

Original Paper

# A Multiview Model for Detecting the Inappropriate Use of Prescription Medication: Machine Learning Approach

Lin Zhuo<sup>1,2\*</sup>, PhD; Yinchu Cheng<sup>3\*</sup>, PhD; Shaoqin Liu<sup>4</sup>, BA; Yu Yang<sup>5</sup>, PhD; Shuang Tang<sup>4</sup>, BA; Jiancun Zhen<sup>6</sup>, MS; Junfeng Zhao<sup>4</sup>, PhD; Siyan Zhan<sup>1,2</sup>, MD, PhD

<sup>1</sup>Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

<sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

<sup>3</sup>Department of Pharmacy, Peking University Third Hospital, Beijing, China

<sup>4</sup>School of Electronics Engineering and Computer Science, Peking University, Beijing, China

<sup>5</sup>Center for Data Science in Medicine and Health, Peking University, Beijing, China

<sup>6</sup>Department of Pharmacy, Ji Shui Tan Hospital and Fourth Medical College of Peking University, Beijing, China

\*these authors contributed equally

**Corresponding Author:**

Siyan Zhan, MD, PhD

Research Center of Clinical Epidemiology

Peking University Third Hospital

49 North Garden Rd, Haidian District

Beijing, 100191

China

Phone: 86 1082805162

Email: [siyan-zhan@bjmu.edu.cn](mailto:siyan-zhan@bjmu.edu.cn)

## Abstract

**Background:** The inappropriate use of prescription medication has recently garnered worldwide attention, but most national policies do not effectively provide for early detection or timely intervention.

**Objective:** This study aimed to develop and assess the validity of a model that can detect the inappropriate use of prescription medication. This effort combines a multiview and topic matching method. The study also assessed the validity of this approach.

**Methods:** A multiview extension of the latent Dirichlet allocation algorithm for topic modeling was chosen to generate diagnosis-medication topics, with data obtained from the Chinese Monitoring Network for Rational Use of Drugs (CMNRUD) database. Topic mapping allowed for calculating the degree to which diagnoses and medications were similarly distributed and, by setting a threshold, for identifying prescription misuse. The Beijing Regional Prescription Review Database (BRPRD) database was used as the gold standard to assess the model's validity. We also conducted a sensitivity analysis using random samples of validated prescriptions and evaluated the model's performance.

**Results:** A total of 44 million prescriptions were used to generate topics using the diagnoses and medications from the CMNRUD database. A random sample (15,000 prescriptions) from the BRPRD was used for validation, and it was found that the model had a sensitivity of 81.8%, specificity of 47.4%, positive-predictive value of 14.5%, and negative-predictive value of 96.0%. The model showed superior stability under different sampling proportions.

**Conclusions:** A method that combines multiview topic modeling and topic matching can detect the inappropriate use of prescription medication. This model, which has mediocre specificity and moderate sensitivity, can be used as a primary screening tool and will likely complement and improve the process of manually reviewing prescriptions.

(*JMIR Med Inform* 2020;8(7):e16312) doi: [10.2196/16312](https://doi.org/10.2196/16312)

**KEYWORDS**

inappropriate use of prescription medication; topic model; latent Dirichlet allocation; multiview learning; prescription review

## Introduction

It is estimated that more than 50% of medicines are inappropriately prescribed, dispensed, or sold, which represents a universal challenge for medical practice [1]. Furthermore, in developing countries, the treatment of about 60% to 70% of patients in primary care does not meet standard treatment guidelines [2]. This inappropriate use of prescription is wasteful and costly, and can increase the risk of adverse drug reactions [3]. Finally, the overuse of antimicrobial and antibiotic injections may result in certain pathogens developing antibiotic resistance [4].

The excessive use of antibiotics is common in China, as is the injection of traditional Chinese medicines [5-7]. Antibiotics are present in 50% of prescriptions and injectable medicines in 30%, exceeding the World Health Organization's standard treatment guidelines [7]. The Chinese government released the *Management Practices of Hospital Prescription Comment (Trial)* in 2010 to assess compliance with rational criteria for using prescription drugs [8]. This document requires that each hospital assign trained pharmacists monthly to review a minimum of 100 randomly sampled prescriptions. However, these reviews are currently associated with limited coverage, high omission rates, a lack of representativeness, and supervisory lag. All of this points to the urgency of improving the review process, especially when Chinese hospitals are witnessing a continuous daily increase in outpatient prescriptions [9,10].

A few knowledge-based approaches have been implemented in the health care information systems (HISs) of Chinese hospitals to screen the appropriateness of prescriptions [11,12]. Prior knowledge, including treatment guidelines, formularies, package inserts, expert knowledge, and published literature, indicates that these systems are generally working well. However, they are time-consuming and costly to establish and maintain [13]. Furthermore, timely updates to these systems are challenging because of the continuous availability of both new drugs and new research.

Currently, both supervised and unsupervised data-driven methodologies, whether as alternatives or supplements to the systems listed above, are being used to identify outliers and detect inappropriate prescriptions. Such approaches remain constrained, however, by the difficulty associated with using supervised methods to obtain high-quality labeled sample data [14,15]. Other limiting factors include defining outliers and considering their association rules, which effectively account for the relationships between features [16]. Usually, diagnosis and medication are closely and consistently related to the clinical condition of the patient. In the absence of this consistency, prescriptions are more likely to be inappropriate (or anomalous). Contextual anomaly detection is one approach to capture the relationship between features (eg, between medication and diagnosis) and to detect exceptions caused by feature mismatch [17,18]. Nonetheless, this does not work well with prescription data or similar information in a high-dimensional sparse space [19].

By contrast, a topic modeling method, the latent Dirichlet allocation (LDA) method [20], has been proven to be useful in dimensional reduction when mining patient records [21]. Here, a "topic" is defined as a collection of semantically related terms that appear frequently and relate to a common subject [22]. LDA, a probabilistic statistical model with the assumption that topic distributions are drawn from their prior distributions, can be used to describe the composition of high-dimensional unstructured text and to capture clusters of words that reveal critical concepts [21,23]. Beginning with its appearance in the biomedical domain, LDA has been used in mining clinical pathway patterns [24-26], image processing [27-29], risk stratification [30], and bioinformatics [31-33]. One drawback of LDA is that it cannot simultaneously consider both diagnosis and medication. We therefore adopted a multiview [23,34-36] concept that enhances the topic modeling capacity of LDA and coordinated it with anomaly detection techniques to build a multiview LDA model (MV-LDA). This model, which was tested in our previous simulation study, had a greater area under the precision-recall curve [37] than the two traditional methods (point anomaly detection and contextual anomaly detection) and had better suitability for high-dimensional sparse data.

## Methods

### Data Sources

One subsystem of the Chinese Monitoring Network for the Rational Use of Drugs (CMNRUD) was the data source for model development. The CMNRUD was launched by the Chinese Ministry of Health in 2010, and it covers over 86% (30/35) of the provinces in China [38], including 60% of the nation's tertiary hospitals (955 hospitals) and 6% of its secondary hospitals (375 hospitals) [39]. Each monitoring hospital must upload encrypted data every month. The system organizes the prescriptions in a stipulated uniform structure, and thereafter, some of the cleaned data are checked by data management professionals. Since 2013, the system has been performing automatic uploading, preliminary cleaning, recoding, and verification.

