Research on Drug Classification Using Machine Learning Model

Changlin Chen *

Shanghai Adcote School, Shanghai, Wai Qing Song Road, China

* Corresponding Author Email: AndrewChen0923@outlook.com

Abstract. In the recent years, more and more fields around the development and application of artificial intelligence have received new improvements and breakthroughs. Among them, drug design and classification are one of the most popular application areas which can provide the greatest help to the society. According to statistics, the number of newly listed drugs in the world is decreasing year by year. At the same time, the risks and costs of drug development are increasing year by year which illustrates a different trend compared with the newly listed drugs. In this case, the application of artificial intelligence technology provides a new idea and opportunity to solve the problem of drug design and classification with higher accuracy than human work. This paper aims to do a research based on using machine learning models to produce suitable outcomes for patients of different drug types in order to reduce the working pressure of doctors in the hospitals.

Keywords: Machine Learning Model, Drugs, Artificial Intelligence.

1. Introduction

In the recent society, medication is the most important and closest fields in human life. With the longer of human longevity, the pressure and work loading of doctors become more and more critical. Nowadays, artificial intelligence is already used in many other fields which includes chess races, auto driving, graphs classification and repairing, but medicinal however, still not using artificial intelligence which can be the best way of using it, developing the qualities of daily life among all the other ways. To use artificial intelligence to release doctors' pressure and work loading is the hot topic and research direction. Along with the development of chemistry and modern technology, the research of medicine goes to the direction of chemical and biology. However, it takes a long cycle time for the company to discover new drugs and the company shall spend a lot for developing new drugs. This becomes the tremendous difficulties to look for new drugs. In the meantime, with the introducing of computer science by using AI, it brings more opportunities for new drugs' industry. G. Shobana and Dr. S. Nikkath Bushra trained the Machine Learning models with observations that define the ADMET properties, pharmacokinetics and physicochemical properties of drugs [1].

The article discusses how to use artificial intelligence to help on drug classification to outcome suitable drugs for the patients.

Artificial intelligence is used to receive and process the information provided by the doctor, which avoids possible safe drug use accidents and greatly speeds up the process of prescription processing [2].

In the practical application of drug research and development, some potential drug candidate compounds can be detected early through the treatment and analysis of different types of compounds. It is then validated and developed in clinical trials. Among them, the application of artificial intelligence technology provides strong support for the quality and efficiency of drug classification. G. Shobana and Dr. S. Nikkath Bushra found that the type of drugs affects the efficiency of absorbing the drugs for patients. It is generally classified as three major types. The rate of absorption of the drug varies depending on the routes [1].

Overall, artificial intelligence technology can comprehensively evaluate multiple indicators in the process of drug classification. Thus, the error caused by misjudgment of a single indicator is effectively avoided. For example, when testing the antiviral effect of a compound, it is possible to analyze the variation of the antiviral activity of different compounds through a large number of

experiments a simulation, and depending on the size of compound molecule, electrical properties and other characteristics to estimate the efficacy and side effects of riding in the human body.

2. Methods

This paper selects to use machine learning model to reduce manual efforts on drug classification There are three purposes to do this research. Abdel Rehman et al. applied Machine Learning techniques like Support Vector Machine, Random Forest and Decision Tree to predict the interaction of drugs with Adenosine Receptors. They also used SMOTE (Synthetic Minority Over Sampling Technique) technique to manage the unbalanced dataset, which had 3990 side effects, 400 drugs and 794 targets. Among the three classifiers Random Forest rank first with 75.09% accuracy [3]. With the original features of 21, Logistic Regression performed well with 96% and Random Forest performed very efficiently with 97% accuracy [4].

Another model which is called Pharmacokinetic simulation. Pharmacokinetic simulation is a three-dimensional model that simulates a drug in the human body. In the model, the effects of drugs on human tissues, organs and proteins can be simulated in the details. The machine learning algorithm combined with the pharmacokinetic simulation and artificial intelligence technology can more accurately predict the degree and duration of drug influence, and eliminate the more coercive drugs as soon as possible, improving the efficiency and success rate of drug development.

Artificial intelligence technology can be based on the patient's genotype disease and other factors through computer simulation real-time monitoring has been reasonable allocation of drugs to each patient. Each patient has a different genotype and drug efficacy, and artificial intelligence technology can be more longitude and efficient. The application of artificial intelligence technology to the clinic can not only promote the precise treatment of drugs, but also reduce the work pressure of doctors and improve the efficiency of medical research.

