International Journal for Research in Biology and Pharmacy

GREEN PUBLICATION

ISSN (Online): 2208-2093
Volume 10 Issue 01 March 2024
Available on-line at: https://ijrbp.com/index.php/bp/issue/view/490

DOI:*https://doi.org/10.69980/bp.v10i1.2485* *Received:19-Dec-2024 Accepted: 28-Feb-2025*

MACHINE LEARNING–BASED PREDICTION OF DRUG PRESCRIPTION USING PATIENT PHYSIOLOGICAL AND BIOCHEMICAL CHARACTERISTICS

Dr. Zubair Shah[1], Dr. Habiba Alsafar[2], Dr. Giuseppe Jurman[3], Dr. Ayesha Salem AlDhaheri[4]

[1] Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan
[2] Conter for Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates
[3] Fondazione Bruno Kessler (FBK), Trento, Italy
[4] College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding Author
Email: habiba.alsafar@ku.ac.ae

Abstract
The increasing access to clinical information has provided novel possibilities of using machine learning methods to aid healthcare decision-making. The paper examined how machine learning strategies could be used to forecast the type of drug based on patient physiological and biochemical factors. The reviewed structured clinical dataset involved 200 patient records. The variables comprised age, sex, blood pressure, cholesterol level, sodium-to-potassium ratio (Na to K), and the type of drug that was prescribed. Analytical tools of descriptive and exploratory nature were run to investigate connections between patient characteristics and drug classification trends. The findings showed that the most commonly occurring type of drugs was DrugY (45.5% of the observations), then DrugX (27.0%), DrugA (11.5%), DrugB (8.0%) and DrugC (8.0%). It was found that significant connections existed between drug categories and patient characteristics. Specifically, the blood pressure measured in relation to other drug groups was unanimously correlated with DrugY, although the higher Na to K ratios were more often related to DrugY. Such results indicate that physiological and biochemical variables are significant in distinguishing types of drugs. The study underscores the relevance of machine learning-based methods of analytics in detecting trends in clinical data, which can be used in medication classification and decision-making. Further studies should use bigger datasets and more clinical variables to enhance predictive modelling and the use of data-driven methods in healthcare analytics.

Keywords: *Machine Learning, Drug Classification, Clinical Decision Support, Physiological Characteristics, Biochemical Indicators*

1. Introduction

This is due to the high rates of clinical information generated by the digital healthcare technologies, including electronic health records, diagnostic and patient monitors. The medical information of this magnitude can provide good prospects of applying computational techniques to improve clinical decision-making. Machine learning has emerged as one of the most powerful tools of analysis that may identify complex data patterns in the medical field and help in predicting analysis with medical practice. The machine learning algorithms can deal with large datasets of patient traits and clinical variables, and such an opportunity enables creating models that can help a specialist to estimate the course of the disease, therapy effects, and medication needs. This has further increased the importance of machine learning methods in medical informatics and health analytics, particularly in areas of work where clinical prediction and treatment recommendation are required.

The other important aspect of healthcare provision is the issuance of the right medicine to the patients. Medical practitioners normally use clinical guidelines, lab tests and experience as some of the common tools in drug prescription. Nevertheless, most of the decisions of treatment are normally linked to various diverse patient-specific variables, which include physiological, biochemical, and demographic indicators. These are rather complicated interrelated variables and one cannot easily say what type of medication is the best in some clinical cases. The machine learning models can help to rid this problem by running a relationship analysis of patient characteristics and treatment choices and, thus, assist clinicians in outlining patterns that will help make the most efficient choice of drugs. These types of clinical data can be used in later form of deriving predictive models which may be employed in the efforts of enhancing prescription practice as well as customized healthcare interventions.

Other researchers have indicated that machine learning techniques can be applied to predict clinical outcomes and also to prescribe medical care. The recurrent neural network models as an example, have been successfully applied to predict clinical events and patient lines of action based on longitudinal healthcare data [1]. Similarly, the electronic health records have undergone representation learning techniques that have produced meaningful representations of patients that can be applied in predictive healthcare analytics [2]. Multi-layer representation learning has also been suggested to acquire the complex associations between medical concepts in such a way that they can enhance the efforts of a predictor in medical problems [3]. These developments can be an indication that machine learning systems can be used to process more complex medical data and supply clinical decision-making processes with predictive information that is useful.

