Scientia Iranica E (2022) 29(6), 3489–3504

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

Detecting factors associated with polypharmacy in general practitioners' prescriptions: A data mining approach

M. Moradi[a,b,*], M. Modarres[a], and M.M. Sepehri[c]

a. *Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran.*
b. *Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran.*
c. *Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.*

Received 3 March 2020; received in revised form 18 October 2020; accepted 7 December 2020

KEYWORDS
Decision tree;
Classification And Regression Tree (CART);
C4.5;
Parameter tuning;
Response Surface Method (RSM);
Rational use of drugs.

Abstract. Prescribing and consuming drugs more than necessary is considered polypharmacy, which is both wasteful and harmful. The purpose of this paper is to establish an innovative data mining framework for analyzing physicians' prescriptions regarding polypharmacy. The approach consists of three main steps: pre-modeling, modeling, and post-modeling. In the first step, after collecting and cleaning the raw data, several novel features of physicians are extracted. In the modeling step, two popular decision trees, i.e., C4.5 and Classification And Regression Tree (CART), are applied to generate a set of If-Then rules in a tree-shaped structure to detect and describe physicians' features associated with polypharmacy. In a novel approach, the Response Surface Method (RSM) as a tool for hyper-parameter tuning is simultaneously applied along with Correlation-based Feature Selection (CFS) to enhance the performance of the algorithms. In the post-modeling step, the discovered knowledge is visualized to make the results more perceptible and is, then, presented to domain experts to evaluate whether they make sense or not. The framework was applied to a real-world dataset of prescriptions. The results were confirmed by the experts, which demonstrated the capabilities of the data mining framework in the detection and analysis of polypharmacy. The derived If-Then rules can be beneficial for healthcare managers and policy-makers to recognize physicians' prescribing patterns and take suitable actions to support medicine management and develop high-quality prescribing guidelines.

1. Introduction

Pharmaceutical spending is a major portion of total healthcare costs. This proportion differs significantly among countries based on their income; pharmaceutical expenditure as a share of total health cost varies from 19.7% in the high-income economies to a mean of 30.4% in low-income ones [1]. World Health Organization (WHO) has announced that about 50% of all drugs are unsuitably prescribed, dispensed, or sold [2]. Irrational use of drugs remains an extremely severe and widespread issue across the world, especially in developing countries [3]. It is both wasteful and harmful [4]. Despite being a global issue, few countries are monitoring medicines prescribing and making sufficient policies and taking appropriate actions to resolve the situation [5]. The essential steps for limiting the irrational use of drugs consist of recognizing the types, extent, and causes for their irrational prescribing and consumption [6]. WHO and the International Network of Rational Use of Drugs (INRUD) have

*. *Corresponding author.*
E-mail address: m.moradi@ie.sharif.edu (M. Moradi)

doi: 10.24200/sci.2020.55569.4283

introduced a group of medicine-prescribing indicators as measures to prescribe quality evaluation [7]. The first indicator, the *average number of drugs prescribed per encounter*, is used to measure the extent of polypharmacy. Polypharmacy can be associated with unnecessary and/or inappropriate drug prescriptions [8]. It does not matter whether the patient has taken them or not [9]. Although various numerical cut-offs have been used to describe polypharmacy, the usual definition is the recommendation of two or three drugs per prescription [9].

"It is very challenging to reach a conclusion about polypharmacy and its related factors because of the wide variety of the definitions applied" [10]. Physicians, pharmacists, and patients are all involved as well as healthcare managers and policy-makers. The authors in Australia [11] and [12] and those in Italy [13] attempted to identify patients' attitudes, opinions, and experiences regarding polypharmacy and readiness to withdraw medications. Quinn and Shah analyzed the prevalence of polypharmacy in outpatient prescription drug claims in the United States containing 4 billion patient-month records [14]. According to Hovstadius and Petersson [15], the factors possibly causing polypharmacy can be divided into four major categories: (1) Factors related to the healthcare system, including the development of society and healthcare services, development of new technologies and therapies, and increased use of drug therapy. (2) Factors related to patients, containing age, sex, ethnicity, socioeconomic status, behavior, clinical condition, and medical therapy. (3) Factors related to physicians, including physician practice environment, guidelines, behavior, and prescribing habits. (4) The interaction between doctor and patient. Analyzing prescription data written by Danish General Practitioners (GPs) utilizing backward stepwise linear multiple regression, Bjerrum et al. identified that the practice structure, workload, clinical work profile, and prescribing profile were the predictors of major polypharmacy (using $\geq$ 5 drugs concurrently). They also claimed that the occurrence of major polypharmacy was meaningfully lower for female GPs than for male GPs; however, the age of physicians or experience did not affect major polypharmacy [16]. Anthierens et al. directed semi-structured interviews to study 65 Belgian GPs' views and beliefs on polypharmacy to classify the role of GPs in improving prescribing patterns [17]. In a study conducted by O'Dwyer et al. demographic variables and reported chronic health conditions were studied in association with polypharmacy in 734 Irish elderlies with intellectual disability using a multinomial logistic regression model [18]. Ie et al. explored factors associated with the prescribing behavior of 61 family physicians prescribed for 932 patients using multivariable regression. They considered patient related characteristics including gender, age, race, and health condition. They also investigated physician related features consisting of gender, age, years since graduation, position, and their responses to a survey of polypharmacy and Potentially Inappropriate Medicines (PIM). The same authors claimed that physicians who care more about the number of prescribed medications and use the Beers List recommended fewer drugs and PIMs. However, physicians' related factors including gender, experience, and perceived confidence were not associated with prescribing behavior in family medicine residency practices [19]. Slater et al. demonstrated that patients' related factors including lower wealth, increasing age, obesity, and occurrence of chronic conditions were associated with polypharmacy prevalence among patients older than 50 years in primary care in England [20].

Furthermore, many studies all around the world, such as [21], [22], [23], [24], and [25], analyzed *the average number of drugs prescribed per encounter* as well as the other indicators of WHO and INRUD to identify the extent of rational use of drugs. On the other hand, Cerrito suggested using a data mining approach to examine the adverse effects caused by polypharmacy in cardiology patients [26]. Data mining is an attractive approach in the pharmacovigilance field of study to identify Adverse Drug Events (ADE) and Adverse Drug Reaction (ADR), which focus on post-marketing surveillance of medicines. For example, readers are referred to [27], [28], [29], and [30] for further relevant research.

Despite the theoretical development of data mining methods and their notable applications in various fields of healthcare studies in recent decades [31], to the best of the authors' knowledge, a data mining approach has not been applied to analyze the rational use of drugs including polypharmacy. Meanwhile, a tremendous amount of drug prescription data is continuously generated by physicians and stored in health information systems. Analysis of this kind of data will offer solutions to rationalize the use of drugs and as a consequence, it can improve the patients' health condition and decrease insurance costs. Therefore, the question that arises is whether data mining can be used in the analysis of prescription quality and rational use of drugs. In addition, one should consider whether an organized and usable data mining framework can be provided for the analysis of huge prescription data to understand the extend of polypharmacy. To answer the questions, the authors have developed an unprecedented data mining framework and applied it to a real-world prescription dataset prescribed by GPs. In other words, in this paper, a data mining approach has been introduced. Then, to study its performance, the GPs' prescribing pattern is monitored and analyzed regarding the average number of drugs recommended per prescription to appraise the extent of polyphar-

macy. Furthermore, the significant physician-related characteristics that are associated with polypharmacy are identified. It is necessary to mention that the choice of GPs is due to the fact that the majority of patients refer to them; consequently, they should be involved in any prescription quality improvement practices [32].

The structure of the rest of the paper is as follows: Section 2 introduces a data mining framework for analyzing physicians' prescribing patterns concerning polypharmacy. Section 3 shows how the proposed method is applied to a real-world prescription dataset. Section 4 contains results and discussion. Finally, Section 5 presents the conclusion and outlines of the future works.

2. Data mining methodology

The data mining framework applied in this paper is inspired by ASUM-MD [33]. As is shown in Figure 1, it is carried out in three main steps: pre-modeling, modeling, and post-modeling. Real world data are mostly poor in quality and also, are often not in the desired format for use in data mining algorithms. Consequently, the use of such data in the machine learning algorithms can lead to misleading results. Thus, before taking any actions, it is necessary to evaluate the data and fix its potential deficiencies and shortcomings. It is, particularly, important for healthcare databases that contain patients, physicians, and their prescriptions records. Identification, collection, preparation, and improvement of the data are made at the pre-modeling phase. Then, in the modeling step, the appropriate machine learning algorithms are applied to diagnosing polypharmacy and identifying the physician-related characteristics associated with it. Because the results of the modeling phase are usually incomprehensible, it is necessary to present them to the domain experts in a more understandable format. Then, their opinions on the approval or rejection of the discovered patterns should be obtained, and their feedback should be applied in different previous stages to achieve better results. This is done at the post-modeling phase. Further details of the data mining framework are provided below.