The CMNRUD consists of the following four monitoring subsystems: outpatient prescriptions, clinical drug use, medical damage, and critical disease. Anonymized data from outpatient prescriptions (from October to December 2016) were used to build the model for detecting prescription misuse. These data include demographic, diagnostic, and drug-related information. Diagnostic information includes the patient ID, diagnosis date, diagnosis description, and diagnostic code (10th revision of the International Classification of Diseases, ICD-10), which are directly related to the purpose of the patient's visit or their condition. It sometimes may not include other complications that do not require further treatment. A higher diagnosis ranking was associated with more visit relevance. Available drug-related information (no more than five medications per prescription) includes information such as the prescription date, generic and brand names, corresponding Anatomical Therapeutic Chemical code, dosage, and administration route. The medications were listed randomly, preventing them from being mapped to the

corresponding diagnoses on a one-to-one basis. The variables taken from the CMNRUD are presented in Table 1.

The data for model validation were randomly selected from the Beijing Regional Prescription Review Database (BRPRD) [40], which was created by the Beijing Municipal Administration of Hospitals in 2010. The BRPRD extracts 1 week of prescriptions every quarter from the HISs of 17 tertiary hospitals and five secondary hospitals in Beijing. A total of 19 hospitals from the BRPRD were included among the 65 CMNRUD monitoring

hospitals from Beijing and accounted for 5.4% (19/349) of the hospitals in the entire CMNRUD database. The prescription variables include treatment type, prescription number, prescription date, age, sex, diagnosis, and medication. As part of the standard procedure, a prescription review board of trained clinicians and clinical pharmacists regularly examines the prescriptions individually based on a standardized guideline and then captures inappropriate data in the BRPRD database [8,41].

**Table 1.** Main variables in the Chinese Monitoring Network for the Rational Use of Drugs outpatient prescription monitoring subsystem.

Information	Variables <sup>a</sup>
Basic information (patients)	Patient ID, treatment card number, sex, age, and age range
Basic information (hospital, department, and doctor)	Hospital name, hospital grade, hospital type, region, department, and doctor ID
Diagnosis	Diagnosis name, and ICD-10 <sup>b</sup> by class, suborder, and type
Medication <sup>c</sup>	Prescription ID, prescription type, prescription date, ATC <sup>d</sup> code, drug trade name, drug generic name, specification, quantity, unit, dosage, usage, price, pharmaceutical company, and individual hospital information

<sup>a</sup>The variables indicate the features of the multiview latent Dirichlet allocation model.

<sup>b</sup>ICD-10: 10th revision of the International Classification of Diseases.

<sup>c</sup>Since a generic medicine works the same as its branded version and owing to the limitation of the Anatomical Therapeutic Chemical's lack of codes for traditional Chinese medicine, we used generic names, which are well recorded in the database, to build the topic model.

<sup>d</sup>ATC: Anatomical Therapeutic Chemical.

## Study Approval

The Institutional Review Board of Peking University reviewed and approved the study protocol before the study commenced, and it determined that informed consent was not required (reference number: IRB00001052-17003-Exempt).

## Study Design

We developed and evaluated an MV-LDA model for detecting the inappropriate use of prescription medications in the following four steps: (1) data preparation, gathering and cleaning data from the CMNRUD and BRPRD database; (2) topic generation, using MV-LDA topic-modeling methods and CMNRUD data to extract associations between diagnoses and medications; (3) inferring and anomaly scoring, using BRPRD data and the topics extracted in step 2 to infer the distribution of each prescription and measuring the degree to which diagnoses and medications show similar distributions (less similarity is associated with more likelihood that the item represents the inappropriate use of prescription medication); (4) model evaluation and sensitivity analysis, evaluating the model for detecting prescription misuse with the results of the BRPRD review.

### Step 1: Data Preparation

Prescriptions between October 2016 and December 2016 from CMNRUD were used, and those with missing prescription identifiers or with medication withdrawal were excluded. The patient ID, treatment card number, prescription date, diagnosis name, and generic drug name were chosen to build the topics.

For model evaluation, considering the sensitivity and specificity (71.5% and 68.8%, respectively) of the Apriori algorithm in previous work [42,43], we set both the expected sensitivity and specificity at 80%. By setting a significance level of .05 and an allowable error of 0.05, we needed at least 14,471 prescriptions given 1.7% prevalence [44] of prescription misuse according to equation 1.

$$n = \left[ \frac{57.3 \times Z_{1-\frac{\alpha}{2}}}{\arcsin\left(\frac{\delta}{\sqrt{p(1-p)}}\right)} \right]^2 \quad (1)$$

Finally, we randomly selected a total of 15,000 prescriptions from 2016 BRPRD data that had already been manually reviewed by experienced pharmacists, with the prescriptions that included the following three variables: prescription ID, diagnosis, and medication.

### Step 2: Topic Generation

The study is based on the assumption that the prescription database is mostly composed of regular instances (ie, rational appropriate prescriptions), and a probabilistic model is fitted to all features.

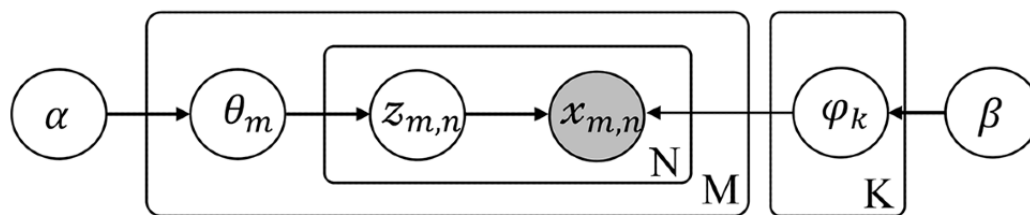
LDA assumes that a set of documents or instances exhibits a specific number of latent independent topics, and then, the given topics generate the terms probabilistically. A graphical representation of the LDA model is given in Figure 1. With specific input records and setting hyperparameters  $\alpha$  and  $\beta$ , LDA can detect  $K$  topics, formally presented as two multinomial

distributions (topic-word distribution  $\phi$  and document-topic distribution  $\theta$ ). The LDA model formula is shown in equation 2.

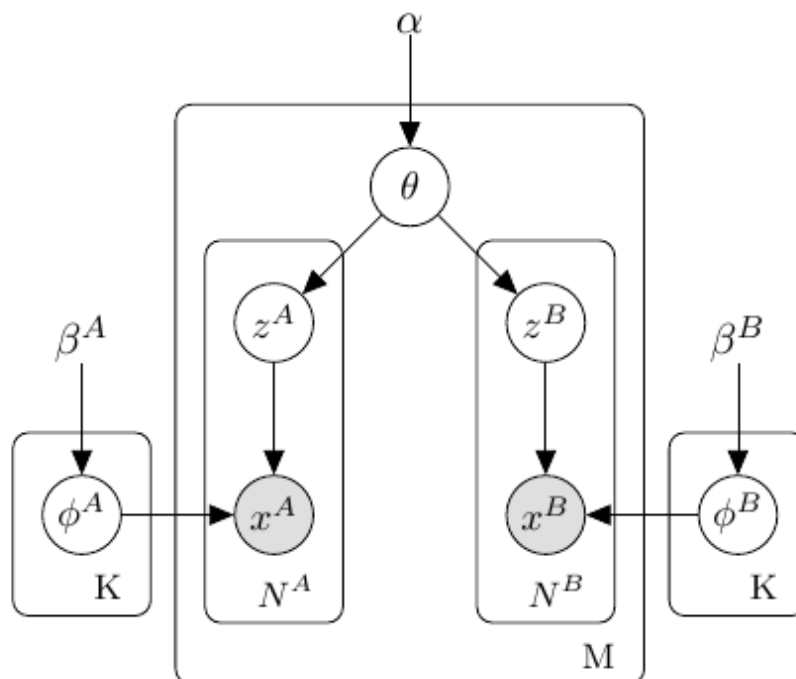
$$P(x, z, \theta, \phi) = \prod_{k=1}^K P(\phi_k | \beta) \prod_{m=1}^M P(\theta_m | \alpha) \prod_{n=1}^N P(z_{m,n} | \theta_m) P(x_{m,n} | \phi_{z_{m,n}}) \quad (2)$$

Like the standard LDA, MV-LDA can be represented as a probability pattern, as shown in Figure 2. In the medical domain, we can consider a clinical condition (topic) as a probability distribution over related diagnostic codes, and a patient's diagnostic record can be regarded as a "document" composed of different clinical topics. The same applies to medication. In our study, we used CMNRUD data to generate MV-LDA topics.

