

Research Article

Predicting the Physician's Specialty Using a Medical Prescription Database

Mahboube Akhlaghi,¹ Hamed Tabesh,² Behzad Mahaki,³ Mohammad-Reza Malekpour,⁴ Erfan Ghasemi,⁴ and Marjan Mansourian ⁵

¹Department of Biostatistics and Epidemiology, School of Health, and Student Research Committee, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran

²Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

³Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

⁴Non-Communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

⁵Epidemiology and Biostatistics Department, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran

Correspondence should be addressed to Marjan Mansourian; j_mansourian@hlth.mui.ac.ir

Received 20 July 2022; Revised 19 August 2022; Accepted 1 September 2022; Published 16 September 2022

Academic Editor: Sheng Du

Copyright © 2022 Mahboube Akhlaghi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. The present study is aimed at predicting the physician's specialty based on the most frequent two medications prescribed simultaneously. The results of this study could be utilized in the imputation of the missing data in similar databases. **Patients and Methods.** The research is done through the KAY-means for MIXed LARge datasets (KAMILA) clustering and random forest (RF) model. The data used in the study were retrieved from outpatients' prescriptions in the second populous province of Iran (Khorasan Razavi) from April 2015 to March 2017. **Results.** The main findings of the study represent the importance of each combination in predicting the specialty. The final results showed that the combination of amoxicillin-metronidazole has the highest importance in making an accurate prediction. The findings are provided in a user-friendly R-shiny web application, which can be applied to any medical prescription database. **Conclusion.** Nowadays, a huge amount of data is produced in the field of medical prescriptions, which a significant section of that is missing in the specialty. Thus, imputing the missing variables can lead to valuable results for planning a medication with higher quality, improving healthcare quality, and decreasing expenses.

1. Introduction

Population aging, urbanization, and consequently noncommunicable disease (NCD) prevalence have led to the escalation of medication consumption per capita globally [1]. According to the United Nations Office on Drugs and Crime (UNODC) report in 2018, about 80% of the population between the ages of 15 and 64 in Asia are taking medications [2]. Considering the prevalence of NCDs and their associated prescribed healthcare services, insurance claimed data, even in a short period, provides an invaluable source of big data.

Heterogeneity is one of the big data characteristics, which leads to obtaining biased or even erroneous results [3, 4]. Although clustering on big data to create homoge-

neous clusters is an essential step for the subsequent analyses [5], it also produces intermediary results that researchers could use to interpret the relation of variables within clusters [6, 7]. In addition, missing data may affect the valence of the data in big data applications [8]. Therefore, imputing missing data leads to reasonable, valuable, and unbiased results. Among numerous studies that investigated medical prescriptions in various ways [9–11], a considerable proportion reported a significant rate of up to one-third for the physician's specialty missingness [12–17]. Despite the provision of a remarkable increase in the accuracy of disease labeling of patients, one of the overlooked applications of studying prescriptions is predicting the physician's specialty.

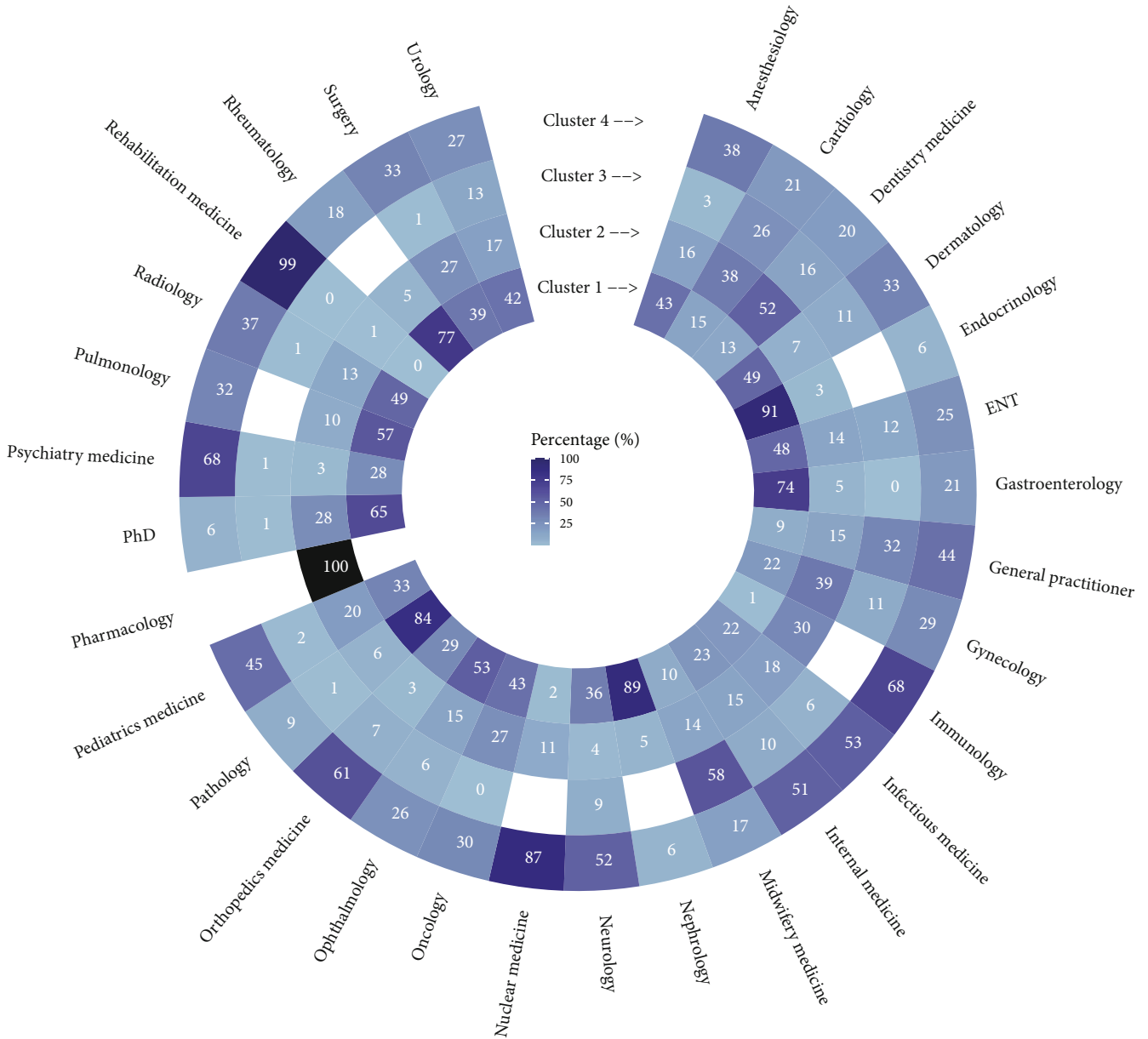


FIGURE 1: The heatmap plot of the specialty variable in four clusters.