2.1. Pre-modeling

2.1.1. Data gathering and understanding

The required raw data should be collected from trusted sources regarding research objectives and business level requirements. Also, the data may be described using visualization techniques and tools. Patient, physician, and prescription data are required to be collected in the study.

2.1.2. Data preparation

In the data preparation task, the quality of the raw data is evaluated and improved. It aims to get rid of irrelevant, noisy, and inconsistent data from the raw data. This phase also consists of feature extraction, data transformation, and handling of missing values. This phase includes data compression and data reduction. Because different data sources, i.e., patient data, prescription data, physician data, and some supplementary data are used, it is necessary to integrate them carefully. At the end of this step, the prepared data are ready to be mined.

2.2. Modeling

In the modeling step, appropriate machine learning algorithms and data mining techniques are applied based on the business goals and the quality of the prepared data. The prepared dataset is split into training and testing subsets based on a predetermined strategy. The training dataset is used to build a model and tune its parameters, whereas the testing dataset is utilized for performance evaluation of the model. In the study, a holdout dataset approach is used, and the division is made up to assign 50% to the training and 50% to the testing subsets using the random stratified sampling strategy. Therefore, the proportion of the two different classes (0 or 1 for polypharmacy) in the primary dataset remains the same in both training and testing datasets.

2.2.1. Model building

In this sub-step, the modeling tool and machine learning algorithms are run on the training dataset to build one or more models. Also, it is essential to tune the parameters of each applied algorithm to appropriate values. Although feature selection is a part of data preparation, it may have an interaction with parameter tuning. Therefore, they are simultaneously applied to consider their possible interaction in the model building sub-step. Since each machine learning algorithm needs particular data requirements, it sometimes is necessary to return to the data preparation phase and make necessary changes to satisfy the desirable requirements.

In the research, the main problem is polypharmacy detection, which is a classification problem. The decision trees are a common and popular way of classification because of their accuracy, simplicity, ease of use, and interpretation [34]. Decision trees build a model demonstrated by a group of If-Then rules in a tree-shaped structure. Thus, decision tree classifiers are applied to recognize polypharmacy and describe the factors associated with it.

2.2.2. Model evaluating

To ensure the effectiveness of the generated models, it is necessary to assess their performance regarding the evaluation criteria such as accuracy. Testing dataset and the 10 fold cross validation method are employed to evaluate the performance of each algorithm. The authors assess each algorithm individually as well as

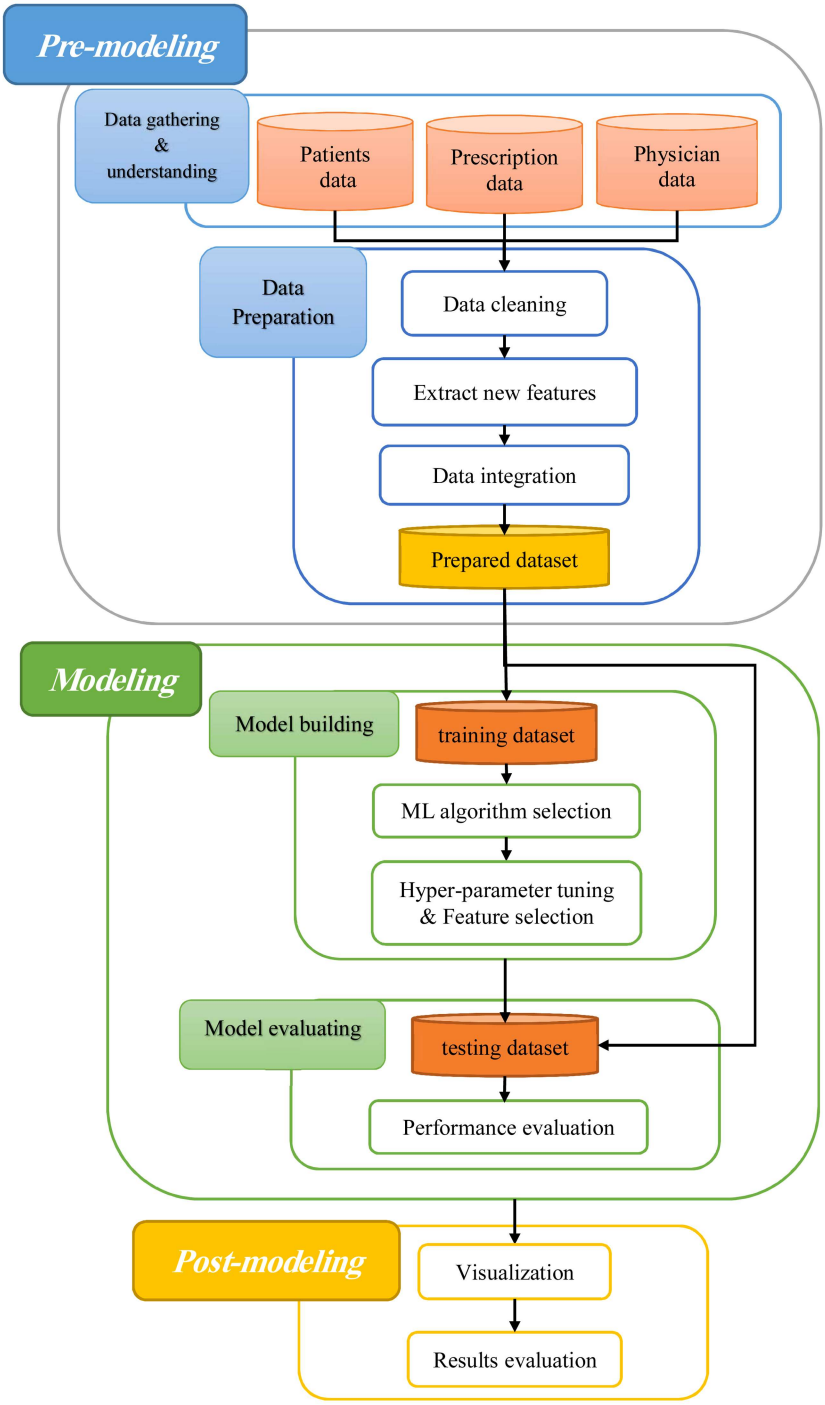


Figure 1. Proposed data mining procedures.

compare them with each other. Also, the transparency and comprehensiveness of each model are evaluated.

2.3. Post-modeling

Finally, the sophisticated results of the modeling phase are presented to domain experts in an understandable format. Their opinions on the discovered patterns are obtained, and their feedback is applied to achieve more rational results.

2.3.1. Visualization

Visualization is the practice of communicating and displaying extracted patterns and discovered knowledge in a way that can be easily understood, analyzed, and remembered. Employing various graphical or visual formats such as charts, graphs, and statistical representation is common in visualization. In this study, the discovered patterns are visualized in the form of tree.

2.3.2. Results evaluation
The data mining results should be interpreted in terms of its application and original business objectives. With the aid of the domain experts, it is necessary to review the discovered knowledge to see whether it makes logical sense and it is useful for the end users. The feedback can be employed to improve the data mining process and achieve more rational outcomes.

3. Detecting polypharmacy in GP's prescriptions

In this section, the developed data mining framework is used to identify polypharmacy and detect factors associated with it. For this purpose, a real-world database that contains the prescription data of GPs is used.

3.1. Pre-modeling

3.1.1. Data gathering and understanding
The raw data used in this research were obtained from the *National Center for Health Insurance Research* of the *Iran Health Insurance Organization* in the format of MS SQL Server 2012. The raw dataset includes patient's insurance code, insurance fund, birthday, sex, the province where the patient's insurance booklet is issued, prescription code, pharmacy code, physician's medical council number, visit date, prescription dispense date, prescription fee, Health Insurance Organization's share of the prescription fee, and the details of the drugs available in each prescription including drug code, amount, and fee.

In the current study, only physicians who have averagely prescribed at least 24 prescriptions per year were considered. In summary, the raw dataset consists of 3,860,835 prescriptions and 13,887,444 medicines. They were prescribed by 9,552 GPs for 1,962,417 different outpatients. The prescriptions were dispensed by 2,213 pharmacies in Tehran province between 21 March 2015 and 20 March 2016 and reimbursed by Iran Health Insurance. The average and the standard deviation of the number of drugs per prescription is 3.16 and 0.85, respectively.