**Figure 1.** Graphical representation of the latent Dirichlet allocation model. K: number of topics; M: number of documents; N: number of words in each document; x: observed words in the document m; z: topic of nth word in a document m;  $\theta$ : topic distribution for document m (document-topic distribution);  $\phi$ : topic-word distribution;  $\alpha$ : hyperparameter of  $\theta$ ;  $\beta$ : hyperparameter of  $\phi$ .



**Figure 2.** Graphical representation of the multiview latent Dirichlet allocation model. K: number of topics; M: number of prescriptions;  $N^A$ : number of diagnoses per prescription;  $N^B$ : number of medications per prescription;  $x^A$ : diagnosis (type A feature);  $x^B$ : medication (type B feature);  $z^A$ : topic of  $x^A$ ;  $z^B$ : topic of  $x^B$ ;  $\phi^A$ : topic-diagnosis distribution;  $\phi^B$ : topic-medication distribution;  $\beta^A$ : hyperparameter of  $\phi^A$ ;  $\beta^B$ : hyperparameter of  $\phi^B$ ;  $\theta$ : prescription-topic distribution;  $\alpha$ : hyperparameter of  $\theta$ .



The MV-LDA model generates topics as follows: (1) For each topic, draw features  $\phi^A \sim \text{Dirichlet}(\beta^A)$  and draw features  $\phi^B \sim \text{Dirichlet}(\beta^B)$ ; (2) For each prescription, draw topic proportions  $\theta \sim \text{Dirichlet}(\alpha)$ ; for each feature A, draw  $z_{m,n} \sim \text{Mult}(\theta_m)$  and draw  $x_{m,n} \sim \text{Mult}(\phi_{z^A}^A)$ ; and for each feature B, repeat the steps for feature A.

We adopted Gibbs sampling to create the model and parameters  $\phi$  and  $\theta$ . Topics were first randomly assigned to all of the features. Every diagnosis (type A feature) or medication (type B feature) in each prescription that corresponds to a topic is iteratively sampled. The calculation of the conditional probability of  $x^A$  is shown in equation 3, with the related

notations shown in Table 2. Here, the first factor of equation 3 only accounts for the type A feature topic-diagnosis counts, whereas the second factor calculates the prescription-topic for all features. This formula also applies to type B features.

The features of both types (A and B) were iteratively sampled for each prescription until the model converged. Thereafter, we utilized the result to calculate the parameters  $\phi^A$  and  $\phi^B$ , which are used for the inferring step, and  $\phi^A$  can be calculated as in equation 4. We set eight topic numbers (15, 20, 25, 30, 35, 40, 45, and 50) and built the MV-LDA model according to previous research and a pilot study revealing that LDA had a moderate

ability in terms of generating topics for electronic medical records with a topic number of around 30 [45].

$$P^A(z = k' | z_{-i}, x) \propto \frac{C_{x^A, k}^{VAK} + \beta^A}{\sum_{x=1}^V C_{x^A, k}^{VAK} + V^A \beta^A} \cdot \frac{C_{m, k}^{MK} + \alpha}{\sum_{x=1}^K C_{m, x}^{MK} + K \alpha} \tag{3}$$

$$\phi_{x^A, k}^A = \frac{C_{x^A, k}^{VAK} + \beta^A}{C_{\cdot, k}^{VAK} + V^A \beta^A} \tag{4}$$

**Table 2.** Notations of the multiview latent Dirichlet allocation model in Gibbs sampling.

Variable	Description
$K$	Topic number
$V^A$	Number of diagnoses (type A feature)
$C_{x^A, k}^{VAK}$	Number of times that $x^A$ is assigned to topic $k$
$C_{\cdot, k}^{VAK}$	Number of times that any feature A is assigned to topic $k$
$C_{m, k}^{MK}$	Number of all features of prescription $m$ (including both A and B) assigned to topic $k$
$C_{m, \cdot}^{MK}$	All features in prescription $m$

**Step 3: Inferring and Anomaly Scoring**

In this step, a separate dataset was used for inferring 15,000 randomly sampled prescriptions from the BRPRD (2016) that had already been manually reviewed by experienced pharmacists. Each feature in the MV-LDA model can be treated as an independent LDA model and can be inferred separately. To be specific, for the MV-LDA model obtained in the previous learning step,  $\phi^A$  can be used to detect the new prescriptions in question, but this only pertains to estimations of the topic distribution under feature A. The equation for this is as follows:

$$P^A(z = k' | z_{-i}, x) \propto \phi_{x, k}^A \cdot \frac{C_{m, k}^{MAK} + \alpha}{C_{m, \cdot}^{MAK} + K \alpha} \tag{5}$$

In this equation,  $\phi_{x, k}^A$  is the value of the topic distribution under the circumstance of topic  $k$  and feature  $x$ . Finally, as in the topic generating step (step 2), we inferred the marginal  $\theta$  based on the Gibbs sampling shown in equation 5, which indicated the proportion of feature A assigned to topic  $k$  for each prescription. Additionally,  $\hat{\theta}^B$  is calculated under type B features. For each test prescription, both  $\hat{\theta}^A$  and  $\hat{\theta}^B$  were inferred and used to calculate the anomaly score.

$$\hat{\theta}^A = \frac{C_{m, k}^{MAK} + \alpha}{C_{m, \cdot}^{MAK} + K \alpha} \tag{6}$$

The assumption mentioned in step 2 is that the given order of diagnoses and prescribed medication should show consistency (ie, the values for  $\hat{\theta}^A$  and  $\hat{\theta}^B$  should be equal or close to each other), and if not, the prescription might be inappropriate. The similarity between  $\hat{\theta}^A$  and  $\hat{\theta}^B$  was measured using novel topic mapping (TM) methods [46]. TM was performed in the following manner: we allocated topic feature distributions from the MV-LDA model for every diagnosis or medication before matching. First, high probability topics were tagged for each diagnosis. Thereafter, we similarly identified the most probable topics for each medication and added up the total. When a topic was not tagged, it was assigned an anomaly score of 1. Finally, the anomaly scores for each prescription were summed, and different thresholds were used to filter potentially inappropriate prescriptions.