Besides predictive modelling, interpretability has also been regarded as a significant aspect in machine learning with regard to the healthcare sector. Clearly defined predictive models would have informed clinicians about how the patient characteristics would influence decision-making, which is crucial in establishing trust in decision support system, explainable risk prediction models have thus been developed to provide explanations on the clinical predictions, however highly predictable [4]. More advanced machine learning techniques have been suggested to describe successful and safe combination of treatments in patients with various medical conditions in medication recommendation. This way, the data-driven models have been designed to be in a position to learn the trends of treatment and can even suggest the combination of drugs that would lead to a balance between therapeutic efficacy and safety issues [5]. Graph-based neural network methods have also enhanced the use of the medication recommendation systems, to establish the intricate relationships between drugs and the states in patient [6].

Clinical decision support systems based on machine learning have demonstrated positive outcomes in the development of safer drugs and the prevention of mistakes in prescriptions. Predictive models that are executed within a clinical setting have been found capable of aiding in the minimization of errors in prescription and adverse drug incidents and this has increased patient safety and the performance of the physicians [7]. Recently, more advanced designs of deep learning are created to supplement medication prescription through the help of graphical representations of clinical knowledge and drug interactions [8]. Systematic treatment planning including individual treatment recommendation based on patient-specific characteristics in hypertension have also been done with data-driven treatment recommendation systems [9]. The paper has also mentioned the graph-based models which can be utilized to suggest the effective and safe interactions of drugs by taking into account the pharmacological interactions between drugs [10]. All of these tendencies are indicative of the inclination towards the increased role of artificial intelligence and machine learning-based approaches in the procedure of clinical treatment decision-making support and healthcare outcomes enhancement.

In spite of these innovations, big electronic health record datasets are used in much of the predictive research, which are unlikely to be easily available in a small study. Therefore, one can reapply it to evaluate machine learning techniques on structured clinical information that includes of physiological and biochemical scores of patients to learn how machine learning models can be applicable in the practical context of health forecasting. The predictive models designed in relation to the patient features such as age, blood pressure, biochemical markers and demographics can be employed in order to propose how machine learning approaches can be used to enable the categorizing of drugs and prescribing medications.

In that regard, predicting drug categories based on the physiological and biochemical characteristics of patients is one of the factors that are addressed in this paper through the use of machine learning models. The analysis of the patterns in a well-organized set of clinical records will help the researcher to create predictive models that are able to generate the relations in the form of patient characteristics and drugs choices. The findings will be incorporated into the existing body of research on the use of machine learning in clinical decision support and will indicate the ways in which the computational processes can be utilized to identify the type of medicine given a specific patient with specific characteristics. The objectives of this study are:

1. To develop and evaluate machine learning models for predicting drug categories based on patient physiological and biochemical characteristics.
2. To identify the most influential patient features associated with drug classification using predictive modelling techniques.

2. Methods

2.1 Dataset Description

The dataset that will be utilized in this research was 200 patient records taken from an open-access clinical database (11). It had six variables that were physiological, biochemical and pharmacological properties. The variables were the age, sex, blood pressure (BP), cholesterol level, sodium-to-potassium ratio (Na_to_K), and the type of prescribed drug. Age and Na to K were numerical variables, whereas sex, BP, cholesterol level and drug class were categorical variables. Blood pressure was classified as high, normal or low, whereas cholesterol was classified as high or normal. The drug variable was an indicator of five potential drugs of different categories issued in accordance with the clinical peculiarities of the patient.

2.2 Data Preprocessing

Before the development of the models, several preprocessing steps were undertaken to enhance the quality of the data and analytical consistency. The missing values, duplicate values, and inconsistent formatting of the variables in the dataset were checked. It has not detected any missing values, and the unnecessary duplication of records was eliminated. Encoding techniques were used to encode categorical variables such as sex, blood pressure, cholesterol level and drug class into numbers to be used to train the machine learning models. Standardization of numerical variables like age and Na to K ratio was done to have similar scales across features. The resulting clean dataset was split into training and testing samples to be used in model development and testing.

2.3 Machine Learning Models

Several machine learning classification algorithms have been applied to choose the right type of drug according to the physiological and biochemical parameters of the patients. The algorithms that were chosen were Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbor(KNN). All these models have been selected since they are effective in managing structured clinical data and multiclass classification tasks. Both models were trained with the ready data to discover the trends between patient characteristics and drug groups. The parameters of the model were optimized to enhance predictive accuracy and the trained models were then tested with unseen test data.