3.1.2. Data preparation
In this research, the quality of the initial data is carefully assessed. The data not in a relevant format were identified and an attempt was made to find the right estimate for them; otherwise, they were excluded from the dataset. While GP's medical council number, patients' insurance code, and drug identification code are of standard format, they did not have this predefined format in some records. For example, while the medicines were recorded as a five-digit number in the Health Insurance database, some of them were logged as a three-digit number, which was invalid. Also, the records including missing values were identified and isolated.

To complete the initial dataset, an attempt was made to collect additional data about physicians, including the gender, age, experience, place of study, and place of work. However, the data was not available to the research group. Therefore, to enrich the dataset, some novel features are extracted from existing ones. In fact, the new features were not directly available in the prescription database. However, they were measured by authors using SQL codes considering the factors, leading to polypharmacy mentioned in the literature, or by approval of the domain experts. For example, in the initial database, it is only specified what doctor prescribed for which patient. However, the total number of patients who referred to a certain physician was extracted by SQL and added to the physician's profile. The feature determines the work burden of the GP. All extracted features are briefly presented as follows:

- *Gender:* The gender of physicians was not declared in the dataset. However, the doctor's name is extracted by matching the medical council number with another supplementary dataset. Next, the gender was specified and added to the dataset using a database of people's first names.
- *Experience:* Given that the medical council number assigned to each physician is an ordinal number starting from number one, it can be claimed that a physician with a smaller medical council number is more experienced. To create a feature that represents *experience*, the medical council number has been discretized at intervals of 40,000 units. By doing this, all physicians are divided into four separate categories 1 to 4. Group 1 identifies the most experienced, while Group 4 identifies the least experienced.
- *Drug Diversity (DD):* The calculated number of different drugs prescribed by each GP. It may represent the range of drugs that each GP is interested to prescribe.
- *Average Cost of Prescriptions (ACPr):* The average cost of prescriptions for each GP.
- *Total Number of Patients (TNPa):* It is clear what each GP prescribes for whom. The total number of patients who referred to each GP is calculated.
- *Percentage of Female Patients (PFPa):* The percentage of female patients of each GP.
- *Percentage of Male Patients (PMPa):* The percentage of male patients of each GP.
- *Percentage of Resident Patients (PRPa):* The calculated percentage of the patients whose insurance booklets have been issued inside the studied region (Tehran Province) for each GP.

- *Percentage of Non-Resident Patients (PNRPa):* The computed percentage of the patients whose insurance booklets issued outside the Tehran province for every GP.
- *Percentage of Special Patients (PSPa):* The percentage of patients with a special disease such as Thalassemia.
- *Percentage of Ordinary Patients (POPa):* The percentage of patients who do not have special conditions.
- *Percentage of Patients in Age Group i ($PPaAG_i$):* WHO uses the following age groups for estimation of the global burden of disease [35]: neonatal (< 28 days), 1–59 months, 5–14, 15–29, 30–49, 50–69, 70 years and older. However, in the study, due to the data conditions, the age is considered for the following age groups: 5 years and younger (AG1), 5–14 (AG2), 15–29 (AG3), 30–49 (AG4), 50–69 (AG5), and 70–94 (AG6). The percentage of patients in these age groups has been calculated for each GP.
- *Percentage of Patients in the ith insurance fund ($PPaF_i$):* There are four main insurances found: Employers (F1), workers (F2), rural (F3), and other categories (F4). The percentages of patients in different *insurance funds* have been computed for each GP.
- *Total Number of Prescriptions (TNPr):* The total number of prescriptions which is prescribed by each GP.
- *Average Number of Prescriptions per Patient (ANPP):* The average number of prescriptions prescribed by the GP for all of his/her patients. In other words, it illustrates the average number of referrals of each patient to the specific GP.
- *Total Number of Pharmacies (TNPh):* Illustrating how many pharmacies have dispensed prescriptions of each GP.
- *Average number of Prescriptions per Pharmacy (APrPh):* On average, illustrating how many prescriptions of each GP are dispensed in the affiliated pharmacies.
- *Polypharmacy:* Illustrating the class variable, which is binary. For each GP, if the average number of prescribed drugs per prescription is less than 3, then polypharmacy is zero; otherwise, it will be 1. Out of 9,552 GPs, 5,252 instances were labeled as 1.

According to domain experts, the two variables, i.e. *drug diversity* and the *average cost of prescriptions*, may have a strong relationship with polypharmacy. In other words, the expensive prescription probably contains more drugs. Also, a physician who has a greater *drug diversity* may have prescribed more drugs per encounter. Therefore, in the following, these two features were discarded. It is necessary to mention that several features contain unknown and missing values. For example, the *gender* of some patients is unknown, or the *age* of them is null or invalid. In this case, the patient is considered in calculating the *total number of patients*; however, it is ignored in the calculation of *gender* or *age group* percentage. Also, the GPs whose patient data, i.e., *gender, age group, and insurance fund*, contained more than 10% of the missing values were discarded. Since residency and specialty of patients contain considerable missing values, their related features have been ignored too. Given these conditions, the prepared dataset contains 5593 GPs, of whom 3295 had an average of more than three drugs per prescription. Table 1 represents details of the features applied in the modeling step.

3.2. Modeling

3.2.1. Model building

In the study, the core problem is the detection of polypharmacy, which is a binary classification. Thus, C4.5 and CART (Classification And Regression Tree)

Table 1. Dataset attributes.

Attribute	Symbol	Data type	Value	Role
Gender	Gender	Binary	0, 1	Input
Experience	Experience	Nominal	1, 2, 3, 4	Input
Total number of patients	TNPa	Numeric	≥ 1	Input
Percentage of female patients	PFPa	Numeric	[0,100]	Input
Percentage of male patients	PMPa	Numeric	[0,100]	Input
Percentage of patients in age group i	$PPaAG_i$	Numeric	[0,100]	Input
Percentage of patients in the ith insurance fund	$PPaF_i$	Numeric	[0,100]	Input
Total number of prescriptions	TNPr	Numeric	≥ 24	Input
Average number of prescriptions per patient	ANPP	Numeric	≥ 1	Input
Total number of pharmacies	TNPh	Numeric	≥ 1	Input
Average number of prescriptions per pharmacy	APrPh	Numeric	≥ 1	Input
Polypharmacy	PP	Binary	0, 1	Class

Table 2. Primary hyper-parameters of C4.5 and CART.

Algorithm	Hyper-parameter	Symbol	Possible value	Type	Default
CART	Minimum number of instances per leaf	MNIL	≥ 1	Integer	2
CART	Percentage of training data used to construct the tree	PTD	(0,1]	Real	1
CART	Using 1 SE rule to make pruning decision	1SEr	{False, True}	Logical	False
C4.5	Confidence factor	CF	[0.001, 0.5]	Real	0.25
C4.5	Minimum number of instances per leaf	MNIL	≥ 1	Integer	2
C4.5	Binary splits	BS	{False, True}	Logical	False

are selected based on their performance and applicability in binary classification. Wu et al. claimed that C4.5 and CART were among the top 10 useful and widespread data mining algorithms [36]. Also, they could generate a set of If-Then rules in a tree-shaped structure, facilitating description of the physician-related factors associated with polypharmacy. It is necessary to mention some efficient methods such as Support Vector Machines (SVMs), Neural Networks (NNs), and ensemble methods that were not applied in the study, since they are black box data mining algorithms and do not provide interpretable and explainable results [37]. In other words, although they may have appropriate performance, their internal logic is hidden to users and cannot describe the factors associated with polypharmacy.

C4.5 is an evolution and extension of ID3 (Iterative Dichotomiser 3) developed by Quinlan [38]. It is probably the most widely used machine learning algorithm in practice to date [39]. This supervised classification algorithm can handle continuous data, missing values, and noisy data and generate rules from trees. C4.5 utilizes *gain rate* as the goodness of a split to choose an attribute for splitting the dataset. The classifier decides the best numerical split point that has the least misclassification error. Then, splitting procedure is applied to the training dataset to generate sub-nodes. Finally, the algorithm applies a pruning step to simplify the classification rules without any loss of accuracy [38]. The post-pruning process in C4.5 is grounded on weak statistical assumptions. However, it is extremely agile and is, therefore, common in practice [39]. The algorithm has been used extensively in recent health studies [40–44].