**Step 4: Model Evaluation and Sensitivity Analysis**

The same prescriptions (15,000 randomly sampled prescriptions from the BRPRD in 2016) were inferred and detected by the MV-LDA model. Multimedia Appendix 1 shows the confusion matrix of the screening test we used. The sensitivity, specificity, positive-predictive value (PPV), negative-predictive value (NPV), and Youden’s index were computed from the results to compare the assessments between the model and the experts and to identify the best performance parameter setting of TM. A sensitivity analysis was performed by randomly sampling 90%, 70%, 50%, 30%, and 10% of prescriptions from the evaluation data of the 15,000 prescriptions. The sensitivity, specificity, PPV, NPV, Youden’s index, and area under the receiver operating characteristic curve were compared. It should



be noted that the overlap between training data and evaluation data was small enough to be ignored.

## Results

### Prescriptions

A total of 44,325,065 prescriptions from 22 million patients (138,535,092 records) at 349 hospitals, including 286 tertiary and 63 secondary hospitals, were used in our topic modeling process. This included 5,653 types of medications and 22,643 diseases or conditions. In the validation dataset, there were 14,166 (94.4%) outpatient prescriptions and 834 (5.6%) emergency prescriptions. Of these, 13,524 (90.2%) prescriptions satisfied the appropriate criteria (marked as “appropriate”) and 1476 (9.8%) failed (marked as “inappropriate”).

### Multiview Latent Dirichlet Allocation Topic Generation Results

By setting the topic parameters, we obtained eight topic models, all with commonly diagnosed diseases in clinical practice. For example, the model ( $K=30$ ) included cardiovascular diseases, diabetes, chronic nephrosis, osteoporosis, and some respiratory infections. Regarding topic 27, hypertension had a 93.3% probability of appearing in this topic, and amlodipine, nifedipine, levamlodipine, and metoprolol had probabilities of 11.7%, 8.9%, 7.7%, and 6.1%, respectively. The top probability diagnoses in topic 23 were bronchitis, pneumonia, and bronchopneumonia, with the proportions of 55.6%, 21.8%, and 12.1%, respectively, whereas the corresponding medications were ambroxol (11.2%), budesonide (10.7%), azithromycin (9.9%), and terbutaline (6.2%). We also obtained topics related to gastrointestinal diseases and mental and dermal disorders. The details pertaining to the top 10 topics and their allocations are shown in [Multimedia Appendix 2](#).

After comparing the training results of the topic models with settings at  $K=15, 20, 25, 30, 35, 40, 45,$  and  $50$  for the training results, it was found that a smaller topic number was associated

with a weaker relation between the topics on one side and diagnoses and medications on the other, which were likely to appear more dispersed and had a lower probability of appearing in a topic. As the set value of the number of topics increased, the ability to summarize the disease was enhanced, that is, the subject-feature distribution of topic learning became more concentrated, the feature became more likely to appear in the topic, and the proportions of diagnosis and medication tended to be uniform.

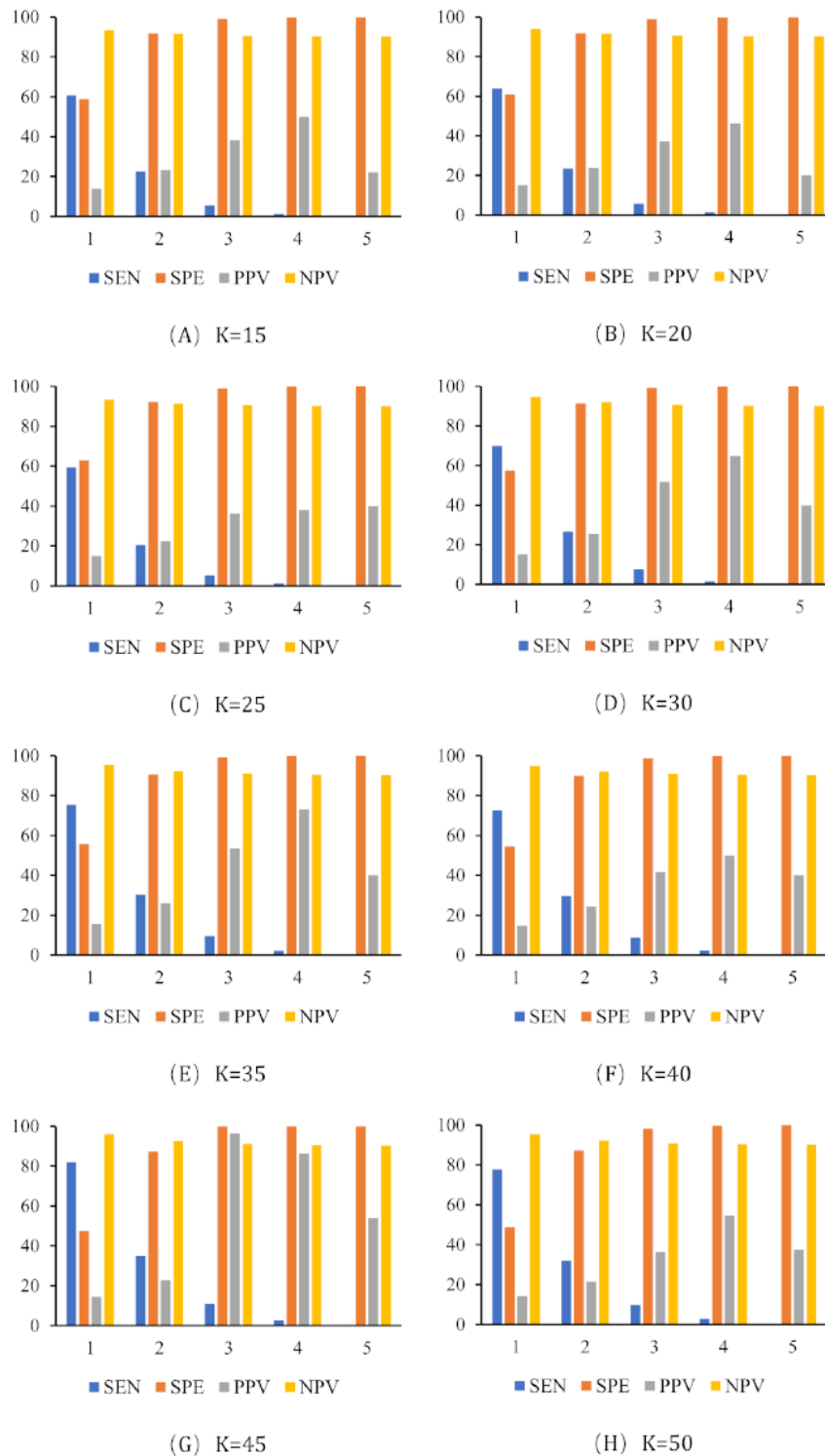
### Multiview Latent Dirichlet Allocation Evaluation Sensitivity Analysis Results

The BRPRD sample data evaluated the MV-LDA model. The performance of the MV-LDA model is shown in [Figure 3](#). Each model showed higher specificity and NPV for some topics, with the NPV reaching more than 90%, and the sensitivity being the highest at a TM threshold of 1. As the threshold value declined, the sensitivity decreased, the specificity and PPV increased, and the NPV showed no relevant change. When the number of topics increased, the sensitivity increased greatly, but the specificity, PPV, and NPV changed little.

Taking all factors (sensitivity, specificity, PPV, NPV, and Youden’s index) into consideration, we set a cutoff of  $\geq 1$  TM anomaly scoring as the threshold for our MV-LDA detection model. The results showed a high sensitivity of 81.8% and a moderate specificity of 47.4%, and the PPV and NPV were 14.5% and 96.0%, respectively. These findings indicate that under the best performance parameter setting, we can find 1208 of 1476 inappropriate prescriptions.