To the best of our knowledge, limited studies have been conducted to predict the specialty of physicians using machine learning methods on big data. The present study is aimed at predicting the physician’s specialty based on the most frequent two medications prescribed simultaneously. The results of this study could be utilized in the imputation of the missing data in similar databases.

2. Materials and Methods

2.1. Data Source. The data used in the study had been registered in the second populous province of Iran (Khorasan Razavi) from April 2015 to March 2017. The data included the fields of prescription ID, patient ID, medication, and prescription date, as well as physician’s specialty, gender,

and age. Also, each record of the data represents a single medication of prescription.

2.2. Data Manipulation and Analysis. An essential step at the beginning of the analysis process is data wrangling. In this study, data was retrieved from 4 SQL tables and preprocessed via R language. As the variables in this step contained both quantitative and qualitative measures, KAY-means for MIXed LARGE datasets (KAMILA) clustering was implemented, which is suitable for clustering mixed-type data [18]. In this step, the variables included in the clustering were age of physician, medication, insurance company, prescription month, sex of physician, physician specialty, corticosteroid, and antibiotic. Moreover, specialties with ages between 22 and 85 years old are included in the study. Also, prescriptions with only one

Five highest frequent medicines in each specialty and cluster

Specialty	Most frequent	2nd most frequent	3th most frequent	4th most frequent	5th most frequent	
Cluster 1	Urology	Diclofenac (N7.1)	Tramadol (N9.3)	Cefixime (N4.9)	Ciprofloxacin (N4.3)	Phenazopyridine (N4.2)
	Urology surgery	Diclofenac (N7.1)	Metronidazole (N3.8)	Cefixime (N4.9)	Actinomophen (N3.4)	Betamethasone (N2.6)
	Rheumatology	Prednisolone (N8.9)	Diclofenac (N8.6)	Hydroxychloroquine (N8.4)	Alendronate (N8.1)	Methotrexate (N7.2)
	Rehabilitation medicine	Mefenamic Acid (N8.5)	Cefixime (N7.1)	Diclofenac (N7.1)	Rufolone (N4.8)	Calcium-D (N4.8)
	Radiotherapy	Subitamol (N7.7)	Iproratium (N8.4)	Magnesium Glycinate (N4.8)	Diclofenac (N4.5)	Hyoscine (N4.3)
	Pulmonology	Salbutamol (N4.9)	Theophylline (N8.7)	Salbutamol Fluocason (N8.5)	Dexamethasone (N6.3)	Bronchitolone (N5.9)
	Psychiatry medicine	Sodium Valproate (N8.7)	Alprazolam (N4.2)	Lamotrigine (N4)	Gabapentin (N3.9)	Clonazepam (N3.9)
	PHD	Salbutamol Fluocason (N4.1)	Subitamol (N9.2)	Prochlorperazine (N8.3)	Theophylline G (N8)	Aluminum Mg (N4)
	Pharmacology	Actinomophen (N11.4)	Ketotifen (N6.2)	Subitamol (N3.7)	Anthromycin (N3.7)	Cefixime (N3.6)
	Pediatrics medicine	Dexamethasone (N5.4)	Cefixime (N5)	Dexamethorphan (N4.1)	Sodium Valproate (N4.1)	Betamethasone (N4)
Cluster 2	Urology	Diclofenac (N7.1)	Betamethasone (N3.9)	Vitamin D3 (Calcitriol) (N6.5)	Methylprednisolone (N6.4)	Ibuprofen (N5.5)
	Urology surgery	Diclofenac (N7.1)	Betamethasone (N3.8)	Chloramphenicol (N4.2)	Artificial Tears (N7)	Erythromycin (N4.1)
	Rheumatology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Sodium Chloride (NaCl) Inf (N3.8)	Tranexamic Acid (N3)	Dexamethasone (N3)
	Rehabilitation medicine	Amoxicillin (N6.7)	Acetaminophen (N6.3)	Adult Gold (N8.3)	Bumex (N8.3)	Ciprofloxacin (N5.5)
	Radiotherapy	Levofloxacine (N6.7)	Acetaminophen (N6.5)	Gabapentin (N4.8)	Proparalolol (N4.8)	Diclofenac (N3.3)
	Pulmonology	Amoxicillin (N3.6)	Hyoscine (N12.8)	Clonazepam (N12.2)	Cephalexin (N9.2)	Folic Acid (N9.1)
	Psychiatry medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Lorazepam (N2.9)
	PHD	Acetaminophen (N6.6)	Anthromycin (N6.3)	Cefixime (N8.1)	Theophylline G (N4.6)	Loratadine (N4.6)
	Pharmacology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Fluconazole (N6.2)	Vitamin D3 (Calcitriol) (N6.2)	Aluminum MgS (N4.2)
	Pediatrics medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Loratadine (N2.9)
Cluster 3	Urology	Ciprofloxacin (N4.5)	Oflazoxin (N11.1)	Cefixime (N12.9)	Hydrochlorothiazide (N8.2)	Fosinotide (N4.3)
	Urology surgery	Cefixime (N11.8)	ASA (N10.4)	Metronidazole (N8.9)	Cephalexin (N8.5)	Acetaminophen (N8.2)
	Rheumatology	Folic Acid (N28.4)	Acetaminophen (N6.2)	Meropenem (N5.4)	Lorazepam (N5.3)	ASA (N5.1)
	Rehabilitation medicine	Acetaminophen (N4.9)	Amoxicillin (N9.6)	ASA (N9.1)	Cefixime (N4.7)	Metoprolol (N4.7)
	Radiotherapy	Cefixime (N18.8)	Beclim G Procaine (N7.3)	Penicillin G 3.3 (N6.6)	Doxycycline (N6.1)	Co-Amoxiclav (N5.5)
	Pulmonology	Anthromycin (N28.1)	Acetaminophen (N8.4)	Metronidazole (N7.9)	Nitroglycerin (N6.6)	Metoprolol (N6.2)
	Psychiatry medicine	ASA (N18.6)	Acetaminophen (N16.4)	Folic Acid (N12.9)	Amoxicillin (N9.7)	Metoprolol (N8.1)
	PHD	Amoxicillin (N3.9)	Chlorhexidine (N2.7)	Metronidazole (N2.1)	Co-Amoxiclav (N4.2)	Cefixime (N4.2)
	Pharmacology	Ciprofloxacin (N4.8)	Ondansetron (N3.12)	Amoxicillin (N11.5)	Cefixime (N11.2)	Co-Amoxiclav (N9.2)
	Pediatrics medicine	Metronidazole (N8.9)	Acetaminophen (N7.2)	Cefixime (N7.5)	Amoxicillin (N6.8)	ASA (N6.5)
Cluster 4	Urology	Ciprofloxacin (N4.5)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)
	Urology surgery	Diclofenac (N6.3)	Rufolone (N6.3)	Cefixime (N4.9)	Neproxen (N4.8)	Fosinotide (N4.7)
	Rheumatology	Prednisolone (N8.7)	Magnesium Glycinate (N6.4)	Magnesium Hydroxide (N6)	Metronidazole (N5.4)	Co-La (N5)