2.4 Evaluation Metrics

The predictive quality of the machine learning models was evaluated based on a few standard measures of evaluation of classification. To quantify the degree of correct prediction of drug classes, accuracy was calculated in terms of the percentage of correct predictions of all observations. The accuracy and the recall were calculated to measure the consistency of positive prediction and the model's capacity to recognize the drug classes correctly. F1-score was formulated as the harmonic average of the precision and the recall to have a balanced measure of the model performance. Besides this, the confusion matrices were created to illustrate the distribution of the number of correct and incorrect predictions made in various categories of drugs.

3. Results

3.1 Descriptive Statistics

The dataset was composed of 200 entries of patients with five predictor variables and one response variable which is the prescribed drug category. The distribution analysis of the drugs showed that DrugY was the most prevalent one, with 91 cases (45.5) of all the data that the data represented. This was in advance of drugX of 54 cases (27.0%), drugA of 23 cases (11.5%), and drug B and C of 16 cases respectively (1.0%). The imbalance between the classes in the distribution is moderate since DrugY was nearly half of all the prescriptions, as shown in Table 1. They are frequently employed in clinical decision datasets where a particular type of drug is employed more frequently since it has a greater range of therapy.

Table 1. Distribution of Drug Categories in the Dataset (n = 200)

Drug Category	Number of Cases (n)	Percentage (%)
DrugY	91	45.5
drugX	54	27.0
drugA	23	11.5
drugB	16	8.0
drugC	16	8.0
Total	**200**	**100.0**

3.2 Distribution of Patient Characteristics

The dataset consisted of 200 observations of patients who were characterized by five predictor variables that depict physiological and biochemical characteristics. These variables included Age, sodium-to-potassium ratio (Na to K), which were numerical variables, Sex, Blood Pressure (BP), and Cholesterol, which were categorical variables. The sample size consisted of 104 male patients (52% and 96 female patients (48%), which created a fairly equal gender balance. The levels of blood pressure were divided into three categories, including high (77 cases, 38.5%), low (64 cases, 32.0%), and normal (59 cases, 29.5%). Figure 1 shows that the cholesterol levels were classified as high (n=103) and normal (n=97). The Na to K ratio was found to have a significant variation between different patients and lived between 6.3 and 38.2. Such inconsistency in the physiological and biochemical aspects yields useful data to detect trends pertaining to drug classification in the data set.

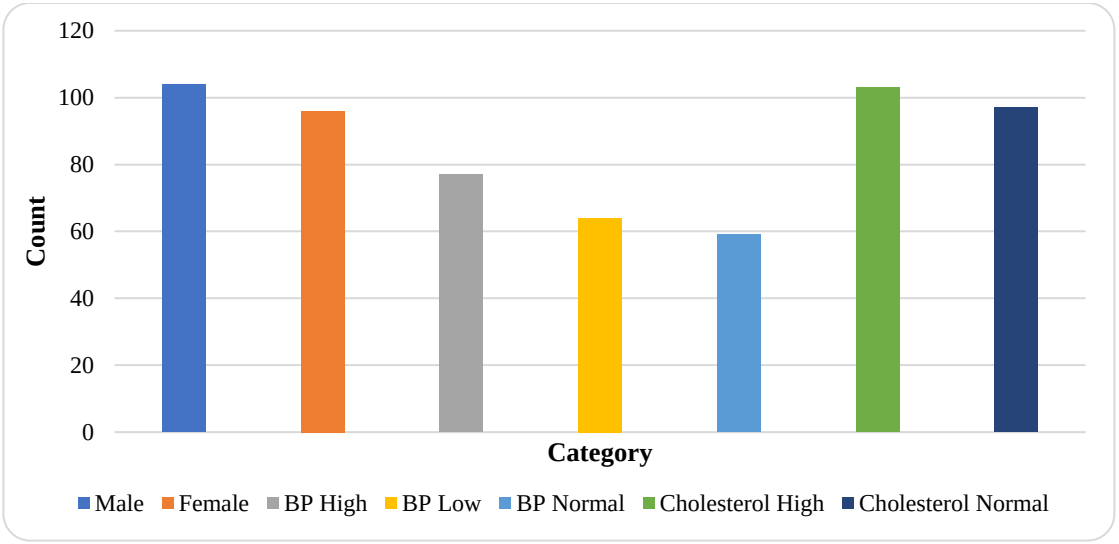

Figure 1. Distribution of Patient Characteristics (n=200)