According to the market nowadays, artificial intelligence technology can be used in the process of drug classification through data mining and machine learning for a better identification of possible risks and root causes. For example, in some drug clinical trials, serious side effects may occur. This situation is usually related to the physical and chemical properties of the compound structure and efficacy. In the process of applying artificial intelligence technology, through long-term data analysis and comparison, researchers can be provided with useful references and suggestions to help them reduce risks and improve the success rate of drug development.

The characteristics and trends of technological development in medical artificial intelligence industry and infrastructure at home and abroad are summarized [5].

In the process of drug development, a large number of data such as drug metabolism products for diseases need to be analyzed. The traditional manual way of working is not only good, but also difficult to deal with a large amount of data. Artificial intelligence technology can process large amounts of data quickly and accurately through supercomputers and intelligent algorithms, and predictive analysis of data to improve the efficiency and quality of research and development.

The automated characteristics of artificial intelligence technology can provide researchers with greater flexibility and innovation. In the process of drug development, a large number of experiments and data analysis require a lot of time and effort. Artificial intelligence technology can automate experiments, reduce the labor burden of researchers and increase research and development efficiency.

Artificial intelligence technology can provide researchers with more accurate and reliable research results. Making the predictive effect of drugs more ideal and getting more accurate and reliable data can also reduce cost and time of drug development.

Artificial intelligence technology in the pharmaceutical chemistry industry can use computer models based on neural networks and deep learning to predict drug efficacy and side effects as early as possible and quickly screen effective compounds.

2.1. Research Aim

Firstly, the research aims to use machine learning model to do drug classification for suitable patients by using Python.

Secondly, the aim of the research is to solve the problem of the sharp growth in patients and using machine learning model to reduce the work for doctors to predict suitable drugs for patients.

The last, to provide a useful solution to solve the problem that the usage of medical machine model is too less in the current situation.

The machine learning model in medical field is quite useful for reducing the work for doctors and according to the papers on Baidu Scholar, that a lot of research had been proved. Although the usage of machine learning model in medical use is very limited, they can do well in the field which can help predicting suitable drugs for patients.

As my research to solve the problem, this paper use Python code to make two models which are random forest model and logistic regression model. For the dataset, it includes six columns which includes gender, age, blood pressure level, cholesterol level and drug type. To train the model, this paper use a dataset with 200 datas which 70% for training and 30% for testing. Since the dataset is not large enough, two models are the most suitable to use.

2.2. Data Analyze

To analyze the data, this paper plot a graph which can directly observe the distribution of the type of drugs. The bar chat clearly illustrates the frequency of drug types taken by patients for both males and females. The drug Y is in the first place which both genders had it the most. The result can be seen in figure 1.