	Rehabilitation medicine	Metoprolol (N6.4)	ASA (N20.5)	Acetaminophen (N9.1)	Folic Acid (N8.1)	Nitroglycerin (N4.5)
	Radiotherapy	Sodium Chloride (NaCl) Inf (N28.1)	Magnesium Glycinate (N13.6)	Doxycycline (N13.2)	Hyoscine (N13.2)	Mefenamic Acid (N10)
	Pulmonology	Citalopram (N8.5)	Gabapentin (N8.5)	ASA (N2.8)	Folic Acid (N7.8)	Acetaminophen (N7.3)
	Psychiatry medicine	Cefixime (N4.8)	Amoxicillin (N17.4)	Metronidazole (N12.2)	Co-Amoxiclav (N10.9)	Acetaminophen Colicine (N2.2)
	PHD	Ondansetron (N10.9)	Acetaminophen (N8.5)	Sodium Chloride (NaCl) Topical (N7.9)	Subitamol (N6.1)	Ketotifen (N5.7)
	Pharmacology	Acetaminophen (N12.9)	ASA (N8.2)	Anthromycin (N7.1)	Hydrochlorothiazide (N7.1)	Amoxicillin (N7.2)
	Pediatrics medicine	Diclofenac (N2.6)	Vitamin D3 (Calcitriol) (N9.8)	Betamethasone (N7.6)	Calcium-D (N7.2)	Mefenamic (N7.2)
Cluster 5	Urology	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Urology surgery	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Rheumatology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Sodium Chloride (NaCl) Inf (N3.8)	Tranexamic Acid (N3)	Dexamethasone (N3)
	Rehabilitation medicine	Amoxicillin (N6.7)	Acetaminophen (N6.3)	Adult Gold (N8.3)	Bumex (N8.3)	Ciprofloxacin (N5.5)
	Radiotherapy	Levofloxacine (N6.7)	Acetaminophen (N6.5)	Gabapentin (N4.8)	Proparalolol (N4.8)	Diclofenac (N3.3)
	Pulmonology	Amoxicillin (N3.6)	Hyoscine (N12.8)	Clonazepam (N12.2)	Cephalexin (N9.2)	Folic Acid (N9.1)
	Psychiatry medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Lorazepam (N2.9)
	PHD	Acetaminophen (N6.6)	Anthromycin (N6.3)	Cefixime (N8.1)	Theophylline G (N4.6)	Loratadine (N4.6)
	Pharmacology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Fluconazole (N6.2)	Vitamin D3 (Calcitriol) (N6.2)	Aluminum MgS (N4.2)
	Pediatrics medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Loratadine (N2.9)
Cluster 6	Urology	Ciprofloxacin (N4.5)	Oflazoxin (N11.1)	Cefixime (N12.9)	Hydrochlorothiazide (N8.2)	Fosinotide (N4.3)
	Urology surgery	Cefixime (N11.8)	ASA (N10.4)	Metronidazole (N8.9)	Cephalexin (N8.5)	Acetaminophen (N8.2)
	Rheumatology	Folic Acid (N28.4)	Acetaminophen (N6.2)	Meropenem (N5.4)	Lorazepam (N5.3)	ASA (N5.1)
	Rehabilitation medicine	Acetaminophen (N4.9)	Amoxicillin (N9.6)	ASA (N9.1)	Cefixime (N4.7)	Metoprolol (N4.7)
	Radiotherapy	Cefixime (N18.8)	Beclim G Procaine (N7.3)	Penicillin G 3.3 (N6.6)	Doxycycline (N6.1)	Co-Amoxiclav (N5.5)
	Pulmonology	Anthromycin (N28.1)	Acetaminophen (N8.4)	Metronidazole (N7.9)	Nitroglycerin (N6.6)	Metoprolol (N6.2)
	Psychiatry medicine	ASA (N18.6)	Acetaminophen (N16.4)	Folic Acid (N12.9)	Amoxicillin (N9.7)	Metoprolol (N8.1)
	PHD	Amoxicillin (N3.9)	Chlorhexidine (N2.7)	Metronidazole (N2.1)	Co-Amoxiclav (N4.2)	Cefixime (N4.2)
	Pharmacology	Ciprofloxacin (N4.8)	Ondansetron (N3.12)	Amoxicillin (N11.5)	Cefixime (N11.2)	Co-Amoxiclav (N9.2)
	Pediatrics medicine	Metronidazole (N8.9)	Acetaminophen (N7.2)	Cefixime (N7.5)	Amoxicillin (N6.8)	ASA (N6.5)
Cluster 7	Urology	Ciprofloxacin (N4.5)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)
	Urology surgery	Diclofenac (N6.3)	Rufolone (N6.3)	Cefixime (N4.9)	Neproxen (N4.8)	Fosinotide (N4.7)
	Rheumatology	Prednisolone (N8.7)	Magnesium Glycinate (N6.4)	Magnesium Hydroxide (N6)	Metronidazole (N5.4)	Co-La (N5)
	Rehabilitation medicine	Metoprolol (N6.4)	ASA (N20.5)	Acetaminophen (N9.1)	Folic Acid (N8.1)	Nitroglycerin (N4.5)
	Radiotherapy	Sodium Chloride (NaCl) Inf (N28.1)	Magnesium Glycinate (N13.6)	Doxycycline (N13.2)	Hyoscine (N13.2)	Mefenamic Acid (N10)
	Pulmonology	Citalopram (N8.5)	Gabapentin (N8.5)	ASA (N2.8)	Folic Acid (N7.8)	Acetaminophen (N7.3)
	Psychiatry medicine	Cefixime (N4.8)	Amoxicillin (N17.4)	Metronidazole (N12.2)	Co-Amoxiclav (N10.9)	Acetaminophen Colicine (N2.2)
	PHD	Ondansetron (N10.9)	Acetaminophen (N8.5)	Sodium Chloride (NaCl) Topical (N7.9)	Subitamol (N6.1)	Ketotifen (N5.7)
	Pharmacology	Acetaminophen (N12.9)	ASA (N8.2)	Anthromycin (N7.1)	Hydrochlorothiazide (N7.1)	Amoxicillin (N7.2)
	Pediatrics medicine	Diclofenac (N2.6)	Vitamin D3 (Calcitriol) (N9.8)	Betamethasone (N7.6)	Calcium-D (N7.2)	Mefenamic (N7.2)
Cluster 8	Urology	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Urology surgery	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Rheumatology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Sodium Chloride (NaCl) Inf (N3.8)	Tranexamic Acid (N3)	Dexamethasone (N3)
	Rehabilitation medicine	Amoxicillin (N6.7)	Acetaminophen (N6.3)	Adult Gold (N8.3)	Bumex (N8.3)	Ciprofloxacin (N5.5)
	Radiotherapy	Levofloxacine (N6.7)	Acetaminophen (N6.5)	Gabapentin (N4.8)	Proparalolol (N4.8)	Diclofenac (N3.3)
	Pulmonology	Amoxicillin (N3.6)	Hyoscine (N12.8)	Clonazepam (N12.2)	Cephalexin (N9.2)	Folic Acid (N9.1)
	Psychiatry medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Lorazepam (N2.9)
	PHD	Acetaminophen (N6.6)	Anthromycin (N6.3)	Cefixime (N8.1)	Theophylline G (N4.6)	Loratadine (N4.6)
	Pharmacology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Fluconazole (N6.2)	Vitamin D3 (Calcitriol) (N6.2)	Aluminum MgS (N4.2)
	Pediatrics medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Loratadine (N2.9)
Cluster 9	Urology	Ciprofloxacin (N4.5)	Oflazoxin (N11.1)	Cefixime (N12.9)	Hydrochlorothiazide (N8.2)	Fosinotide (N4.3)
	Urology surgery	Cefixime (N11.8)	ASA (N10.4)	Metronidazole (N8.9)	Cephalexin (N8.5)	Acetaminophen (N8.2)
	Rheumatology	Folic Acid (N28.4)	Acetaminophen (N6.2)	Meropenem (N5.4)	Lorazepam (N5.3)	ASA (N5.1)