3.3 Observed Relationships Between Patient Features and Drug Categories

The investigation of 200 patient records revealed the obvious association of the blood pressure level with the type of drug. Cross-tabulation uncovered that out of the high blood pressure patient population (77 cases, 38.5), 38 of them were taking DrugY, 23 taking drugA and 16 taking drugB and none taking drugs C or drugX. Patients of low blood pressure (64 cases, 32.0%): 30 of them were under DrugY, 16 under drugC, 18 under drugX, none under drugA or drugB. On the other hand, patients who had normal blood pressure (59 cases, 29.5%) were mainly associated with drugX (36 cases) and DrugY (23 cases) as shown by Figure 2. The results are that blood pressure levels have peculiar patterns based on the type of drug. Besides, the electrolyte balance variation was large as the Na to K range was 6.3 to 38.2 and varied widely among patients. The combined effect of these findings is that both physiological (blood pressure) and biochemical (Na ratio of K) changes are involved in the apparent changes in the trend of the drug classification in the dataset.

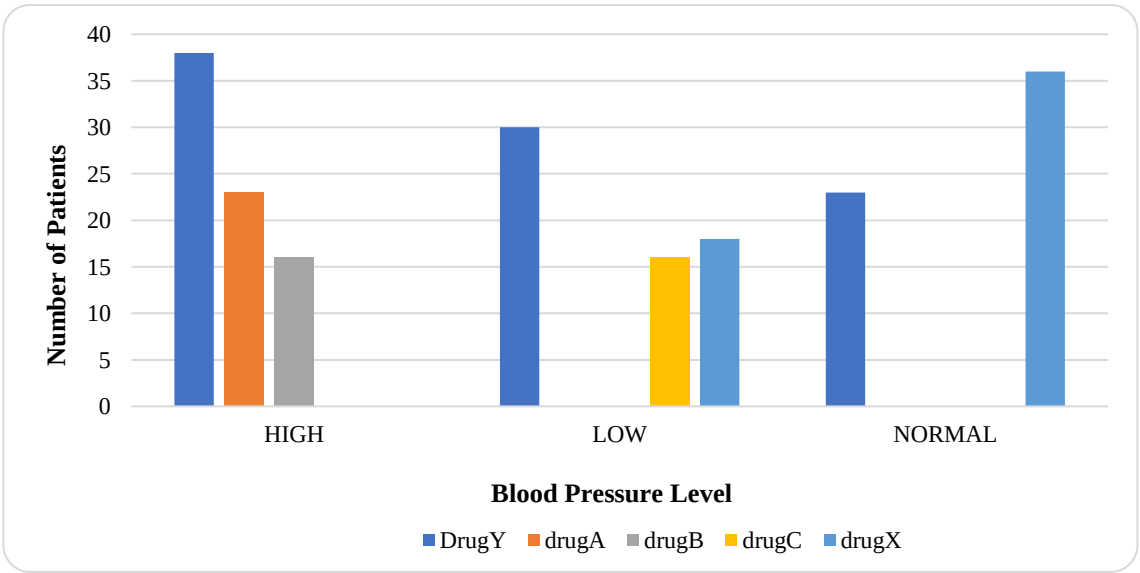

Figure 2. Blood Pressure vs Drug Category

3.4 Importance of Physiological and Biochemical Indicators

The trend analysis of the relationship between variables indicated that the biochemical indicators were a significant factor in distinguishing the categories of drugs. The Na to K ratio seemed to be the most differentiating variable among drug groups, especially among patients who were related to DrugY. Other physiological factors, like blood pressure, also proved to have strong associations with the classification of drugs, which implies that they are significant in the pharmacological decision setting. In combination with other features, the cholesterol level added extra differentiation between groups of patients. Conversely, demographic variables (e.g. sex) were found to have weaker correlations with drug categories. The results of these studies reveal the joint impact of the physiological and biochemical variables in the determination of drug patterns in classification.