CART was developed by Leo Breiman in the early 1980s. The algorithm, as its name implies, can produce both CARTs. It generates binary trees. More specifically, by applying CART, the dataset is divided into two subsets and their difference is mostly based on the *Gini index*. It uses the cost-complexity pruning method, which is a post-pruning approach [39]. The CART classifier has the option of producing multivariate tests, i.e., the algorithm can use a linear combination of attributes in splitting procedure [34]. It also can handle the missing values. CART is a popular machine learning algorithm in the field of healthcare and medical research for more than two decades and has been utilized in various medical researches [42,45–49].

To apply C4.5 and CART algorithms, *J48* and *SimpleCart* were employed from the *classify* tab of the *Explorer* interface of *WEKA* (8.3.2). J48 is an open-source java reimplementation of C4.5 revision 8, the last non-commercial version of C4.5 [39]. SimpleCart is also the implementation of CART in WEKA. C4.5 and CART, like many other machine learning algorithms, contain a set of hyper-parameters that can affect their performance significantly. Therefore, their values must be set carefully. Table 2 provides concise information about hyper-parameters to be considered and tuned.

There are several methods for parameter tuning including grid search, random search, Gaussian process, Bayesian optimization, and DOE-based method. Recently, Lujan-Moreno et al. proposed a method for parameter tuning using DOE and *Response Surface Method* (RSM) [50]. In this method, interaction among factors (parameters) is carefully considered; however, it is often ignored in hyper-parameter optimization studies [50]. RSM develops a response surface by creating regression models to characterize the dependence of performance of the applied algorithm on parameter configuration. Since the shape of the response surface has been formerly indistinct, it is necessary to find a polynomial model that fits the relationship between the predictors and the response. A low-order model such as first- and second-order polynomial models is often appropriate to illustrate such a relationship [51]. The second-order or "quadratic" polynomial model is usually favored as it delivers the best tradeoff between the modeling accuracy and computational effort and provides perfect response surface curvature around intended regions. The second-order RSM can be defined as [52]:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum \sum_{i<j}^{k} \beta_{ij} x_i x_j + \varepsilon. \quad (1)$$

RSM is a sequential method that moves in the direction of improvement at each step for optimizing its objective. Since RSM creates a surrogate model, it can obtain valuable results with only limited experiments.

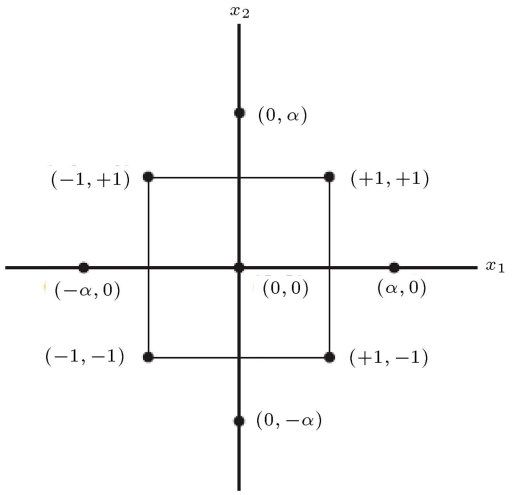


Figure 2. Central composite designs for $k = 2$.

The Central Composite Design (CCD) is one of the most prevalent RSM designs. As a rotatable design, it can be considered a factorial design in combination with some additional points, i.e., center point, and star points. They expand the cuboidal region of the initial factorial design. At each factor level, the curvature of the response surface can be estimated by additional points. Figure 2 depicts a CCD for two continuous factors with two levels, where -1 and $+1$ indicate lower and upper levels, respectively. Moreover, 0 demonstrates central point and star points are denoted by α [51].

Sometimes, it is necessary to optimize multiple objectives (responses) simultaneously. The *desirability function* is one of the most extensively applied approaches to optimizing multiple responses [52]. Each response R is transformed into an individual desirability function d ranging from 0 to 1, where 1 is the most desirable. If response R is considered to have a maximum value of objective or target T, we have:

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^t & L \leq y \leq T \\ 1 & y > L \end{cases} \quad (2)$$

If the objective or the target of the response is to be minimized, we have:

$$d = \begin{cases} 1 & y < T \\ \left(\frac{U-y}{U-T}\right)^t & T \leq y \leq U \\ 0 & y > U \end{cases} \quad (3)$$

By setting $t = 1$, the desirability function is linear. Setting $t > 1$ results in a greater emphasis on getting close to the target value, and setting $0 < t < 1$ makes this unimportant. If R_i is completely undesirable outside an acceptable region $d_i = 0$ and if the response R_i has a completely desirable value at its goal or target, then $d_i = 1$. The geometric mean of the individual desirability functions is the overall desirability function that should be maximized:

$$D = (d_1 * d_2 * ... * d_k)^{1/k}, \quad (4)$$

where k denotes the number of responses.

On the other hand, since there are many features in the prepared dataset, it should be decided whether to use feature selection techniques or not. Feature selection is productive in reducing dimensionality, taking away irrelevant and noisy data and simplifying models without much loss of the total information. In this work, *Correlation-based Feature Selection* (CFS) is utilized for feature selection. CFS is a fast algorithm, which makes it applicable to large datasets. It evaluates the efficacy of individual features for predicting the class in the company of the level of inter-correlation among them using Eq. (5) [53]:

$$M_S = \frac{k\overline{r_{fc}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}, \quad (5)$$

where M_S is the score of a feature subset S, including k features, $\overline{r_{fc}}$ the average feature-class correlation, and $\overline{r_{ff}}$ the average feature-feature correlation. If each member of a subset of features is extremely correlated with the class feature while they are uncorrelated to each other, the subset receives a high score. CFS distinguishes irrelevant features as they are not correlated with the class. Moreover, it handles redundant attributes as they are highly correlated with one or more of the other features. However, CFS cannot recognize an intense interaction among features since they are treated independently. It merely identifies suitable features at moderate levels of interaction [53].

Applying RSM to parameter tuning makes it possible to consider the use of the feature selection method as a factor. It allows deciding whether to utilize the feature selection or not when the parameters of the algorithm are tuned. Since the novel approach uses the two methods simultaneously, the possible interaction between them can be distinguished.

3.2.2. Model evaluating

To prevent overfitting and evaluate the performance of each algorithm, it is necessary to re-apply C4.5 and CARD to a completely independent testing dataset. In order to achieve more accurate results, the 10-fold cross validation method is employed. Moreover, each classifier is evaluated in terms of *accuracy, sensitivity (or recall), specificity, precision, and F-measure* using indicators from the confusion matrix displayed in Figure 3 and the following mathematical equations according to [34] and [39].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (6)$$

		Predicted class	
		Positive	Negative
Actual class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 3. Confusion matrix.

$$Sensitivity\ (or\ Recall\ or\ true\ positive\ rate) = \frac{TP}{TP+FN}, \quad (7)$$

$$Specificity\ (true\ negative\ rate) = \frac{TN}{FP+TN}, \quad (8)$$

$$Precision = \frac{TP}{TP+FP}, \quad (9)$$

$$F-measure = \frac{2*Recall*Precision}{Recall+Precision} = \frac{2*TP}{2*TP+FP+FN}. \quad (10)$$

Another metric for determining the quality of a data mining model is the *Receiver Operating Characteristic* (ROC) curve. It is constructed by mapping *True Positive Rate* (TPR), which is '*sensitivity*', and the *false positive rate* (FPR), which is '$1-specificity$'. In the study, the Area Under the ROC Curve (AUC) is used as a measure of the quality of the applied classifiers. The area closer to 1, the higher the quality of the classifier [39]. Moreover, *Size of Tree* (SoT) is used to measure the simplicity and interpretability of the results. A smaller tree generated by a decision tree represents a simpler set of If-Then rules. Also, the *support and confidence* are used as objective *interestingness* measures of extracted rules. Support of the rule $A \Rightarrow B$ is the percentage of transactions in a dataset in which the antecedent-consequent pair occurs concurrently (Eq. (11)). The rule confidence indicates the proportion of cases including the consequent among those comprising the antecedent; thus, it illustrates the rule's reliability (Eq. (12)) [34]. To calculate the support, the entire transactions in a dataset are counted; however, if the total number of instances in a given class is considered, then another measure can be calculated as Eq. (13), which the authors have called *coverage*. In fact, it shows the power of a rule in recognizing a particular class.

$$Support\,(A \Rightarrow B) = P\,(A \cup B)\,, \quad (11)$$

$$Confidence(A \Rightarrow B) = P(A|B) = \frac{Support(A \cup B)}{Support(A)}, \quad (12)$$

$$Coverage(A \Rightarrow B) = P(A \cup B|C) = \frac{Support(A \cup B)}{Support(C)}. \quad (13)$$

The support-confidence framework cannot evaluate the real strength/lack of strength of the correlation between A and B. Therefore, alternative indicators such as *cosine* can be useful in detecting interesting data relationships. The cosine of A and B defined as Eq. (14) ranges from 0 to 1. The higher value of the measure demonstrates a stronger relationship between A and B [34].

$$\text{Cosine}(A \Rightarrow B) = \frac{P(A \cup B)}{\sqrt{P(A)*P(B)}} = \sqrt{P(A|B)*P(B|A)} = \frac{Support(A \cup B)}{\sqrt{Support(A)*Support(B)}}. \quad (14)$$

All classifiers are compared with each other based on the above-mentioned evaluation criteria.