	Rehabilitation medicine	Acetaminophen (N4.9)	Amoxicillin (N9.6)	ASA (N9.1)	Cefixime (N4.7)	Metoprolol (N4.7)
	Radiotherapy	Cefixime (N18.8)	Beclim G Procaine (N7.3)	Penicillin G 3.3 (N6.6)	Doxycycline (N6.1)	Co-Amoxiclav (N5.5)
	Pulmonology	Anthromycin (N28.1)	Acetaminophen (N8.4)	Metronidazole (N7.9)	Nitroglycerin (N6.6)	Metoprolol (N6.2)
	Psychiatry medicine	ASA (N18.6)	Acetaminophen (N16.4)	Folic Acid (N12.9)	Amoxicillin (N9.7)	Metoprolol (N8.1)
	PHD	Amoxicillin (N3.9)	Chlorhexidine (N2.7)	Metronidazole (N2.1)	Co-Amoxiclav (N4.2)	Cefixime (N4.2)
	Pharmacology	Ciprofloxacin (N4.8)	Ondansetron (N3.12)	Amoxicillin (N11.5)	Cefixime (N11.2)	Co-Amoxiclav (N9.2)
	Pediatrics medicine	Metronidazole (N8.9)	Acetaminophen (N7.2)	Cefixime (N7.5)	Amoxicillin (N6.8)	ASA (N6.5)
Cluster 10	Urology	Ciprofloxacin (N4.5)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)
	Urology surgery	Diclofenac (N6.3)	Rufolone (N6.3)	Cefixime (N4.9)	Neproxen (N4.8)	Fosinotide (N4.7)
	Rheumatology	Prednisolone (N8.7)	Magnesium Glycinate (N6.4)	Magnesium Hydroxide (N6)	Metronidazole (N5.4)	Co-La (N5)
	Rehabilitation medicine	Metoprolol (N6.4)	ASA (N20.5)	Acetaminophen (N9.1)	Folic Acid (N8.1)	Nitroglycerin (N4.5)
	Radiotherapy	Sodium Chloride (NaCl) Inf (N28.1)	Magnesium Glycinate (N13.6)	Doxycycline (N13.2)	Hyoscine (N13.2)	Mefenamic Acid (N10)
	Pulmonology	Citalopram (N8.5)	Gabapentin (N8.5)	ASA (N2.8)	Folic Acid (N7.8)	Acetaminophen (N7.3)
	Psychiatry medicine	Cefixime (N4.8)	Amoxicillin (N17.4)	Metronidazole (N12.2)	Co-Amoxiclav (N10.9)	Acetaminophen Colicine (N2.2)
	PHD	Ondansetron (N10.9)	Acetaminophen (N8.5)	Sodium Chloride (NaCl) Topical (N7.9)	Subitamol (N6.1)	Ketotifen (N5.7)
	Pharmacology	Acetaminophen (N12.9)	ASA (N8.2)	Anthromycin (N7.1)	Hydrochlorothiazide (N7.1)	Amoxicillin (N7.2)
	Pediatrics medicine	Diclofenac (N2.6)	Vitamin D3 (Calcitriol) (N9.8)	Betamethasone (N7.6)	Calcium-D (N7.2)	Mefenamic (N7.2)
Cluster 11	Urology	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Urology surgery	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Rheumatology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Sodium Chloride (NaCl) Inf (N3.8)	Tranexamic Acid (N3)	Dexamethasone (N3)
	Rehabilitation medicine	Amoxicillin (N6.7)	Acetaminophen (N6.3)	Adult Gold (N8.3)	Bumex (N8.3)	Ciprofloxacin (N5.5)
	Radiotherapy	Levofloxacine (N6.7)	Acetaminophen (N6.5)	Gabapentin (N4.8)	Proparalolol (N4.8)	Diclofenac (N3.3)
	Pulmonology	Amoxicillin (N3.6)	Hyoscine (N12.8)	Clonazepam (N12.2)	Cephalexin (N9.2)	Folic Acid (N9.1)
	Psychiatry medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Lorazepam (N2.9)
	PHD	Acetaminophen (N6.6)	Anthromycin (N6.3)	Cefixime (N8.1)	Theophylline G (N4.6)	Loratadine (N4.6)
	Pharmacology	Folic Acid (N4.8)	Vitamin D3 (Calcitriol) (N3.9)	Fluconazole (N6.2)	Vitamin D3 (Calcitriol) (N6.2)	Aluminum MgS (N4.2)
	Pediatrics medicine	Amoxicillin (N4.3)	ASA (N4.1)	Vitamin D3 (Calcitriol) (N3.3)	Mefenamic (N2.9)	Loratadine (N2.9)
Cluster 12	Urology	Ciprofloxacin (N4.5)	Oflazoxin (N11.1)	Cefixime (N12.9)	Hydrochlorothiazide (N8.2)	Fosinotide (N4.3)
	Urology surgery	Cefixime (N11.8)	ASA (N10.4)	Metronidazole (N8.9)	Cephalexin (N8.5)	Acetaminophen (N8.2)
	Rheumatology	Folic Acid (N28.4)	Acetaminophen (N6.2)	Meropenem (N5.4)	Lorazepam (N5.3)	ASA (N5.1)
	Rehabilitation medicine	Acetaminophen (N4.9)	Amoxicillin (N9.6)	ASA (N9.1)	Cefixime (N4.7)	Metoprolol (N4.7)
	Radiotherapy	Cefixime (N18.8)	Beclim G Procaine (N7.3)	Penicillin G 3.3 (N6.6)	Doxycycline (N6.1)	Co-Amoxiclav (N5.5)
	Pulmonology	Anthromycin (N28.1)	Acetaminophen (N8.4)	Metronidazole (N7.9)	Nitroglycerin (N6.6)	Metoprolol (N6.2)
	Psychiatry medicine	ASA (N18.6)	Acetaminophen (N16.4)	Folic Acid (N12.9)	Amoxicillin (N9.7)	Metoprolol (N8.1)
	PHD	Amoxicillin (N3.9)	Chlorhexidine (N2.7)	Metronidazole (N2.1)	Co-Amoxiclav (N4.2)	Cefixime (N4.2)
	Pharmacology	Ciprofloxacin (N4.8)	Ondansetron (N3.12)	Amoxicillin (N11.5)	Cefixime (N11.2)	Co-Amoxiclav (N9.2)
	Pediatrics medicine	Metronidazole (N8.9)	Acetaminophen (N7.2)	Cefixime (N7.5)	Amoxicillin (N6.8)	ASA (N6.5)
Cluster 13	Urology	Ciprofloxacin (N4.5)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)	Cefixime (N4.1)
	Urology surgery	Diclofenac (N6.3)	Rufolone (N6.3)	Cefixime (N4.9)	Neproxen (N4.8)	Fosinotide (N4.7)
	Rheumatology	Prednisolone (N8.7)	Magnesium Glycinate (N6.4)	Magnesium Hydroxide (N6)	Metronidazole (N5.4)	Co-La (N5)
	Rehabilitation medicine	Metoprolol (N6.4)	ASA (N20.5)	Acetaminophen (N9.1)	Folic Acid (N8.1)	Nitroglycerin (N4.5)
	Radiotherapy	Sodium Chloride (NaCl) Inf (N28.1)	Magnesium Glycinate (N13.6)	Doxycycline (N13.2)	Hyoscine (N13.2)	Mefenamic Acid (N10)
	Pulmonology	Citalopram (N8.5)	Gabapentin (N8.5)	ASA (N2.8)	Folic Acid (N7.8)	Acetaminophen (N7.3)
	Psychiatry medicine	Cefixime (N4.8)	Amoxicillin (N17.4)	Metronidazole (N12.2)	Co-Amoxiclav (N10.9)	Acetaminophen Colicine (N2.2)
	PHD	Ondansetron (N10.9)	Acetaminophen (N8.5)	Sodium Chloride (NaCl) Topical (N7.9)	Subitamol (N6.1)	Ketotifen (N5.7)
	Pharmacology	Acetaminophen (N12.9)	ASA (N8.2)	Anthromycin (N7.1)	Hydrochlorothiazide (N7.1)	Amoxicillin (N7.2)
	Pediatrics medicine	Diclofenac (N2.6)	Vitamin D3 (Calcitriol) (N9.8)	Betamethasone (N7.6)	Calcium-D (N7.2)	Mefenamic (N7.2)
Cluster 14	Urology	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Urology surgery	Diclofenac (N7.1)	Chloramphenicol (N2.8)	Betamethasone (N4.8)	Artificial Tears (N8.8)	Erythromycin (N4.1)
	Rheumatology	Folic Acid (N4				

