilsey, Michele Bohm, Meghan Weyrich, Kakoli Roy, Dominique Ritley, Christopher Jones, and Joy Melnikow. 2015. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *The Journal of Pain* 16, 5 (2015), 445–453.

[49] Wenchao Yu, Wei Cheng, Charu C. Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. 2018. NetWalk: A Flexible Deep Embedding Approach for Anomaly Detection in Dynamic Networks. In *KDD*. 2672–2681.

[50] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. 2019. Graph Transformer Networks. In *NIPS*. 11960–11970.

[51] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *KDD*. 793–803.

[52] Ziwei Zhang, Peng Cui, Jian Pei, Xiao Wang, and Wenwu Zhu. 2018. TIMERS: Error-Bounded SVD Restart on Dynamic Networks. In *AAAI*. 224–231.