3.3. Post-modeling

3.3.1. Visualization

To make the results and finding of this research clearer and more understandable for healthcare policy-makers and other end-users, C4.5 and CART performance metrics were picturized. Moreover, the discovered patterns were visualized in the form of a tree.

3.3.2. Results evaluation

The extracted knowledge from the data mining framework was presented to the experts of the Iran Health Insurance Organization. Then, their opinions on the discovered patterns and the generated rules for polypharmacy identification were received. Finally, the results were modified based on their feedbacks.

4. Results and discussion

The experimental results of applying the data mining framework on GP's prescriptions dataset are presented and discussed in this section. Recalling the parameters of the CART algorithm mentioned in Table 2, after some preliminary experiments, the region of the *Minimum Number of Instances per Leaf (MNIL)* was considered to be between 1 and 49. To create an appropriate CCD design, the low and high levels of the parameter were set to 8 and 42, respectively. Moreover, the low and high levels of the parameter of the *percentage of training data used to construct the tree (PTD)* were adjusted to 0.312 and 0.882, respectively. Furthermore, 5 central points were selected and α was assigned the value of 1.414. Considering that *using 1 SE rule to make pruning decision (1SEr)* was set to either *false* or *true* and deciding on the use of CFS as the *feature selection* method was set to either *yes* or *no*, we reached

a CCD for RSM which required 52 runs of the CART in WEKA. The factors and corresponding responses, i.e., AUC and SoT, were analyzed by RSM using Minitab software. The stepwise method with default settings was used to recognize a useful subset of predictors during each step. The results of the ANOVA analysis of AUC of CART algorithm demonstrate that the overall model is significant at $p < 0.001$ and $R^2 = 0.83$, $R^2_{\text{adj}} = 0.81$, and $R^2_{\text{pred}} = 0.76$. The lack-of-fit test is not significant ($p = 0.330$), which means the model has a good quality of fit. At a confidence level of 99.9%, all of the main factors other than FS are significant. However, FS is significant at $p < 0.02$. On the other hand, only the quadratic term of PTD is statistically relevant at a confidence level of 99%. Also, its interaction with the MNIL is significant too. Similarly, the ANOVA of CCD for a total of 52 runs of SoT of CART shows that the overall model is significant with $p < 0.001$ and $R^2 = 0.68$, $R^2_{\text{adj}} = 0.65$, and $R^2_{\text{pred}} = 0.59$. Besides, the lack-of-fit test was not significant ($p = 0.665$), demonstrating the model goodness of fit. At a confidence level of 99.9%, just FS is not significant. Moreover, only the quadratic term of MNIL is significant. Furthermore, its interaction with $1SEr$ is statistically considerable. It can be concluded that the feature selection did not affect the tree size generated by CART. The optimum values of the parameters of CART were calculated to maximize AUC and minimize SoT using the desirability function approach. The optimum values of $MNIL$ and PTD are 33 and 0.82, respectively; however, their default values are 2 and 1, respectively. The optimum value of $1SEr$ is false, which is its default value. Besides, feature selection is recommended. The response surface plot of AUC and SoT demonstrates the curvature from second-order effects.

Similarly, the above steps were repeated for parameter tuning of C4.5. Considering $\alpha = 1.414$, the low and high levels of $MNIL$ were considered as 8 and 42, respectively, to cover its region of interest as shown in [1,49]. Moreover, the low and high levels of Confidence Factor (CF) were set to 0.074 and 0.426, respectively, to cover its range of (0,0.5]. Furthermore, Binary Splitting (BS) was set to either *false* or *true*, and deciding on the use of CFS was set to either *yes* or *no*. Considering 5 central points, a CCD for RSM, which entails 52 runs, was generated. ANOVA table of AUC of C4.5 depicts that the overall model is significant at $p < 0.001$ and $R^2 = 0.66$, $R^2_{\text{adj}} = 0.61$, and $R^2_{\text{pred}} = 0.46$, thus indicating that the model has quite a good quality of fit. The p-Value for the lack of fit test cannot be calculated as the pure error is zero. Both MNIL and CF are statistically relevant, considering a significance level of 0.01. Meanwhile, only the quadratic term of MNIL is significant at $p < 0.001$. The fascinating finding is that there is an interaction between MNIL and FS at $p < 0.001$,

```
APrPh < 5.135
|   PPaAG3 < 0.058055: 1 (152.0/110.0); R1
|   PPaAG3 >= 0.058055
|   |   APrPh < 2.075: 0 (289.0/91.0); R2
|   |   APrPh >= 2.075
|   |   |   PPaAG2 < 0.0670405: 0 (263.0/132.0); R3
|   |   |   PPaAG2 >= 0.0670405: 1 (152.0/115.0); R4
APrPh >= 5.135
|   PFPa < 0.7051145: 1 (1089.0/293.0); R5
|   PFPa >= 0.7051145
|   |   PPaAG1 < 0.0226805: 0 (66.0/10.0); R6
|   |   PPaAG1 >= 0.0226805: 1 (22.0/13.0); R7
```

Figure 4. The decision tree generated by tuned CART.

which demonstrates that the hyper-parameter should be tuned along with the feature selection strategy because they have an interaction with each other. This issue is not recognizable in other popular methods of hyper-parameter tuning. The ANOVA table for RSM of SoT generated by C4.5 shows that the overall model is significant at $p < 0.001$ and $R^2 = 0.72$, $R^2_{\text{adj}} = 0.69$, and $R^2_{\text{pred}} = 0.57$, thus proving that the model is fitted quite well. The results show that only CF is not significant at $p < 0.001$; however, it is significant at $p < 0.1$. Considering the quadratic term, only MNIL is significant. At a confidence level of 99.9%, the interaction between MNIL and FS is statistically meaningful. Again, it can be claimed that the hyper-parameter cannot be optimized independently from the feature selection strategy, as there is a considerable interaction, which is a contribution of this study. Considering the maximization of AUC and the minimization of SoT, the optimum values of the parameters of C4.5 differ from their default values. The optimum values of MNIL and CF are 42 and 0.31, respectively. Moreover, feature selection was not suggested and the BS did not affect the performance of C4.5.

The decision tree created by the tuned CART is shown in Figure 4. The SoT is 13, which consists of 5 features and 7 leaves. Each leaf provides a rule for profiling GPs' prescribing patterns. For each rule, the first number in parenthesis indicates the total number of instances (GPs) identified by the rule, while the second one is the number of instances by which the rule is misclassified (number of instances/number of misclassifications). For example, it can be seen that rule No. 5 (R5) distinguished 1089 GPs who have prescribed more than 3 drugs per encounter averagely. However, 293 GPs who prescribed normally were misclassified by the rule. More information on the 4 rules that identified polypharmacy is given in Table 3. It can be seen that only R5, which consists of only two features, has been

Table 3. Rules generated by CART for identifying polypharmacy,

Rule no.	Rule	No. of features	Support (%)	Coverage (%)	Confidence (%)
R1	APrPh < 5.135 and PPaAG3 < 0.058055	2	2.1	3.8	27.6
R4	$2.075 \leq$ APrPh < 5.135 and PPaAG2 ≥ 0.0670405 and PPaAG3 > 0.058055	3	1.8	3.3	24.3
R5	APrPh ≥ 5.135 and PFPa < 0.7051145	2	39.2	71.3	73.1
R7	APrPh ≥ 5.135 and PPaAG1 ≥ 0.0226805 and PFPa ≥ 0.7051145	3	0.4	0.8	40.9

able to detect more than 71% of polypharmacy cases with a confidence of 73%.