Our model evaluation results revealed that the MV-LDA model had a better ability to detect inappropriate prescriptions when the TM threshold was set to 1. However, for a better understanding of the robustness of the results at this parameter setting, we performed a sensitivity analysis, repeating the experiments with separate sampling proportions of 90%, 70%, 50%, 30%, and 10%. [Table 3](#) presents the findings. There were no relevant differences between the two experiments.

**Figure 3.** Summary of the performance of multiview latent Dirichlet allocation model with TM detection methods under different thresholds. Horizontal axis: thresholds of TM methods (from 1 to 5). Vertical axis: percentage of SEN, SPE, PPV, and NPV. K: number of topics; NPV: negative-predictive value; PPV: positive-predictive value; SEN: sensitivity; SPE: specificity; TM: topic mapping.



**Table 3.** Sensitivity analysis for multiview latent Dirichlet allocation with topic mapping detection methods (threshold=1).

Sampling proportion	TP <sup>a</sup> , n (%)	FP <sup>b</sup> , n (%)	FN <sup>c</sup> , n (%)	TN <sup>d</sup> , n (%)	SEN <sup>e</sup> (%)	SPE <sup>f</sup> (%)	PPV <sup>g</sup> (%)	NPV <sup>h</sup> (%)	Youden's index	AUROC <sup>i</sup>
90%	1073 (7.9)	5752 (42.6)	249 (1.8)	6435 (47.6)	81.2	47.2	14.3	95.9	14.3	0.689
70%	823 (7.9)	4446 (42.7)	179 (1.7)	4954 (47.6)	82.1	47.3	14.2	96.1	14.2	0.695
50%	570 (7.6)	3235 (43.1)	143 (1.9)	3562 (47.4)	79.9	47.6	13.8	95.8	13.8	0.686
30%	383 (8.6)	1958 (43.9)	93 (2.1)	2030 (45.5)	80.5	49.1	15.9	95.5	15.9	0.705
10%	109 (7.3)	622 (41.9)	21 (1.4)	734 (49.4)	83.8	45.9	12.9	96.7	12.9	0.693

<sup>a</sup>TP: true positive.

<sup>b</sup>FP: false positive.

<sup>c</sup>FN: false negative.

<sup>d</sup>TN: true negative.

<sup>e</sup>SEN: sensitivity.

<sup>f</sup>SPE: specificity.

<sup>g</sup>PPV: positive-predictive value.

<sup>h</sup>NPV: negative-predictive value.

<sup>i</sup>AUROC: area under the receiver operating characteristic curve.

## Discussion

### Principal Findings

The study drew upon the data of almost 45 million prescriptions obtained from the CMNRUD database (between October and December 2016). It then used the MV-LDA combination of TM anomaly detection and LDA topic modeling to build a model for detecting the inappropriate use of prescription medication. The model had a sensitivity of 81.8% and a specificity of 47.4% with 45 topics, and it had an anomaly threshold of 1 and showed stability in the sensitivity analysis. The topics that were already built into our study included most disorders, and the topics that were generated included noncommunicable diseases, such as cardiovascular diseases, which appeared in the largest proportions, consistent with clinical practice. The model also accommodated the tendency for many disorders to be seen in winter. For instance, upper respiratory infections, fever, and acute bronchitis were determined to be highly probable in winter. It took only 3.5 hours to generate the topics and only seconds to detect an anomaly, much quicker than the system of manual review or knowledge-based prescription review.

### Limitations

The present work has several limitations. The MV-LDA model has more features than were used in this study, including usage, dosage, cost, and even laboratory test results, when available. Moreover, it was challenging to clean accessional variables from the data source. For example, there are multiple modes of recording dose packaging because the composition and dosage forms of medicines differ from each other. Because of the difficulties noted above, the first limitation is that our method ignored the medication's usage and dosage and only addressed the medication itself when building the MV-LDA model and validating the results. Second, the current model is still not supported for indicating the specific medication but tells us which prescriptions do not comply with the prescription review criteria. Besides, limited by the failure to obtain a labeled

training dataset, the current model is not able to classify prescription misuse by criteria, such as the absence of proper indications, violation of clinical guidelines, and misuse of dosage, and can only detect the appropriateness of prescriptions. This study is also limited by the diverse structure of the model training and evaluation database and a minor overlap between the datasets used. However, we thought that a minor overlap of the data might not be associated with a major change in the results. Meanwhile, the two databases showed a commonality in their treatment patterns, and this, in fact, could be a topic for exploratory research on methodology. While this study focused on the development of a model, in the future, we will address a diverse range of parameters to determine the most effective MV-LDA model for detecting prescription misuse.

### Comparison With Prior Work

Studies, such as those encouraging regular medication review and introducing automated information systems [47,48], have been conducted with the aim of controlling the inappropriate use of medications in China. However, the increasing number of new drugs entering the market, delays in updating the databases, and insufficient knowledge of medications all raise the probability of nonideal use [49]. Knowledge-based and experience-based software has relevant limitations, including efficiency constraints. However, data mining techniques are customizable and can identify inappropriate prescriptions. For example, association rule mining has been used to find inappropriate prescriptions by calculating the co-occurrence of medications and diseases, resulting in a sensitivity of 75.9% and a specificity of 89.5% [42,43]. These methods however do have disadvantages, including inefficiency in the generation of candidate item sets because they require vast data sources and the frequent scanning of databases.

Furthermore, these methods often fail to explore latent structures and are prone to making spurious associations that can mislead clinical practitioners. Besides, in a previous study, a model combining natural language processing with guidelines based on expert knowledge was used to detect medication overuse,



and it showed degrees of sensitivity and specificity that were similar to those in our study [50]. Despite using different data sources and operating under diverse study conditions, we noted a higher sensitivity as compared with the association rule mining method. Although we failed to obtain a higher PPV, which is strongly related to the prevalence of inappropriate prescriptions, we think the MV-LDA model is suitable for preliminary screening and can be an alternative detection method, allowing clinical practice to flag potentially inappropriate prescriptions for manual review. Such a step could save a large amount of working time and reduce labor intensity.

A topic model is a multiple machine learning method and is used to reveal the semantics in the body of the text. With its advantages of topic extraction and model expansibility, LDA has become a commonly used topic modeling method. It was first used to extract underlying semantics and was then optimized to become a robust means of text mining analysis for social media [51-53]. Recently, topic modeling methods, particularly LDA, have been used for both structured and unstructured clinical data [21,26,54-56]. Several studies have attempted to scale the efficiency of LDA's topic generation. The symptom-herb-diagnosis topic model, which was proposed to determine the association between treatment with Chinese medicine and diabetes, can be used to find herbs to treat specific symptoms [55]. Multiple-channel LDA [57] focused on the support system for clinical decisions and based itself on a similar concept that the coupling of diagnoses and medications reflects the health status of patients at the time of seeing a doctor. However, we were not able to obtain a piece of well-recorded

contextual information or additional information, but we could leverage two variables (diagnosis and medication) and realize the aim of the study. Besides, given the shortcomings of miscellaneous algorithms and more extended calculations of an LDA-based model, various measures were taken to improve the MV-LDA algorithm in our project, which is not the focus of this study. The multiview topic modeling approach used here has been previously tested in different languages [34]. This allowed us to take both medications and diagnoses into consideration simultaneously. We leveraged these advantages and processed only half of the data used in a previous study to determine the association between diagnoses and medications for each prescription [58]. Prescriptions in our study were considered the equivalent of articles in that previous study, reflecting the particular situation of those patients. The topics and their allocations were consistent with clinical practice, providing proof of the robustness of our method.