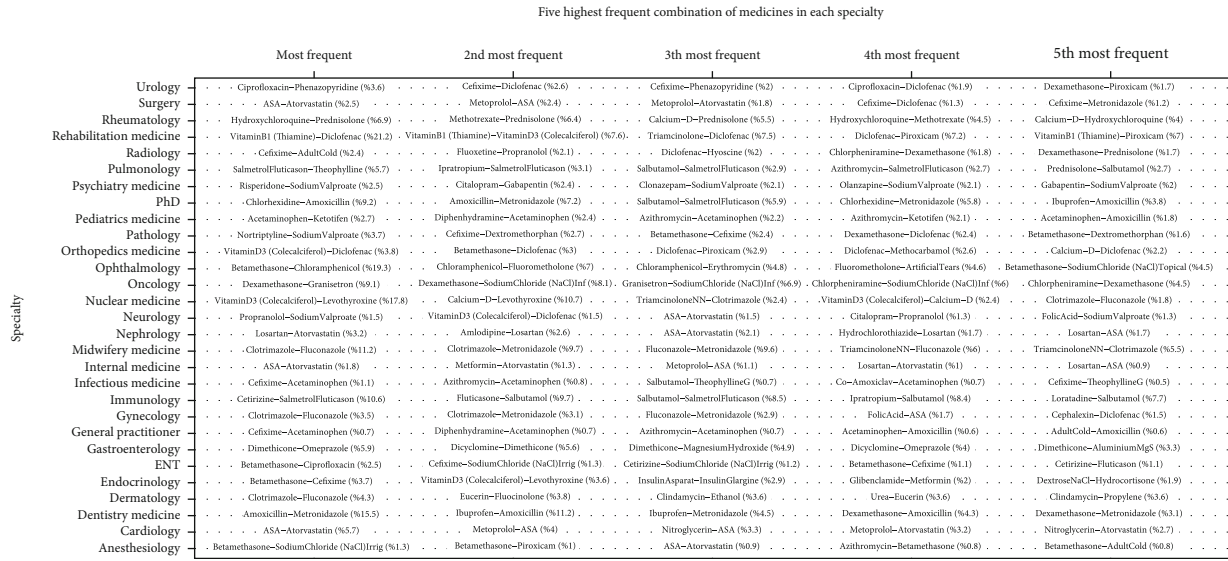


FIGURE 3: The five highest frequent 2-combs in each specialty.