Similarly, the decision tree generated by the tuned C4.5 is depicted in Figure 5. The SoT is 23, including 7 leaves and 9 features. The decision tree is relatively more complex than the one created by the tuned CART. However, there are only 4 rules for identifying polypharmacy that is similar to CART. Table 4 gives more information on the *interestingness* of these rules. Only $R11$ composed of only two features identified about 71% of the cases of polypharmacy. It has more than 78% confidence.

```
APrPh <= 5.13
|   PPaAG1 <= 0.300884
|   |   PPaAG6 <= 0.243055
|   |   |   PMPa <= 0.235294: 0 (143.0/22.0); R1
|   |   |   PMPa > 0.235294
|   |   |   |   TNPh <= 23
|   |   |   |   |   PPaAG2 <= 0.066815: 0 (224.0/95.0); R2
|   |   |   |   |   PPaAG2 > 0.066815
|   |   |   |   |   |   TNPr <= 32: 0 (47.0/21.0); R3
|   |   |   |   |   |   TNPr > 32: 1 (100.0/21.0); R4
|   |   |   |   TNPh > 23
|   |   |   |   |   PPaAG4 <= 0.179487: 1 (59.0/26.0); R5
|   |   |   |   |   PPaAG4 > 0.179487: 0 (507.0/147.0); R6
|   |   PPaAG6 > 0.243055
|   |   |   PPaF3 <= 0.046875: 1 (140.0/30.0); R7
|   |   |   PPaF3 > 0.046875: 0 (47.0/20.0); R8
|   PPaAG1 > 0.300884: 0 (37.0); R9
APrPh > 5.13
|   PMPa <= 0.040816: 0 (46.0/1.0); R10
|   PMPa > 0.040816
|   |   PPaAG1 <= 0.261718: 1 (1422.0/308.0); R11
|   |   PPaAG1 > 0.261718: 0 (25.0/6.0); R12
```

Figure 5. The decision tree generated by tuned C4.5.

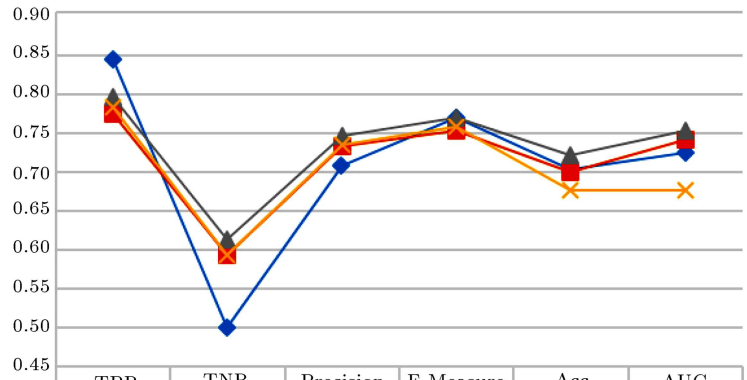
Furthermore, the rules generated by CART and C4.5 and recognized polypharmacy are expressed in a verbal form in order of *interestingness* measures as follows to make them more transparent for healthcare experts and policy-makers.

- CART:
  - R5: If the *average number of dispensed prescriptions per pharmacy* of a GP is greater than or equal to 5.135 and the *percentage of female patients* is not too much (approximately less than 70%), then polypharmacy occurs.
  - R1: If *the average number of dispensed prescriptions per pharmacy* of a GP is less than 5.135 and the percentage of his young adult *patients aged between* 15 and 29 years old is small (approximately less than 6%), then he prescribes excessively (more than 3 drugs per encounter).
  - R4: If *the average number of dispensed prescriptions per pharmacy* for a GP is between 2.075 and 5.135 and the percentage of his patients aged between 5 and 14 years old and the percentage of his patients aged between 15 and 29 years old are not small (approximately more than 6.7% and 5.8% respectively), then polypharmacy occurs.
  - R7: If *the average number of dispensed prescriptions per pharmacy* for a GP is greater than or equal to 5.135, the percentage of his patients aged less than 5 years old is not very small (approximately more than 2.2%), and the percentage of *female patients* is not too high (approximately less than 70%), then he prescribes abnormally.
- C4.5:

Table 4. Rules generated by C4.5 for identifying polypharmacy.

Rule no.	Rule	No. of features	Support (%)	Coverage (%)	Confidence (%)
R4	APrPh $\leq$ 5.13 and PPaAG1 $\leq$ 0.300884 and PPaAG2 $>$ 0.066815 and PPaAG6 $\leq$ 0.243055 and PMPa $\geq$ 0.235294 and TNPh $\leq$ 23 and TNPr $>$ 32	7	2.82	4.8	79.0
R5	APrPh $\leq$ 5.13 and PPaAG1 $\leq$ 0.300884 and PPaAG4 $\leq$ 0.179487 and PPaAG6 $\leq$ 0.243055 and PMPa $\geq$ 0.235294 and TNPh $>$ 23	6	1.18	2.0	55.9
R7	APrPh $\leq$ 5.13 and PPaAG1 $\leq$ 0.300884 and PPaAG6 $>$ 0.243055 and PPaF3 $\leq$ 0.046875	4	3.93	6.7	78.6
R11	APrPh $>$ 5.13 and PPaAG1 $\leq$ 0.261718 and PMPa $\geq$ 0.040816	3	39.83	67.8	78.3

- R11: If *the average number of dispensed prescriptions per pharmacy* of a GP is greater than 5.13, the percentage of his patients aged less than 5 years old is approximately less than one quarter ($\leq$26.17%), and the *percentage of male patients* is not too insignificant (approximately more than 4.1%), then polypharmacy occurs.
- R7: If *the average number of dispensed prescriptions per pharmacy* of a GP is less than or equal to 5.13, the percentage of his patients aged less than 5 years old is less than or equal to 30%, and the percentage of his elder patients aged between 70 and 94 years old is roughly more than 24.3%, and the *percentage of patients in the rural insurance fund* is insignificant (approximately less than 4.7%), then he prescribes excessively.
- R4: If the *average number of dispensed prescriptions per pharmacy* of a GP is less than or equal to 5.13, the percentage of his child patients (less than 5 years old) and elderly patients (between 70 and 94) is roughly less than or equal to 30% and 24.3%, respectively, while the percentage of his young adult patients (between 15 and 29 years old) is not insignificant (is more than 6%), the *percentage of male patients* is approximately higher than or equal to 23.5%, the *total number of pharmacies* that dispensed his prescriptions is less than 24, and the *total number of prescriptions* is more than 32, then the GP prescribes abnormally.
- R5: If *the average number of dispensed prescriptions per pharmacy* of a GP is less than or equal to 5.13, the percentage of his child patients (less than 5 years old), middle-aged patients (between 30 and 49), and elderly patients (between 70 and 94 years old) is roughly less than or equal to 30%, 18%, and 24.3%, respectively, the *percentage of male patients* is approximately higher than or equal to 23.5%, and the *total number of pharmacies* that dispensed his prescriptions is larger than 23, then polypharmacy occurs.



	TPR	TNR	Precision	F-Measure	Acc	AUC
Tuned CART	0.845	0.5	0.708	0.77	0.703	0.725
Tuned C4.5	0.775	0.594	0.733	0.753	0.701	0.742
Default CART	0.797	0.614	0.747	0.771	0.721	0.753
Default C4.5	0.784	0.594	0.735	0.759	0.676	0.676

Figure 6. Performance measures of applied decision trees.

On the other hand, various performance measures have been calculated for the applied algorithms with the tuned configurations as well as default ones, the summary of which is presented in Figure 6. It can be seen that parameter tuning had a significant improvement in the AUC and the accuracy of the C4.5 and resulted in a smaller tree. The default setting of C4.5 generated a huge tree with 443 nodes and 233 leaves. In other words, by adjusting the parameters, not only the performance indicators of the C4.5 were improved but also a smaller and more understandable tree was created. The CART parameters setting resulted in a smaller tree, whereas the untuned CART generated a tree with 21 nodes and 11 leaves. However, it caused a decline in AUC and TNR (specificity), while it increased TPR (recall). It is important to note that in this case, larger TPR is preferable to larger TNR. Because the costs of misclassification are not the same in the two classes, in other words, if a GP who prescribes normally be misclassified as a physician who prescribes excessively, then it will eventually lead to more training or some strict policies, causing greater rationalization of drugs. Conversely, if a GP with an anomalous prescribing pattern is not correctly distinguished, he may keep his abnormal behavior.