## Conclusions

Our MV-LDA model can train the distribution of diagnosis-medication topics from a large number of prescriptions and can detect the potentially inappropriate use of prescription medications when combined with the TM method. Considering its mediocre specificity and moderate sensitivity, this model can be used as a primary screening tool and will likely complement and improve manual review. The model still needs more extension of views (introduction of more variables) to make full use of the information in the prescription and further improve the ability to identify prescription misuse.

## Acknowledgments

This work was supported by the National Natural Foundation of China (grant number 91646107) and the Beijing Municipal Science and Technology Project (grant number D151100002215002).

## Authors' Contributions

LZ was the principal investigator for this study; she contributed to the design, analysis, and interpretation of the study. YY and YC contributed to the analysis and interpretation of data and provided clinical support. SL, ST, and JZ contributed to the model's development. YH and JZ contributed to data extraction. LZ and YC drafted the manuscript. SZ provided overall supervision of the study and critically edited the manuscript. All authors reviewed the manuscript critically for scientific content, and all authors gave final approval of this version for publication.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Confusion matrix of model evaluation (multiview latent Dirichlet allocation model versus experts).

[\[DOCX File , 14 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Example results of multiview latent Dirichlet allocation topic modeling (K=30).

[\[DOCX File , 25 KB-Multimedia Appendix 2\]](#)

## References

1. World Health Organization. 2004. World Medicines Situation URL: [https://www.who.int/medicines/areas/policy/world\\_medicines\\_situation/en/](https://www.who.int/medicines/areas/policy/world_medicines_situation/en/) [accessed 2019-11-14]

2. Holloway K, Dijk LV. World Health Organization. 2011. The World Medicines Situation 2011: Rational Use of Medicines URL: [https://www.who.int/medicines/areas/policy/world\\_medicines\\_situation/WMS\\_ch14\\_wRational.pdf](https://www.who.int/medicines/areas/policy/world_medicines_situation/WMS_ch14_wRational.pdf) [accessed 2019-10-12]
3. Hamilton HJ, Gallagher PF, O'Mahony D. Inappropriate prescribing and adverse drug events in older people. *BMC Geriatr* 2009 Jan 28;9:5 [FREE Full text] [doi: [10.1186/1471-2318-9-5](https://doi.org/10.1186/1471-2318-9-5)] [Medline: [19175914](https://pubmed.ncbi.nlm.nih.gov/19175914/)]
4. Lederberg J. Infectious history. *Science* 2000 Apr 14;288(5464):287-293. [doi: [10.1126/science.288.5464.287](https://doi.org/10.1126/science.288.5464.287)] [Medline: [10777411](https://pubmed.ncbi.nlm.nih.gov/10777411/)]
5. Reynolds L, McKee M. Factors influencing antibiotic prescribing in China: an exploratory analysis. *Health Policy* 2009 Apr;90(1):32-36. [doi: [10.1016/j.healthpol.2008.09.002](https://doi.org/10.1016/j.healthpol.2008.09.002)] [Medline: [18849089](https://pubmed.ncbi.nlm.nih.gov/18849089/)]
6. Song Y, Bian Y, Petzold M, Li L, Yin A. The impact of China's national essential medicine system on improving rational drug use in primary health care facilities: an empirical study in four provinces. *BMC Health Serv Res* 2014 Oct 25;14:507 [FREE Full text] [doi: [10.1186/s12913-014-0507-3](https://doi.org/10.1186/s12913-014-0507-3)] [Medline: [25344413](https://pubmed.ncbi.nlm.nih.gov/25344413/)]
7. Li Y, Xu J, Wang F, Wang B, Liu L, Hou W, et al. Overprescribing in China, driven by financial incentives, results in very high use of antibiotics, injections, and corticosteroids. *Health Aff (Millwood)* 2012 May;31(5):1075-1082. [doi: [10.1377/hlthaff.2010.0965](https://doi.org/10.1377/hlthaff.2010.0965)] [Medline: [22566449](https://pubmed.ncbi.nlm.nih.gov/22566449/)]
8. Ministry of Health of the People's Republic of China. 2010. Management Practices of Hospital Prescription Comment (trial) URL: [http://www.gov.cn/gzdt/2010-03/04/content\\_1547080.htm](http://www.gov.cn/gzdt/2010-03/04/content_1547080.htm) [accessed 2018-02-06]
9. Zhang Y, Li P, Li J, Wang D, Mei D, Zhang B. Influence of automated pharmacy system on waiting time in outpatient pharmacy. *Chinese Journal of Hospital Pharmacy* (1) 2014:63-66. [doi: [10.13286/j.cnki.chinhosp pharmacy.2014.01.19](https://doi.org/10.13286/j.cnki.chinhosp pharmacy.2014.01.19)]
10. Peking University Third Hospital. 2018. Outpatient pharmacy URL: <https://www.puh3.net.cn/yjk/ksbm/153254.shtml> [accessed 2018-11-04]
11. Wang X, Gong Z, Zhou Y, Huang Y. Research of real-time review system of electronic prescriptions for outpatient and emergency in hospital. *Science Mosaic* 2016(5):33-35.
12. Gao Y, Fu L, Zhong X, Liu Z. Discussions on Problems about the Monitoring System for Rational Drug Use and Relevant Countermeasures. *China Pharmacy* 2015(22):3159-3161. [doi: [10.6039/j.issn.1001-0408.2015.22.42](https://doi.org/10.6039/j.issn.1001-0408.2015.22.42)]
13. Meyer J, Ostrzinski S, Fredrich D, Havemann C, Krafczyk J, Hoffmann W. Efficient data management in a large-scale epidemiology research project. *Comput Methods Programs Biomed* 2012 Sep;107(3):425-435. [doi: [10.1016/j.cmpb.2010.12.016](https://doi.org/10.1016/j.cmpb.2010.12.016)] [Medline: [21256617](https://pubmed.ncbi.nlm.nih.gov/21256617/)]
14. Peikari M, Salama S, Nofech-Mozes S, Martel AL. A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification. *Sci Rep* 2018 May 08;8(1):7193 [FREE Full text] [doi: [10.1038/s41598-018-24876-0](https://doi.org/10.1038/s41598-018-24876-0)] [Medline: [29739993](https://pubmed.ncbi.nlm.nih.gov/29739993/)]
15. Hu X, Gallagher M, Loveday W, Connor J, Wiles J. Detecting anomalies in controlled drug prescription data using probabilistic models. : Springer, Cham; 2015 Presented at: Australasian Conference on Artificial Life and Computational Intelligence; February 5-7, 2015; Newcastle, NSW, Australia p. 337-349. [doi: [10.1007/978-3-319-14803-8\\_26](https://doi.org/10.1007/978-3-319-14803-8_26)]
16. Nirad D, Surendro K. Outlier detection using association rule mining for information quality improvement. 2017 Presented at: International Conference on Recent Trends in Science, Engineering and Technology; July 10-11, 2017; Bangkok, Thailand. [doi: [10.17758/eap.dir0717002](https://doi.org/10.17758/eap.dir0717002)]
17. Valko M, Kveton B, Valizadegan H, Cooper G, Hauskrecht M. Conditional anomaly detection with soft harmonic functions. : IEEE; 2011 Presented at: 11th International Conference on Data Mining; December 11-14, 2011; Vancouver, BC, Canada p. 735-743. [doi: [10.1109/icdm.2011.40](https://doi.org/10.1109/icdm.2011.40)]
18. Song X, Wu M, Jermaine C, Ranka S. Conditional Anomaly Detection. *IEEE Trans Knowl Data Eng* 2007 May;19(5):631-645. [doi: [10.1109/tkde.2007.1009](https://doi.org/10.1109/tkde.2007.1009)]
19. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012 May 02;13(6):395-405. [doi: [10.1038/mrg3208](https://doi.org/10.1038/mrg3208)] [Medline: [22549152](https://pubmed.ncbi.nlm.nih.gov/22549152/)]
20. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3(4-5):993-1022. [doi: [10.1056/NEJM196803212781204](https://doi.org/10.1056/NEJM196803212781204)] [Medline: [5637250](https://pubmed.ncbi.nlm.nih.gov/5637250/)]
21. Park S, Choi D, Kim M, Cha W, Kim C, Moon I. Identifying prescription patterns with a topic model of diseases and medications. *J Biomed Inform* 2017;75:35-47 [FREE Full text] [doi: [10.1016/j.jbi.2017.09.003](https://doi.org/10.1016/j.jbi.2017.09.003)] [Medline: [28958484](https://pubmed.ncbi.nlm.nih.gov/28958484/)]
22. Zeng QT, Redd D, Rindfleisch T, Nebeker J. Synonym, topic model and predicate-based query expansion for retrieving clinical documents. *AMIA Annu Symp Proc* 2012;2012:1050-1059 [FREE Full text] [Medline: [23304381](https://pubmed.ncbi.nlm.nih.gov/23304381/)]
23. Shivashankar S, Srivathsan S, Ravindran B, Tendulkar AV. Multi-view methods for protein structure comparison using latent dirichlet allocation. *Bioinformatics* 2011 Jul 01;27(13):i61-i68 [FREE Full text] [doi: [10.1093/bioinformatics/btr249](https://doi.org/10.1093/bioinformatics/btr249)] [Medline: [21685102](https://pubmed.ncbi.nlm.nih.gov/21685102/)]
24. Huang Z, Dong W, Ji L, Gan C, Lu X, Duan H. Discovery of clinical pathway patterns from event logs using probabilistic topic models. *J Biomed Inform* 2014 Feb;47:39-57 [FREE Full text] [doi: [10.1016/j.jbi.2013.09.003](https://doi.org/10.1016/j.jbi.2013.09.003)] [Medline: [24076435](https://pubmed.ncbi.nlm.nih.gov/24076435/)]
25. Huang Z, Dong W, Ji L, He C, Duan H. Incorporating comorbidities into latent treatment pattern mining for clinical pathways. *J Biomed Inform* 2016 Feb;59:227-239 [FREE Full text] [doi: [10.1016/j.jbi.2015.12.012](https://doi.org/10.1016/j.jbi.2015.12.012)] [Medline: [26719169](https://pubmed.ncbi.nlm.nih.gov/26719169/)]
26. Zhang L, Zhao J, Wang Y, Xie B. Mining Patterns of Disease Progression: A Topic-Model-Based Approach. *Stud Health Technol Inform* 2016;228:354-358. [Medline: [27577403](https://pubmed.ncbi.nlm.nih.gov/27577403/)]