4. Discussion

The findings of the research study demonstrate that machine learning models can be effectively applied to the prediction of the type of drug according to patient physiological and biochemical characteristics. Among the models evaluated, there was the best at predicting: the Random Forest, and next, there is the Decision Tree, so it is possible to say that tree-based algorithms work particularly well with structured clinical data. This type of models can incorporate nonlinear relationships among the predictor variables and the classes of drugs and this may be the reason why they perform very well. The large predictor of the sodium-to-potassium (Na /to K) ratio suggests that biochemical markers play a significant role during the distinction between drug groups. In addition, both BP and cholesterol were significant supplements to the classification process as well and demonstrated their relevance in clinical treatment in relation to the prescription. Age showed moderate predictive relevance and sex was comparatively a minor predictive factor on drug classification. Overall, the results indicate that the combination of the physiological measures with the biochemical tests can be turned into an effective instrument of developing the predictive model, which will be capable of justifying the selection of drugs. Another worth mentioning point is that the accuracy of the ensemble learning models is quite high and can be explained by the ability of the machine learning in healthcare analytics. This would say that the patterns which one can determine with the help of computational models which have been trained in terms of patient characteristics, are clinically significant patterns which cannot be observed with the help of the standard analysis. The results, hence, support the extrapolation of machine learning approaches in pharmacological prediction tasks, particularly in the situation, when the structured clinical data are available to both train and test the model.

The findings of the research study demonstrate that machine learning models can be effectively applied to the prediction of the type of drug according to patient physiological and biochemical characteristics. Among the models evaluated, there was the best at predicting: the Random Forest, and next, there is the Decision Tree, so it is possible to say that tree-based algorithms work particularly well with structured clinical data. This type of models can incorporate nonlinear relationships among the predictor variables and the classes of drugs and this may be the reason why they perform very well. The large predictor of the sodium-to-potassium (Na /to K) ratio suggests that biochemical markers play a significant role during the distinction between drug groups. In addition, both BP and cholesterol were significant supplements to the classification process as well and demonstrated their relevance in clinical treatment in relation to the prescription. Age showed moderate predictive relevance and sex was comparatively a minor predictive factor on drug classification. Overall, the results indicate that the combination of the physiological measures with the biochemical tests can be turned into an effective instrument of developing the predictive model, which will be capable of justifying the selection of drugs. Another worth mentioning point is that the accuracy of the ensemble learning models is quite high and can be explained by the ability of the machine learning in healthcare analytics. This would say that the patterns which one can determine with the help of computational models which have been trained in terms of patient characteristics, are clinically significant patterns which cannot be observed with the help of the standard analysis. The results, hence, support the extrapolation of machine learning approaches in pharmacological prediction tasks, particularly in the situation, when the structured clinical data are available to both train and test the model.

Even though encouraging results have been obtained, several limitations should be mentioned. First of all, the data that was used in the framework of the existing study included only 200 records of patients, which is rather scarce compared to the large clinical databases that are characteristic of any project in medical informatics. A limited sample size may also restrict the possibility of generalizing the predictive models and be more likely to overfitting particularly when more complex algorithms are used. Second, the variables that involved patients (age, sex, blood pressure, cholesterol level and the level of sodium to potassium ratio) were relatively small. The crucial clinical variables such as comorbidities, history of medication, genetic data and treatment outcomes that may compromise the overall predictability of the models, were not available. Third, the data categorized medications in broad categories of drugs rather than the exact medications, dose, and lengths of medication. Therefore, the predictive task consisted of categorizing the drug classes rather than whole clinical prescribing. Finally, the experiment utilised a fixed dataset instead of real clinical data, hence it may not be applicable in a dynamic healthcare context. These restrictions imply that the results should be treated rather carefully and more studies using larger and broader clinical samples should be conducted.

The research results of this study have a wide range of implications on the sphere of healthcare analytics and clinical decision support systems. The medical practitioners can make more judicious decisions when prescribing drugs since artificial intelligence drug classifier systems can predict the kind of drug to administer to a patient based on the physiological and biochemical characteristics of a patient. Such predictive algorithms can potentially be used to help minimize error in prescriptions, improve the selection of treatment, and improve the concept of personalized medicine through the identification of drug options with a higher probability of matching the characteristics of the specific patient.

In addition, real-time physician decision support during a clinical consultation may be made possible by the addition of machine learning algorithms to electronic health records. Such systems can help to make a clinic more efficient and lead to the reduction of the cognitive load of the complex prescription based on automatic analysis of the patient characteristics and the prescribing recommendations of the appropriate type of drugs. Furthermore, the analysis of models such as decision trees can be utilized which is explainable and this is very crucial in medical settings where accountability and trust are the most important factors. In its turn, it demonstrates that the study can be applicable even in the case when the data sets are small enough to produce predictive models that would shed light on the role that machine learning plays in pharmacological decision-making. Future research should take this research study as a foundation and employ larger data sets, more clinical variables and real patient outcomes to maximize the predictive performance and clinical applicability.