Since end-users should evaluate whether or not the extracted rules are interesting and useful, they were subjected to judgment by experts from the *Iran Healthcare Insurance Organization*. Given that they did not experience use of such a data mining approach to examine the rational use of drugs, the results were generally welcomed by them. Moreover, many extraction rules were considered useful. For example, they confirmed that the existence of a strong relationship between a physician and some pharmacies could lead to excessive drug prescription and abnormal behavior. Although some rules such as R5 of C4.5, stating that the association with more pharmacies (more than 23) might cause polypharmacy, are not tangible. In other words, it was expected that a partnership with fewer pharmacies would cause polypharmacy. In general, the experts preferred more easy-to-follow CART rules that made it more perceptible.

5. Conclusions

World Health Organization (WHO) and International Network of Rational Use of Drug (INRUD) suggest that the '*average number of drugs prescribed per encounter*' could be used to appraise the extent of polypharmacy. Reduction of polypharmacy exposure can reduce risks of adverse drug reactions, adverse drug events, prescribing potentially inappropriate medications, mortality, and other dangerous consequences. In this paper, to the best of authors' knowledge, the first data mining framework was developed to study polypharmacy. The approach contains three steps: pre-modeling, modeling, and post-modeling. Pre-modeling includes collecting data, handing inconsistent data, integrating different data sources, and extracting several features. In the modeling step, C4.5 and Classification and Regression Tree (CART) were applied to build models to elicit a set of rules for polypharmacy detection and explanation. To improve the performance of utilized decision trees, their parameters were tuned by Response Surface Method (RSM) to maximize Area Under the ROC Curve (AUC) and minimize SoT using the desirability function approach along with the Correlation-based Feature Selection (CFS) method of feature selection simultaneously. The results demonstrated that the hyper-parameter cannot be independently optimized from the feature selection

strategy, as there is a considerable interaction between them. Moreover, the tuned C4.5 created a smaller tree and improved the accuracy and AUC compared to the untuned one. Furthermore, tuned CART generated a smaller tree; however, its TNR (specificity), accuracy, and AUC declined. On the other hand, its True Positive Rate (TPR) (sensitivity) increased to 84.5%. TPR measures the algorithm capability of detecting polypharmacy. According to both C4.5 and CART, '*the average number of dispensed prescriptions per pharmacy*' is the key feature for identifying polypharmacy. It shows the extent of the relationship between physicians and pharmacies. Moreover, patient age groups and patient gender are effective features. However, physician gender and experience did not have a noticeable effect on prescribing patterns. In the final step, the extracted rules for polypharmacy detection were visualized and simplified and then, were presented to domain experts to evaluate them. In general, they preferred the rules created by CART because of its transparency and simplicity.

This study demonstrates the capabilities of the data mining approach to analyze and describe the prescribing pattern of General Practitioners (GPs) in association with polypharmacy. The approach determines GPs' profile as a set of If-Then rules; therefore, it can be used as a part of an automatic drug prescription monitoring system. It could also be helpful for healthcare managers and policy-makers to understand the factors associated with physicians' prescribing pattern to take suitable actions and improve appropriate prescribing affairs to reduce the average number of drugs prescribed per encounter, optimize drug regimens for patients, support high-quality prescription, and develop or revise prescribing guidelines.

In the future studies, in the case of access to physicians' supplementary data including age, academic records, work history, and place of prescription, they can be included in the data mining procedure. Also, Over The Counter (OTC) drugs, dietary supplements, and complementary and alternative medicine (CAM) can be considered in the analysis; however, gathering this kind of data is very difficult and time-consuming. Furthermore, medical diagnosis and inpatient prescription data are involved in enriching the results. Since different cutoff numbers are used for defining polypharmacy, they can be applied to sensitivity analysis. In this research, the threshold of three drugs is used in the definition of polypharmacy. The consideration of five or more drugs as definition threshold results in an imbalanced dataset, thus making the analysis more technically appealing. Besides, the proposed method can be generalized to analyze WHO's other measures of rational use of drugs such as *percentage of prescriptions containing antibiotics, percentage of encounters prescribed injection*, and *percentage of drugs prescribed from PHC formulary*. The values of some features can be fuzzified to make the results more understandable for the healthcare managers and policy-makers.

Acknowledgment

The authors would like to express their special thanks of gratitude to the *National Center for Health Insurance Research* of the *Iran Health Insurance Organization* for providing the raw dataset. They are particularly grateful for the assistance given by Dr. Ali Shojaei, Mr. Mehdi Eghlidi, and Mrs. Zahra Shahali. Also, they wish to acknowledge the valuable help provided by Mrs. Maryam Arab for preprocessing the insurers' demographic data.

References

1. Lu, Y., Hernandez, P., Abegunde, D., et al. "Medicine expenditures", In *The World Medicines Situation 2011.*, 3th Edn., pp. 35–38, World Health Organization, Geneva, Switzerland (2011).
2. Ofori-Asenso, R. "A closer look at the World Health Organization's prescribing indicators", *J. Pharmacol. Pharmacother.*, **7**(1), p. 51 (2016).
3. Mao, W., Vu, H., Xie, Z., et al. "Systematic review on irrational use of medicines in China and Vietnam", *PLoS One*, **10**(3), pp. 1–16 (2015).
4. Holloway, K. and Van Dijk, L. "Rational use of medicines", In *The World Medicines Situation 2011.*, 3th Edn., World Health Organization, Geneva, Switzerland (2011).
5. Holloway, K.A. and Henry, D. "WHO essential medicines policies and use in developing and transitional countries: An analysis of reported policy implementation and medicines use surveys", *PLoS Med.*, **11**(9), pp. 1–16 (2014).
6. WHO "The World Health Report 2002: reducing risks, promoting healthy life", World Health Organization, Geneva, Switzerland (2002).
7. WHO, "How to investigate drug use in health facilities: selected drug use indicators", WHO/DAP/93.1, World Health Organization (1993).
8. Rambhade, S., Shrivastava, A., Rambhade, A., et al. "A survey on polypharmacy and use of inappropriate medications", *Toxicol. Int.*, **19**(1), pp. 68–73 (2012).
9. WHO "The world medicines situation", (No. WHO/EDM/PAR/2004.5), World Health Organization, Geneva, Switzerland (2004).
10. Jokanovic, N., Tan, E.C.K., Dooley, M.J., et al. "Prevalence and factors associated with polypharmacy in long-term care facilities: A systematic review", *J. Am. Med. Dir. Assoc.*, **16**(6), pp. 535–e1 (2015).
11. Qi, K., Reeve, E., Hilmer, S.N., et al. "Older peoples' attitudes regarding polypharmacy, statin use and willingness to have statins deprescribed in Australia", *Int. J. Clin. Pharm.*, **37**(5), pp. 949–957 (2015).