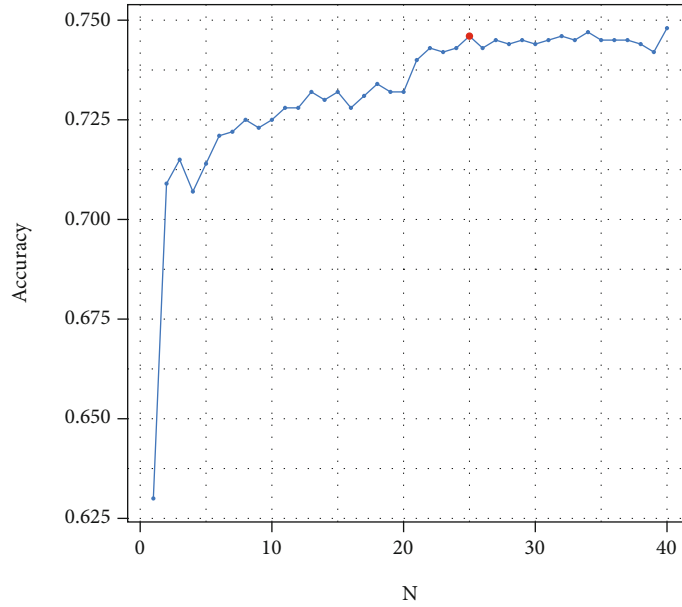


FIGURE 4: The accuracy of the RF models against the number of most frequent 2-combs used as predictors.

including more than two classes [23]. When specialties are missed in a database, there is no information available about specialist characteristics. So, the study has proposed an RF model that uses characteristic-free variables as predictors, i.e., 2-combs.

At last, in order to predict the specialty based on prescription patterns, the following actions have been done.

The RF method has been fitted as the following phases:

- (i) The count of the 2-combs was calculated and sorted in a descending order
- (ii) For each specialty, the selected proportions of the prescribed 2-combs were calculated

- (iii) Data were restructured as wide format
- (iv) To select the appropriate variables as predictors in the RF model, several RF models were fitted on the data using the first n most frequent 2-combs ($n = 1, 2, \dots, 40$)
- (v) Selecting variables to achieve appropriate accuracy
- (vi) Checking the validation of the model using the accuracy index

In this approach, the first 25 most frequent combinations have been used as variables, and the physician's specialty was considered the response. The 7354 specialties

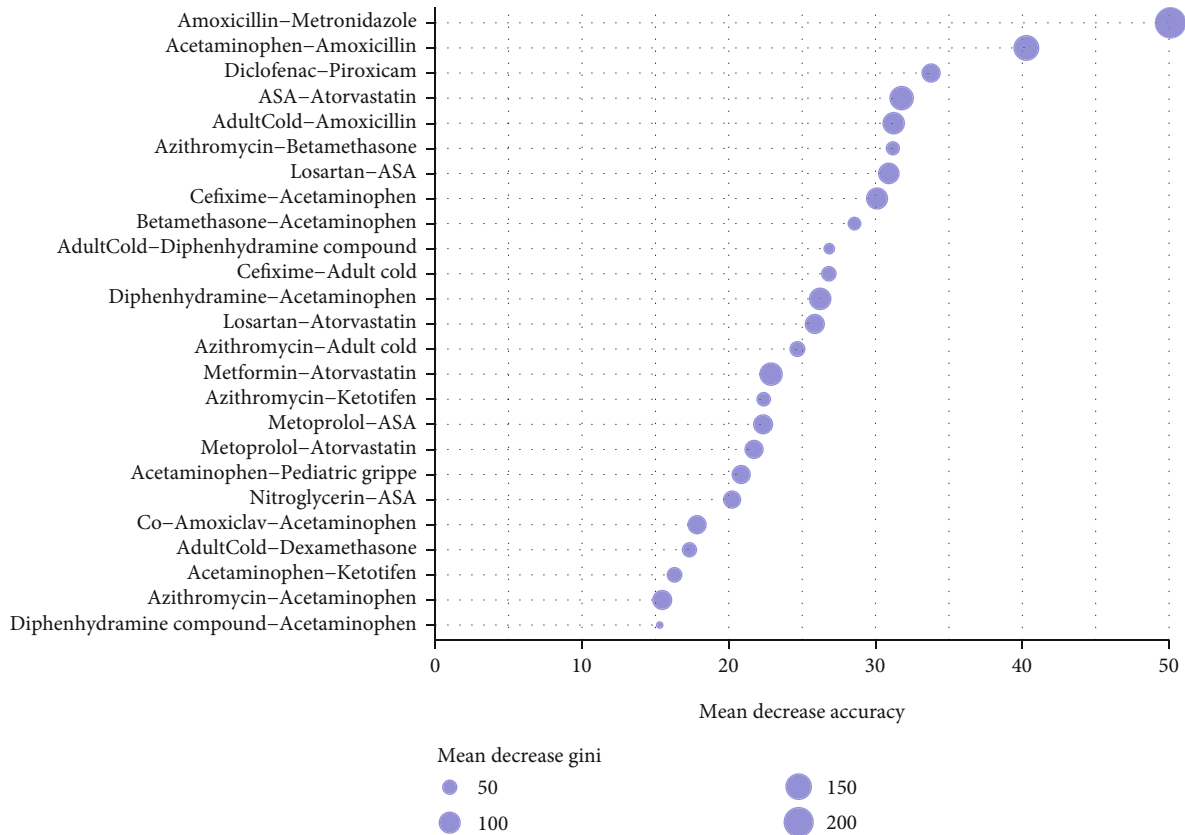


FIGURE 5: The mean decreased accuracy by deleting each predictor of the model.

have been split into training and validation datasets by a 70/30 proportion. Therefore, 5160 and 2194 observations were randomly allocated to training and test subsets, respectively.

To carry out the research aim, R 4.1.2 programming language has been used to develop codes, and the RStudio environment was applied to implement and execute them. As well, Apache Spark framework 3.1.0 via “sparklyr” R library along with “tidyverse” and “vroom” libraries was used for data wrangling. Besides, other libraries like “kamila” and “RandomForest” were used for analyses. Finally, “ggplot2” and “ggtext” have been applied to make visualization. A configured computer with 64 GB of RAM and a 32-core Intel CPU was used.

3. Results

The study results were obtained from 17,137,949 prescriptions prescribed by 30,544 physicians. 18,620 (61%) were related to male practitioners. The means of physician’s age were 51.9 and 42.1 for males and females, respectively. In addition, general practitioners (GPs) constituted 20,081 (65.7%) of total physicians.

In the first step, KAMILA clustering was performed on prescription records. Figure 1 provides overall sight of the distribution of each specialty in four clusters. For example, endocrinology and nephrology were concentrated in cluster 1.

As shown in Figure 2, the high-prescribed antibiotics (i.e., ciprofloxacin, azithromycin, cephalexin, chlorampheni-

col, and doxycycline) were concentrated in cluster 2. Results represented that GPs have prescribed the majority of the antibiotics. In addition, ibuprofen and diclofenac, two well-known nonsteroidal anti-inflammatory drugs (NSAIDs), were primarily categorized in clusters 1 and 4, respectively. Furthermore, dexamethasone was mainly categorized in clusters 3 and 4.