27. Chong W, Blei D, Li FF. Simultaneous image classification and annotation. 2009 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 20-25, 2009; Miami, FL, USA p. 1903-1910. [doi: [10.1109/cvpr.2009.5206800](https://doi.org/10.1109/cvpr.2009.5206800)]
28. Wang X, Ma X, Grimson E. Unsupervised activity perception by hierarchical bayesian models. *IEEE Trans Pattern Anal Mach Intell* 2009;31(3):539-555. [doi: [10.1109/cvpr.2007.383072](https://doi.org/10.1109/cvpr.2007.383072)]
29. Poldrack RA, Mumford JA, Schonberg T, Kalar D, Barman B, Yarkoni T. Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS Comput Biol* 2012;8(10):e1002707 [FREE Full text] [doi: [10.1371/journal.pcbi.1002707](https://doi.org/10.1371/journal.pcbi.1002707)] [Medline: [23071428](https://pubmed.ncbi.nlm.nih.gov/23071428/)]
30. Huang Z, Dong W, Duan H. A probabilistic topic model for clinical risk stratification from electronic health records. *J Biomed Inform* 2015 Dec;58:28-36 [FREE Full text] [doi: [10.1016/j.jbi.2015.09.005](https://doi.org/10.1016/j.jbi.2015.09.005)] [Medline: [26370451](https://pubmed.ncbi.nlm.nih.gov/26370451/)]
31. Liu B, Liu L, Tsykin A, Goodall G, Green J, Zhu M, et al. Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation. *Bioinformatics* 2010 Dec 15;26(24):3105-3111 [FREE Full text] [doi: [10.1093/bioinformatics/btq576](https://doi.org/10.1093/bioinformatics/btq576)] [Medline: [20956247](https://pubmed.ncbi.nlm.nih.gov/20956247/)]
32. Pivovarov R, Perotte AJ, Grave E, Angiolillo J, Wiggins CH, Elhadad N. Learning probabilistic phenotypes from heterogeneous EHR data. *J Biomed Inform* 2015 Dec;58:156-165 [FREE Full text] [doi: [10.1016/j.jbi.2015.10.001](https://doi.org/10.1016/j.jbi.2015.10.001)] [Medline: [26464024](https://pubmed.ncbi.nlm.nih.gov/26464024/)]
33. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. *Springerplus* 2016;5(1):1608 [FREE Full text] [doi: [10.1186/s40064-016-3252-8](https://doi.org/10.1186/s40064-016-3252-8)] [Medline: [27652181](https://pubmed.ncbi.nlm.nih.gov/27652181/)]
34. Zhang G, Iwata T, Kashima H. Robust multi-view topic modeling by incorporating detecting anomalies. 2017 Presented at: Joint European Conference on Machine Learning and Knowledge Discovery in Databases; September 18-22, 2017; Skopje, Macedonia p. 238-250. [doi: [10.1007/978-3-319-71246-8\\_15](https://doi.org/10.1007/978-3-319-71246-8_15)]
35. Sun S. A survey of multi-view machine learning. *Neural Comput & Applic* 2013 Feb 17;23(7-8):2031-2038. [doi: [10.1007/s00521-013-1362-6](https://doi.org/10.1007/s00521-013-1362-6)]
36. Xu C, Tao D, Xu C. A survey on multi-view learning. *arXiv preprint* 2013 [FREE Full text]
37. Zhang L, Li X, Liu H, Mei J, Hu G, Zhao J. Probabilistic-mismatch anomaly detection: do one's medications match with the diagnoses. 2016 Presented at: IEEE 16th International Conference on Data Mining (ICDM); December 12-15, 2016; Barcelona, Spain p. 659-668. [doi: [10.1109/icdm.2016.0077](https://doi.org/10.1109/icdm.2016.0077)]
38. Hu Y. Establishment and application of Chinese monitoring network for the rational use of drugs system. *Exploration of Rational Drug Use in China* 2009;6(8):5-8.
39. Yang Y, Zhou X, Gao S, Lin H, Xie Y, Feng Y, et al. Evaluation of Electronic Healthcare Databases for Post-Marketing Drug Safety Surveillance and Pharmacoepidemiology in China. *Drug Saf* 2018 Jan;41(1):125-137. [doi: [10.1007/s40264-017-0589-z](https://doi.org/10.1007/s40264-017-0589-z)] [Medline: [28815480](https://pubmed.ncbi.nlm.nih.gov/28815480/)]
40. Zhen J, Bian B, Kong F, Yan B. Effect evaluation of regional prescription review on rational clinical drug use. *Chinese Journal of Hospital Administration* 2015(7):531-533.
41. Ministry of Health of the People's Republic of China. Prescription Administrative Policy 2007 URL: [http://www.gov.cn/flfg/2007-03/13/content\\_549406.htm](http://www.gov.cn/flfg/2007-03/13/content_549406.htm) [accessed 2019-11-08]
42. Nguyen PA, Syed-Abdul S, Iqbal U, Hsu M, Huang C, Li H, et al. A probabilistic model for reducing medication errors. *PLoS One* 2013;8(12):e82401 [FREE Full text] [doi: [10.1371/journal.pone.0082401](https://doi.org/10.1371/journal.pone.0082401)] [Medline: [24312659](https://pubmed.ncbi.nlm.nih.gov/24312659/)]
43. Yang H, Iqbal U, Nguyen PA, Lin S, Huang C, Jian W, et al. An automated technique to identify potential inappropriate traditional Chinese medicine (TCM) prescriptions. *Pharmacoepidemiol Drug Saf* 2016 Apr;25(4):422-430. [doi: [10.1002/pds.3976](https://doi.org/10.1002/pds.3976)] [Medline: [26910512](https://pubmed.ncbi.nlm.nih.gov/26910512/)]
44. Yang M, Wang D, Wang X, Zhang Y. Prescription review and inappropriate prescription analysis in our hospital in 2013. *Chinese Medical Science* 2014(16):129-131.
45. Li DC, Thermeau T, Chute C, Liu H. Discovering associations among diagnosis groups using topic modeling. *AMIA Jt Summits Transl Sci Proc* 2014;2014:43-49 [FREE Full text] [Medline: [25954576](https://pubmed.ncbi.nlm.nih.gov/25954576/)]
46. Liu S, Tang S, Zhao J, Wang Y, Zhuo L. An extended topic model based abnormal medical prescription detection method. 2018 Presented at: National Conference on Pervasive Computing; September 14-16, 2018; Tianjin, China.
47. World Health Organization. 2016. Medication Errors: Technical Series on Safer Primary Care URL: <https://apps.who.int/iris/bitstream/handle/10665/252274/9789241511643-eng.pdf?sequence=1> [accessed 2020-04-24]
48. Velo GP, Minuz P. Medication errors: prescribing faults and prescription errors. *Br J Clin Pharmacol* 2009 Jun;67(6):624-628 [FREE Full text] [doi: [10.1111/j.1365-2125.2009.03425.x](https://doi.org/10.1111/j.1365-2125.2009.03425.x)] [Medline: [19594530](https://pubmed.ncbi.nlm.nih.gov/19594530/)]
49. Yuan N, Chen N. Reasons of irrational drug use in medical institutions and its countermeasures: a case study of irrational drug use in department of gastroenterology. *Medicine and Philosophy* 2015;36(15):51-53.
50. Salmasian H, Freedberg DE, Abrams JA, Friedman C. An automated tool for detecting medication overuse based on the electronic health records. *Pharmacoepidemiol Drug Saf* 2013 Feb;22(2):183-189 [FREE Full text] [doi: [10.1002/pds.3387](https://doi.org/10.1002/pds.3387)] [Medline: [23233423](https://pubmed.ncbi.nlm.nih.gov/23233423/)]
51. Zhao W, Jiang J, Weng J, He J, Lim E, Yan H, et al. Comparing twitter and traditional media using topic models. 2011 Presented at: European Conference on Information Retrieval; April 18-21, 2011; Dublin, Ireland p. 338-349. [doi: [10.1007/978-3-642-20161-5\\_34](https://doi.org/10.1007/978-3-642-20161-5_34)]