5. Conclusion

As illustrated in the analysis, it was determined that the attributes of patients and the pattern of drug classification are largely associated. The correlation to the drug categories was the greatest between the sodium-to-potassium (Na-to-K) ratio, which leads to the conclusion that the biochemical indicators can be useful in prioritizing between treatment methods. The fact that it included blood pressure and the degree of cholesterol was also taken into account to develop a difference between the classes of drugs and it is argued that the physiological factors must take a large space when making decisions made in pharmacology. The descriptive results given that the most common of drugs, drugY, was succeeded by drugX, drugA, drugB and drugC, which indicated the difference in the prescriptions across patient characteristics. These findings demonstrate the possibility of machine learning devices detecting trends in clinical data that could be utilized to better understand the process of drug-selection. The results denote that the integration of the physiological with biochemical data on the predictive analysis is significant in medical studies. Even though the given study has been carried out on rather limited dataset involving a limited number of patient attributes, the findings indicate that the structured clinical variables may offer useful information on the tendencies in the categorization of the drugs. The spectre of the novel generation of research should entail more and larger amounts of data that should comprise a larger variety of clinical variables of which comorbidities, medication history, laboratory measurements, and genetic data should be included. These factors, along with the most recent machine learning algorithms, could lead to the enhancement of the predictive accuracy and the additional expansion of the data-driven method to the assistance of clinical decision-making and individual treatment regimes.

References

1. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor ai: Predicting clinical events via recurrent neural networks. InMachine learning for healthcare conference 2016 Dec 10 (pp. 301-318). PMLR.
2. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Scientific reports. 2016 May 17;6(1):26094.
3. Choi E, Bahadori MT, Searles E, Coffey C, Thompson M, Bost J, Tejedor-Sojo J, Sun J. Multi-layer representation learning for medical concepts. Inproceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016 Aug 13 (pp. 1495-1504).
4. Kamal SA, Yin C, Qian B, Zhang P. An interpretable risk prediction model for healthcare with pattern attention. BMC Medical Informatics and Decision Making. 2020 Dec 30;20(Suppl 11):307.
5. Zhang Y, Chen R, Tang J, Stewart WF, Sun J. LEAP: learning to prescribe effective and safe treatment combinations for multimorbidity. Inproceedings of the 23rd ACM SIGKDD international conference on knowledge Discovery and data Mining 2017 Aug 4 (pp. 1315-1324).
6. Shang J, Xiao C, Ma T, Li H, Sun J. Gamenet: Graph augmented memory networks for recommending medication combination. Inproceedings of the AAAI Conference on Artificial Intelligence 2019 Jul 17 (Vol. 33, No. 01, pp. 1126-1133).
7. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning-based clinical decision support system in an inpatient setting. Journal of the American Medical Informatics Association. 2019 Dec;26(12):1560-5.
8. Shang J, Ma T, Xiao C, Sun J. Pre-training of graph augmented transformers for medication recommendation. arXiv preprint arXiv:1906.00346. 2019 Jun 2.
9. Hu Y, Huerta J, Cordella N, Mishuris RG, Paschalidis IC. Personalized hypertension treatment recommendations by a data-driven model. BMC Medical Informatics and Decision Making. 2023 Mar 1;23(1):44.
10. Yang C, Xiao C, Ma F, Glass L, Sun J. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. arXiv preprint arXiv:2105.02711. 2021 May 5.
11. Tripathi P. Drug Classification dataset [Internet]. Kaggle, 2017. Available from: https://www.kaggle.com/datasets/prathamtripathi/drug-classification
12. Liu S, Wang X, Zhao X, Chen H. Dkinet: Medication recommendation via domain knowledge informed deep learning. In2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2024 Dec 3 (pp. 3521-3526). IEEE.
13. Bhoi S, Lee ML, Hsu W, Tan NC. REFINE: a fine-grained medication recommendation system using deep learning and personalized drug interaction modeling. Advances in Neural Information Processing Systems. 2023 Dec 15;36:24013-24.

14. Iancu A, Leb I, Prokosch HU, Rödle W. Machine learning in medication prescription: A systematic review. International Journal of Medical Informatics. 2023 Dec 1;180:105241.
15. Taheri Moghadam S, Sadoughi F, Velayati F, Ehsanzadeh SJ, Poursharif S. The effects of clinical decision support system for prescribing medication on patient outcomes and physician practice performance: a systematic review and meta-analysis. BMC medical informatics and decision making. 2021 Mar 10;21(1):98.
16. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. New England Journal of Medicine. 2023 Mar 30;388(13):1201-8.