Variations and overlaps of medicines across specialty levels became a motivation factor in evaluating the 2-combs in another way. So, Figure 3 represents the five highest frequent 2-combs in each specialty.

The results of the RF classification model are presented as follows.

According to Figure 4, the value of accuracy in the 25th 2-combs tended to be a constant. The final RF model was fitted using the 25 most frequent 2-combs, which has the optimal accuracy of 0.74 in the validation dataset.

Figure 5 presents the mean decreased accuracy by deleting each predictor of the model, and it also represents each predictor’s importance. So, the combination of amoxicillin and metronidazole between the 25 first combinations contributed the most to accuracy. Also, the combination of ASA and atorvastatin in Figure 3 appeared in several specialties as the most frequent 2-combs have the fourth accuracy rank in Figure 5. To predict the specialty, the findings of the study can be applied to any similar medical prescription database, which is provided in a user-friendly R-shiny web application (<https://mahboube-akhlghi.shinyapps.io/SpecialtyPrediction>).

4. Discussion

The present study focused on predicting specialty based on the RF model on 17,265,238 prescriptions. Also, the method of this study is in line with numerous previous studies which have focused on disease prediction [19, 24]. Many studies presented the RF model's superiority over some machine learning methods [19].

Shirazi et al. used a community detection algorithm to identify the real specialty of physicians in prescription databases [25], which was a subjective method. In the present study, the authors tried to predict missing specialties differently. In the study, the 2-combs were utilized to predict the physician's specialty using the RF model.

However, the main findings of the study showed that the RF method could precisely predict the physician's specialty using the proportion of their most frequent 2-comb prescriptions. In addition, knowledge about the patterns of the prescription data can help better treatment and investigate the side effects of 2-combs. In this case, the combination of amoxicillin-metronidazole has the 21st rank among the 25 most frequently selected variables, which has the highest importance in making an accurate prediction. This is despite the fact that the recent systematic review and meta-analysis research about simultaneously amoxicillin/metronidazole usage in periodontitis treatment has shown positive short-term effects [26]. In addition, considering the extracted results showed that acetaminophen existed in several 2-combs. Several studies presented evidence that acetaminophen has adverse effects on certain human organs (for example, the liver) [27–29]. Moreover, some recent studies have presented simultaneously prescribing ASA and atorvastatin. He et al.'s research demonstrated the strong effect of ASA and atorvastatin 2-combs on inhibiting the growth of prostate cancer cells [30]. Other studies provided evidence of the merit of using these 2-combs in preventing and treating cardiovascular disease and severe sepsis [31–34]. Also, in many 2-combs, one of the medications was antibiotics. However, adverse events and antibiotic resistance are a well-acknowledged global problem [35, 36].

Finally, we found that general practitioners had the most outpatients. One reason for it is family physicians' issues [37]. Also, the services of general practitioners have a less economic burden on the patients than specialty physicians' services. Further, the prescribed medications were more related to the diseases such as colds and allergies, which patients often refer to a general practitioner. The observations in this study are limited to the insurance claimed database registry. However, in Iran, many prescriptions do not cover by insurance companies, and this issue is a specific limitation of the study.

5. Conclusion

In the study, we predict the physician's specialty by using the proportion of their most frequent 2-comb prescriptions. Today, a huge amount of data is produced in the field of medical prescriptions, which a significant section of that is missing in the specialty [12–17]. So, imputing the missing

can lead to valuable results for planning a medication with higher quality, improving healthcare quality, and decreasing expenses [38–40].

Abbreviations

NCDs:	Noncommunicable diseases
UNODC:	The United Nations Office on Drugs and Crime
KAMILA:	KAY-means for MIXed LARge datasets clustering
RF:	Random forest
SVM:	Support vector machine
GPs:	General practitioners
NSAIDs:	Nonsteroidal anti-inflammatory drugs.

Data Availability

The raw data used to support the findings of this study are restricted in order to protect patient privacy, but the aggregated data are available from the first author upon request.

Disclosure

This current study is a part of a Biostatistics PhD thesis at the School of Health, Isfahan University of Medical Sciences, with project number 397804.

Conflicts of Interest

The authors report no conflicts of interest in this work.

Acknowledgments

This study was funded partially by the Isfahan University of Medical Sciences (IUMS).

References

- [1] V. Yazdi Feyzabadi, M. Mehroolhassani, and M. Iranmanesh, "Evaluation of medication consumption indices in Iran from 2012 to 2015: a descriptive study," *Iranian Journal of Epidemiology*, vol. 14, pp. 72–81, 2019.
- [2] Collaborator WDR, *Executive Summary Policy Implications*, The United Nations Office on Drugs and Crime (UNODC), 2021.
- [3] F. Cappa, R. Oriani, E. Peruffo, and I. McCarthy, "Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance," *Journal of Product Innovation Management*, vol. 38, no. 1, pp. 49–67, 2021.
- [4] J. Björkdahl and M. Holmén, "Exploiting the control revolution by means of digitalization: value creation, value capture, and downstream movements," *Industrial and Corporate Change*, vol. 28, no. 3, pp. 423–436, 2019.
- [5] B. Zerhari, A. A. Lahcen, and S. Mouline, "Big data clustering: algorithms and challenges," in *International Conference on Big Data, Cloud and Applications BDCA'15*, pp. 1–6, Tetuan, Morocco, 2015.
- [6] Y. Li, X. Deng, S. Ba et al., "Cluster-based data filtering for manufacturing big data systems," *Journal of Quality Technology*, vol. 54, no. 3, pp. 1–13, 2021.