12. Reeve, E., Wiese, M.D., Hendrix, I., et al. "People's attitudes, beliefs, and experiences regarding polypharmacy and willingness to deprescribe", *J. Am. Geriatr. Soc.*, **61**(9), pp. 1508–1514 (2013).
13. Galazzi, A., Lusignani, M., Chiarelli, M.T., et al. "Attitudes towards polypharmacy and medication withdrawal among older inpatients in Italy", *Int. J. Clin. Pharm.*, **38**(2), pp. 454–461 (2016).
14. Quinn, K.J. and Shah, N.H. "A dataset quantifying polypharmacy in the United States", *Sci. Data*, **4**, p. 167 (2017).
15. Hovstadius, B. and Petersson, G. "Factors leading to excessive polypharmacy", *Clin. Geriatr. Med.*, **28**(2), pp. 159–172 (2012).
16. Bjerrum, L., Sϕgaard, J., Hallas, J., et al. "Polypharmacy in general practice: differences between practitioners", **49**(440), pp. 195–198 (1999).
17. Anthierens, S., Tansens, A., Petrovic, M., et al. "Qualitative insights into general practitioners views on polypharmacy", *BMC Fam. Pract.*, **11**(1), p. 65 (2010).
18. O'Dwyer, M., Peklar, J., McCallion, P., et al. "Factors associated with polypharmacy and excessive polypharmacy in older people with intellectual disability differ from the general population: a cross-sectional observational nationwide study", *BMJ Open*, **6**(4), e010505 (2016).
19. Ie, K., Felton, M., Springer, S., et al. "Physician factors associated with polypharmacy and potentially inappropriate medication use", *J. Am Board Fam Med.*, **30**(4), pp. 528–536 (2017). DOI: 10.3122/jabfm.2017.04.170121
20. Slater, N., White, S., Venables, R., et al. "Factors associated with polypharmacy in primary care: a cross-sectional analysis of data from the english longitudinal study of ageing (ELSA)", *BMJ Open*, **8**(3), e020270 (2018).
21. Hogerzeil, H.V. "Promoting rational prescribing: an international perspective", *Br. J. Clin. Pharmacol.*, **39**(1), pp. 1–6 (1995).
22. Kumari, R., Idris, M.Z., Bhushan, V., et al. "Assessment of prescription pattern at the public health facilities of Lucknow district", *Indian J. Pharmacol.*, **40**(6), pp. 243–247 (2008).
23. Desalegn, A.A. "Assessment of drug use pattern using WHO prescribing indicators at Hawassa University teaching and referral hospital, south Ethiopia: A cross-sectional study", *BMC Health Serv. Res.*, **13**(1), p. 170 (2013).
24. Atif, M., Sarwar, M.R., Azeem, M., et al. "Assessment of core drug use indicators using WHO/INRUD methodology at primary healthcare centers in Bahawalpur, Pakistan", *BMC Health Serv. Res.*, **16**(1), p. 684 (2016).
25. Bastani, P., Barfar, E., Rezapour, A., et al. "Rational prescription of drug in Iran: Statistics and trends for policymakers", *JHMI J. Heal. Manag. Informatics*, **5**(April), pp. 35–40 (2018).
26. Cerrito, P. "Application of data mining for examining polypharmacy and adverse effects in cardiology patients", *Cardiovasc. Toxicol.*, **1**(3), pp. 177–179 (2001).
27. Ji, Y., Ying, H., Tran, J., et al. "A functional temporal association mining approach for screening potential drug-drug interactions from electronic patient databases", *Informatics Heal. Soc. Care*, **41**(4), pp. 387–404 (2016).
28. Held, F., Le Couteur, D.G., Blyth, F.M., et al. "Polypharmacy in older adults: Association rule and frequent-set analysis to evaluate concomitant medication use", *Pharmacol. Res.*, **116**(April), pp. 39–44 (2017).
29. Poluzzi, E., Raschi, E., Piccinni, C., et al. "Data mining techniques in pharmacovigilance: Analysis of the publicly accessible FDA Adverse Event Reporting System (AERS)", In *Data Mining Applications in Engineering and Medicine*, Karahoca, A., Ed., BoD (2012).
30. Vilar, S., Friedman, C., and Hripcsak, G. "Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media", *Brief. Bioinform.*, **19**(5), pp. 1–15 (2018).
31. Sohail, M.N., Jiadong, R., Uba, M.M., et al. "A comprehensive looks at data mining techniques contributing to medical data growth: A survey of researcher reviews", *Recent Dev. Intell. Comput. Commun. Devices*, **752**, pp. 21–26 (2019).
32. Kann, I.C., Lundqvist, C., and Luras, H. "Polypharmacy among the elderly in a list-patient system", *Drugs-Real World Outcomes*, **2**(3), pp. 193–198 (2015).
33. IBM, "ASUM-DM", URL http://gforge.icesi.edu.co/ASUM-DM_External/index.htm. (2015)
34. Han, J., Kamber, M., and Pei, J., *Data Mining: Concepts and Techniques*, Elsevier (2011).
35. WHO "WHO methods and data sources for global burden of disease estimates 2000-2016", World Health Organization, Geneva, Switzerland (2018).
36. Wu, X., Kumar, V., Quinlan, J.R., et al. "Top 10 algorithms in data mining", *Knowl. Inf. Syst.*, **14**(1), pp. 1–37 (2008).
37. Guidotti, R., Monreale, A., Ruggieri, S., et al. "A survey of methods for explaining black box models" (2018). https://doi.org/10.48550/arXiv.1802.01933
38. Quinlan, J.R., *C4. 5 Programs for Machine Learning*, Morgan Kaufmann, San Mateo, USA (2014).
39. Witten, I.H., Frank, E., Hall, M.A., et al., *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd Edn., Morgan Kaufmann (2016).
40. Lamy, J.B., Ellini, A., Ebrahiminia, V., et al. "Use of the C4.5 machine learning algorithm to test a clinical guideline-based decision support system", *Stud. Health Technol. Inform.*, **136**, pp. 223–228 (2008).
41. Tu, M.C., Shin, D., and Shin, D. "A comparative study of medical data classification methods based on decision tree and bagging algorithms", *8th IEEE*

Int. Symp. Dependable, Auton. Secur. Comput. DASC 2009, Chengdu, China, pp. 183–187 (2009).

42. Lavanya, D. and Rani, K.U. "Performance evaluation of decision tree classifiers on medical datasets", *Int. J. Comput. Appl.*, **26**(4), pp. 1–4 (2011).
43. Solanki, A.V. "Data mining techniques using WEKA classification for sickle cell disease", *Int. J. Comput. Sci. Inf. Technol.*, **5**(4), pp. 5857–5860 (2014).
44. Wiharto, W., Kusnanto, H., and Herianto, H. "Interpretation of clinical data based on C4.5 algorithm for the diagnosis of coronary heart disease", *Healthc. Inform. Res.*, **22**(3), pp. 186–195 (2016).
45. Pecchia, L., Melillo, P., and Bracale, M. "Remote health monitoring of heart failure with data mining via CART method on HRV features", *IEEE Trans. Biomed. Eng.*, **58**(3), pp. 800–804 (2011).
46. Diessner, J., Wischnewsky, M., Stüber, T., et al. "Evaluation of clinical parameters influencing the development of bone metastasis in breast cancer", *BMC Cancer*, **16**(1), pp. 307–320 (2016).
47. Zimmerman, R.K., Balasubramani, G.K., Nowalk, M.P., et al. "Classification and regression tree (CART) analysis to predict influenza in primary care patients", *BMC Infect. Dis.*, **16**(1), pp. 503–519 (2016).
48. Cheng, Z., Nakatsugawa, M., Hu, C., et al. "Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy", *Adv. Radiat. Oncol.*, **3**(3), pp. 346–355 (2018).
49. Mburu, J.W., Kingwara, L., Ester, M., et al. "Use of classification and regression tree (CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor tuberculosis treatment outcomes and associated risk factors", *J. Clin. Tuberc. Other Mycobact. Dis.*, **11**(1), pp. 10–16 (2018).
50. Lujan-Moreno, G.A., Howard, P.R., Rojas, O.G., et al. "Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study", *Expert Syst. Appl.*, **109**(1), pp. 195–205 (2018).
51. Myers, R.H., Montgomery, D.C., and Anderson-Cook, C.M., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 4th Edn., John Wiley & Sons, New York, USA (2016).
52. Montgomery, D.C., *Design and Analysis of Experiments*, 10th Edn., John Wiley & Sons, USA (2019).
53. Hall, M.A. and Holmes, G. "Benchmarking attribute selection techniques for data mining", *IEEE Trans. Knowl. Data Eng.*, **15**(6), pp. 1437–1447 (2003).

Biographies

Morteza Moradi is a faculty member at the Faculty of Industrial Engineering, Urmia University of Technology, Iran. He received his PhD in Industrial Engineering from Sharif University of Technology in 2021. His current research focuses on data mining for healthcare and financial applications.

Mohammad Modarres is a Professor at Department of Industrial Engineering, Sharif University of Technology, Iran. He received his PhD in Systems Engineering and Operations Research from University of California, Los Angles (UCLA) in 1975. His research interests are operations research, revenue management, and robust optimization. He has published in European Journal of Operational Research, IEEE Transactions on Power Systems, IEEE Transactions on Reliability, IEEE Transactions on Fuzzy Systems, Naval Research Logistics Quarterly, Fuzzy sets and systems, International Journal of Production Research Journal of Operational Research Society, Transportation Research, Journal of Computer and Operations Research, Computers & Industrial Engineering, Scientia Iranica, and Iranian Journal of Operations Research.

Mohammad Mehdi Sepehri Professor of Healthcare Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University (TMU), Tehran, Iran; Founder and Director of the Center of Excellence in Healthcare Systems Engineering; Founder and Head of the Laboratory for Healthcare Systems Optimization, Engineering, and Informatics; Editor-in-Chief of International Journal of Hospital Research (IJHR), the official peer review publication of Iran University of Medical Sciences. Dr. Sepehri obtained his MSc and PhD in Management Science from the University of Tennessee, Knoxville, Tennessee, USA in 1987 and 1991. His current research focuses on Healthcare Systems Engineering, mHealth, Healthcare Internet of Things, Network Optimization, Lean Healthcare, and Masstech: Technology for mass of people.