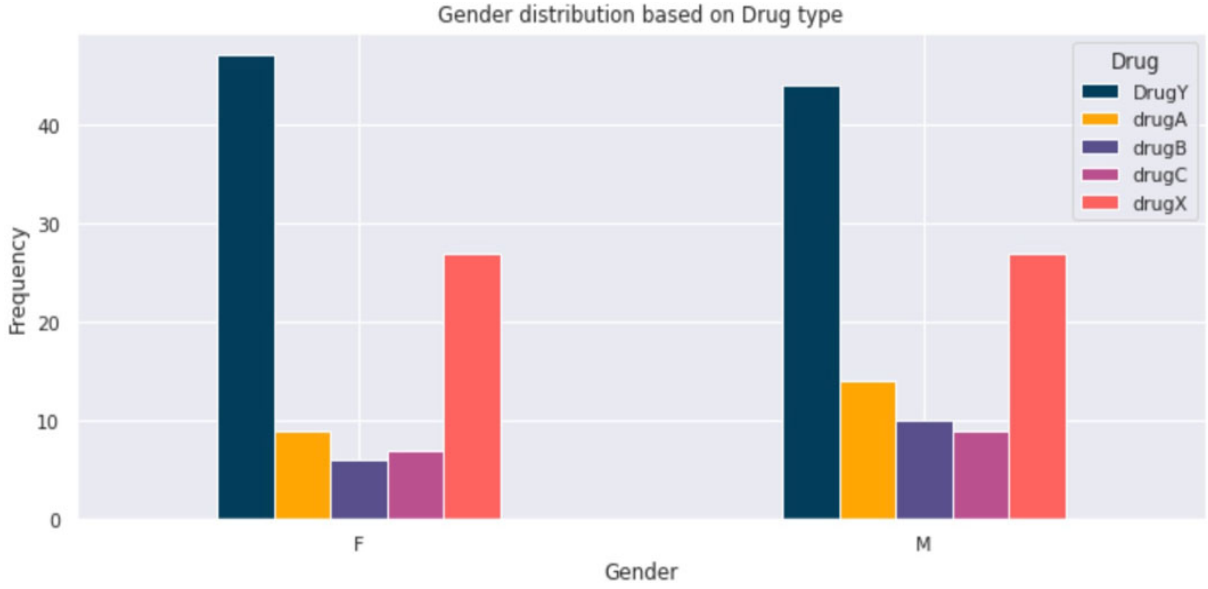

Fig. 1 Bar chat of drugs type against gender

Clemence et al.[6] and Hongming et al.[7] have stated the performance of deep learning models on huge dataset.

According to the accuracy of machine learning model, the logistic regression model has a higher accuracy with 85% than the random forest model with 83%.

These conclusions are summarized in table 1 and table 2. With the dataset with 200 samples, the accuracy illustrates that the random forest model is the most suitable model for the case. The training success of the machine learning model solve the problem of the increasing work for doctors as the population grows directly cause the growth of patients.

Table. 1 Accuracy of Logistic Regression Model

	precision	recall	f1-score	support
DrugY	1.00	0.70	0.82	30
DrugA	0.71	1.00	0.83	5
DrugB	0.75	1.00	0.86	3
DrugC	0.67	1.00	0.80	4
DrugX	0.82	1.00	0.90	18
Accuracy			0.85	60
Macro avg	0.79	0.94	0.84	60
Weighted avg	0.89	0.85	0.85	60

Table. 2 Accuracy of Random Forest Model

	precision	recall	f1-score	support
DrugY	1.00	0.67	0.80	30
DrugA	0.62	1.00	0.77	5
DrugB	0.75	1.00	0.86	3
DrugC	0.67	1.00	0.80	4
DrugX	0.82	1.00	0.90	18
Accuracy			0.83	60
Macro avg	0.77	0.93	0.83	60
Weighted avg	0.88	0.83	0.83	60

For larger datasets, the random forest model also remains the highest accuracy among other models. Zhaoying et al. predict the potential users and estimated the users of drugs using Machine Learning algorithms. They employe XGBoost, Random Forest and LightGBM Machine Learning models for prediction. LightGBM has the best performance score. The dataset has 1885 observations and 12 features. They concluded that demographic information is also one of the key features [8].

3. Discussion

Although the application of artificial intelligence technology has brought many substantial improvements and breakthroughs, there are still some limitations and challenges in the field of drug development. Specifically, these limitations include.

1. Lack of good data source. The amount of data involve in drug development is huge. Its quality and accuracy are often determined by the stability and practicality of the data source. This puts forward higher requirements for the application of artificial intelligence technology. Therefore, how to effectively acquire and manage large amounts of reliable data is crucial to the application of artificial intelligence technology.
2. Uncertainty and risk. Drug development tends to have a low success rate, at the same time, drug discovery and research and development often take a long time and may even have certain uncertainties and risks. Although the application of artificial intelligence technology can effective improve the efficiency and success rate of drug development, it cannot completely eliminate the risks and uncertainties involved.
3. Generalization problem of machine learning techniques. Due to the high complexity of chemical and biological experiments in drug development, in the application of machine learning to drug classification and design, there may be some challenges including overfitting and data bias, etc. It is the problem to be solved.

In general, artificial intelligence technology is used in drug design and classification, despite some limitations and challenges, some results and progress have been achieved. Issues such as equitable benefits, unemployment, patient privacy, medical security, division of responsibility and regulation.