- [7] R. Chitta, *Kernel-Based Clustering of Big Data*, Michigan State University, 2015.
- [8] G. Jeon, A. K. Sangaiah, Y.-S. Chen, and A. Paul, "Special issue on machine learning approaches and challenges of missing data in the era of big data," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2589–2591, 2019.
- [9] E. Barbieri, C. Liberati, A. Cantarutti et al., "Antibiotic prescription patterns in the paediatric primary care setting before and after the COVID-19 pandemic in Italy: an analysis using the AWaRe metrics," *Antibiotics*, vol. 11, no. 4, p. 457, 2022.
- [10] Y. T. Wondmkun and A. G. Ayele, "Assessment of prescription pattern of systemic steroidal drugs in the outpatient department of Menelik II Referral Hospital, Addis Ababa, Ethiopia, 2019," *Patient Preference and Adherence*, vol. 15, pp. 9–14, 2021.
- [11] K. Kurata, E. Taniai, K. Nishimura, K. Fujita, and A. Dobashi, "A prescription survey about combined use of acetylcholinesterase inhibitors and anticholinergic medicines in the dementia outpatient using electronic medication history data from community pharmacies," *Integrated Pharmacy Research & Practice*, vol. 4, p. 133, 2015.
- [12] A. Patel, R. Medhekar, M. Ochoa-Perez et al., "Care provision and prescribing practices of physicians treating children and adolescents with ADHD," *Psychiatric Services*, vol. 68, no. 7, pp. 681–688, 2017.
- [13] C. Ringwalt, H. Gugelmann, M. Garretson et al., "Differential prescribing of opioid analgesics according to physician specialty for Medicaid patients with chronic noncancer pain diagnoses," *Pain Research and Management*, vol. 19, no. 4, Article ID 857952, 2014.
- [14] S. G. Weiner, O. Baker, A. F. Rodgers et al., "Opioid prescriptions by specialty in Ohio, 2010–2014," *Pain Medicine*, vol. 19, no. 5, pp. 978–989, 2018.
- [15] S. G. Weiner, S. C. Chou, C. Y. Chang et al., "Prescription and prescriber specialty characteristics of initial opioid prescriptions associated with chronic use," *Pain Medicine*, vol. 21, no. 12, pp. 3669–3678, 2020.
- [16] C. Ringwalt, A. W. Roberts, H. Gugelmann, and A. C. Skinner, "Racial disparities across provider specialties in opioid prescriptions dispensed to Medicaid beneficiaries with chronic noncancer pain," *Pain Medicine*, vol. 16, no. 4, pp. 633–640, 2015.
- [17] B. C. Sun, N. Lupulescu-Mann, C. J. Charlesworth et al., "Variations in prescription drug monitoring program use by prescriber specialty," *Journal of Substance Abuse Treatment*, vol. 94, pp. 35–40, 2018.
- [18] A. Foss, M. Markatou, B. Ray, and A. Heching, "A semiparametric method for clustering mixed data," *Machine Learning*, vol. 105, no. 3, pp. 419–458, 2016.
- [19] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–16, 2019.
- [20] A. D. Shah, J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway, "Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study," *American Journal of Epidemiology*, vol. 179, no. 6, pp. 764–774, 2014.
- [21] M. Rakhra, P. Soniya, D. Tanwar et al., "Crop price prediction using random forest and decision tree regression:-a review," *Materials Today: Proceedings*, 2021.
- [22] L. Yang, H. Wu, X. Jin et al., "Study of cardiovascular disease prediction model based on random forest in eastern China," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.
- [23] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 1, pp. 1–13, 2006.
- [24] Y. Jiang, X. Zhang, R. Ma et al., "Cardiovascular disease prediction by machine learning algorithms based on cytokines in Kazakhs of China," *Clinical Epidemiology*, vol. 13, pp. 417–428, 2021.
- [25] S. Shirazi, A. Albadvi, E. Akhondzadeh, F. Farzadfar, and B. Teimourpour, "A new application of community detection for identifying the real specialty of physicians," *International Journal of Medical Informatics*, vol. 140, p. 104161, 2020.
- [26] M. Karrabi and Z. Baghani, "Amoxicillin/metronidazole dose impact as an adjunctive therapy for stage II-III grade C periodontitis (aggressive periodontitis) at 3-and 6-month follow-ups: a systematic review and meta-analysis," *Journal of Oral & Maxillofacial Research*, vol. 13, no. 1, 2022.
- [27] Y. Tsuji, M. Kuramochi, H. M. Golbar, T. Izawa, M. Kuwamura, and J. Yamate, "Acetaminophen-induced rat hepatotoxicity based on M1/M2-macrophage polarization, in possible relation to damage-associated molecular patterns and autophagy," *International Journal of Molecular Sciences*, vol. 21, no. 23, p. 8998, 2020.
- [28] D. Kim, B. S. Moon, S. M. Park et al., "Feasibility of TSPO-specific positron emission tomography radiotracer for evaluating paracetamol-induced liver injury," *Diagnostics*, vol. 11, no. 9, p. 1661, 2021.
- [29] C. M. Herndon and D. M. Dankenbring, "Patient perception and knowledge of acetaminophen in a large family medicine service," *Journal of Pain & Palliative Care Pharmacotherapy*, vol. 28, no. 2, pp. 109–116, 2014.
- [30] Y. He, H. Huang, C. Farischoon et al., "Combined effects of atorvastatin and aspirin on growth and apoptosis in human prostate cancer cells," *Oncology Reports*, vol. 37, no. 2, pp. 953–960, 2017.
- [31] D. Mahtta, D. J. Ramsey, M. Al Rifai et al., "Evaluation of aspirin and statin therapy use and adherence in patients with premature atherosclerotic cardiovascular disease," *JAMA Network Open*, vol. 3, no. 8, article e2011051, 2020.
- [32] C. H. Hennekens and W. R. Schneider, "The need for wider and appropriate utilization of aspirin and statins in the treatment and prevention of cardiovascular disease," *Expert Review of Cardiovascular Therapy*, vol. 6, no. 1, pp. 95–107, 2008.
- [33] M. A. Sanchez, C. B. Thomas, and H. R. O'Neal, "Do aspirin and statins prevent severe sepsis?," *Current Opinion in Infectious Diseases*, vol. 25, no. 3, pp. 345–350, 2012.
- [34] S. S. Patel, L. A. Guzman, F. P. Lin et al., "Utilization of aspirin and statin in management of coronary artery disease in patients with cirrhosis undergoing liver transplant evaluation," *Liver Transplantation*, vol. 24, no. 7, pp. 872–880, 2018.
- [35] B. Lund, A. Cederlund, M. Hultin, and F. Lundgren, "Effect of governmental strategies on antibiotic prescription in dentistry," *Acta Odontologica Scandinavica*, vol. 78, no. 7, pp. 529–534, 2020.
- [36] L. M. King, M. Bartoces, K. E. Fleming-Dutra, R. M. Roberts, and L. A. Hicks, "Changes in US outpatient antibiotic prescriptions from 2011–2016," *Clinical Infectious Diseases*, vol. 70, no. 3, pp. 370–377, 2020.

- [37] R. E. Rakel and R. E. Rakel, *Textbook of Family Practice*, Wb Saunders, 1984.
- [38] P. S. Duggal and S. Paul, "Big data analysis: challenges and solutions," *International Conference on Cloud, Big Data and Trust*, vol. 15, pp. 269–276, 2013.
- [39] P. S. Mathew and A. S. Pillai, "Big data solutions in healthcare: problems and perspectives," in *2015 International conference on innovations in information, embedded and communication systems (ICIIECS)*, pp. 1–6, Coimbatore, India, 2015.
- [40] A. Heinrich, A. Lojo, and F. Xu, "Big data technologies in healthcare: needs, opportunities and challenges," *White Paper, Big Data Value Association, TF7 Healthcare Subgroup*, vol. 21, 2016.