52. Xianghua F, Guo L, Yanyan G, Zhiqiang W. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems* 2013 Jan;37(2):186-195. [doi: [10.1016/j.knosys.2012.08.003](https://doi.org/10.1016/j.knosys.2012.08.003)]
53. Fu X, Liu G, Guo Y, Guo W. Multi-aspect blog sentiment analysis based on LDA topic model and Hownet Lexicon. 2011 Presented at: International Conference on Web Information Systems and Mining; September 24-25, 2011; Taiyuan, China p. 131-138. [doi: [10.1007/978-3-642-23982-3\\_17](https://doi.org/10.1007/978-3-642-23982-3_17)]
54. Lin F, Xiahou J, Xu Z. TCM clinic records data mining approaches based on weighted-LDA and multi-relationship LDA model. *Multimed Tools Appl* 2016 Apr 13;75(22):14203-14232. [doi: [10.1007/s11042-016-3363-9](https://doi.org/10.1007/s11042-016-3363-9)]
55. Zhang X, Zhou X, Huang H, Feng Q, Chen S, Liu B. Topic model for Chinese medicine diagnosis and prescription regularities analysis: case on diabetes. *Chin J Integr Med* 2011 Apr;17(4):307-313. [doi: [10.1007/s11655-011-0699-x](https://doi.org/10.1007/s11655-011-0699-x)] [Medline: [21509676](https://pubmed.ncbi.nlm.nih.gov/21509676/)]
56. Cohen R, Aviram I, Elhadad M, Elhadad N. Redundancy-aware topic modeling for patient record notes. *PLoS One* 2014;9(2):e87555 [FREE Full text] [doi: [10.1371/journal.pone.0087555](https://doi.org/10.1371/journal.pone.0087555)] [Medline: [24551060](https://pubmed.ncbi.nlm.nih.gov/24551060/)]
57. Lu H, Wei C, Hsiao F. Modeling healthcare data using multiple-channel latent Dirichlet allocation. *J Biomed Inform* 2016 Apr;60:210-223 [FREE Full text] [doi: [10.1016/j.jbi.2016.02.003](https://doi.org/10.1016/j.jbi.2016.02.003)] [Medline: [26898516](https://pubmed.ncbi.nlm.nih.gov/26898516/)]
58. Nguyen C, Zhan D, Zhou Z. Multi-modal image annotation with multi-instance multi-label LDA. 2013 Presented at: 23rd International Joint Conference on Artificial Intelligence; August 3-9, 2013; Beijing, China.

## Abbreviations

- BRPRD:** Beijing Regional Prescription Review Database  
**CMNRUD:** Chinese Monitoring Network for the Rational Use of Drugs  
**LDA:** latent Dirichlet allocation  
**MV-LDA:** multiview latent Dirichlet allocation  
**NPV:** negative-predictive value  
**PPV:** positive-predictive value  
**TM:** topic mapping

*Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by Z Yang, A Aminbeidokhti; comments to author 25.11.19; revised version received 18.01.20; accepted 24.03.20; published 06.07.20*

*Please cite as:*

Zhuo L, Cheng Y, Liu S, Yang Y, Tang S, Zhen J, Zhao J, Zhan S

*A Multiview Model for Detecting the Inappropriate Use of Prescription Medication: Machine Learning Approach*

*JMIR Med Inform* 2020;8(7):e16312

URL: <https://medinform.jmir.org/2020/7/e16312>

doi: [10.2196/16312](https://doi.org/10.2196/16312)

PMID:

©Lin Zhuo, Yinchu Cheng, Shaoqin Liu, Yu Yang, Shuang Tang, Jiancun Zhen, Junfeng Zhao, Siyan Zhan. Originally published in *JMIR Medical Informatics* (<http://medinform.jmir.org>), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.