The reasons may include non-compliance with basic ethical principles, technical deficiencies, lack of legislation and regulation, implicit algorithmic bias, poor data quality, and lack of public literacy. Solutions to the causes are proposed, including clear human priority, transparent and traceable faults, representative samples and researchers to avoid bias, and effectively improve the quality of medical services on the premise of reducing the repetitive work of medical staff to improve efficiency and reduce misdiagnosis and missed diagnosis. Artificial intelligence technology can model and simulate the structure and function of drugs to provide researchers with reference and optimization. For example, researchers can find better structural designs by simulating the mechanism of action of drugs in the body, thus accelerating the process of drug innovation and optimization. In addition, artificial intelligence technology can also predict and analyze the characteristics and conformation of drug compounds, helping researchers better understand the interaction and chemical reactions between compounds, so as to better guide the drug development process. The process of drug development involves screening and analyzing millions of compounds to find effective drug candidates. The process often takes years and is expensive to test. Artificial intelligence technology can use algorithms such as machine learning to analyze and predict a large number of compounds and pharmacological data, quickly determine good drug candidate data, and analyze possible side effects and safety hazards, so as to improve the safety and effect of the business. The application scenario of medical artificial intelligence technology during the epidemic covers the risk assessment of diseases and has been widely used in public health and medical fields from the beginning to the present, and artificial intelligence continues to play a role in the establishment of digital health libraries and the expansion of telemedicine [9]. In the future research, this field can learn more about the advantages and opportunities of artificial intelligence technology. In addressing drug research and classification, it could continue to improve efficiency and success rates to better serve the medical needs of public. An experience proves that the proposed methodology with reduced feature elimination has increased the prediction accuracy of the Machine learning models. Initially, the 150 observations with 21 were used for the experiment. The dataset was pre-processed, and several categorical data were converted to numerical values for the Machine learning training procedure. The dataset is a balanced dataset of 50 observations for each class. The samples are trained with the three models like LR, RF and DT [10].

4. Conclusion

In conclusion, the research uses the machine learning model to do drug classification based on the blood pressure level, cholesterol and age of the patients to make outcomes of suitable drugs. On the other hand, by using machine learning model, doctors could reduce the human error and to avoid medical negligence which can help increasing the efficiency of them. Using artificial intelligence technology, the drug development process can be faster and more accurate, and the quality and safety of drugs have a higher guarantee. Of course, the application of artificial intelligence technology also has to face these challenges, including the quality and diversity of data, and the establishment of machine learning algorithms. With the development of medical artificial intelligence, it still faces many problems and challenges at the theoretical level, technical level, ethical legal level and application level. The adverse impact of medical artificial intelligence products on patient safety needs to be analyzed from four perspectives: medical artificial intelligence developers, medical institutions, health technicians, and the government. For the future research, this paper should aim to use machine learning model to predict drugs for patients based on other datas such as, weight, elements and diet habit, furthermore, the physical properties and chemical properties of the drugs also influence the choices of the drug types. Artificial intelligence can scientifically manage and utilize the consulting room of the hospital, which relieves the problem of the shortage of medical resources. However, due to the constraints of many factors, the application of artificial intelligence in the field of medical health has encountered many problems. Artificial intelligence technology has developed rapidly, and it has completed the landing application in the corresponding scene in many industries.

The pharmaceutical industry has the characteristics of strong product specificity, high high-tech absorption capacity, high investment and high risk, so it is very suitable for the use of artificial intelligence technology to develop new research and new technologies. Meanwhile, the route of taking drugs is another point which influence the outcomes of the suitable drugs, collecting the data for situations when patients having drugs with their skin or with other ways and test the data with the machine learning models. In this way, the machine learning model can predict all kinds of drugs for different patients.

References

[1] G. Shobana and S. N. Bushra, "Drug Administration Route Classification using Machine Learning Models," *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, 2020, pp. 654-659, doi: 10.1109/ICISS49785.2020.9315975.

[2] Wei Zhang,Jing Shang,Lei Guo et al.The invention relates to an intelligent prescription review system and method based on artificial intelligence .CN202211133766.8[2023-08-08].

[3] A. I. Saad, Y. M. K. Omar and F. A. Maghraby, "Predicting Drug Interaction with Adenosine Receptors Using Machine Learning and SMOTE Techniques," in IEEE Access, vol. 7, pp. 146953-146963, 2019, doi: 10.1109/ACCESS.2019.2946314.

[4] Jiawei et al., “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, (2006)

[5] YangGuang Zhao.Research on Medical artificial Intelligence technology and application [J]. Information and Communication technology 2018,12(3):5.DOI:CNKI:SUN:OXXT.0.2018-03-008.

[6] ClémenceRéda,EmilieKaufmann,AndréeDelahaye-Duriez,Machine learning applications in drug development,Computational and Structural Biotechnology Journal,Volume 18,2020,Pages 241-252, doi.org/10.1016/j.csbj.2019.12.006.

[7] Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, Thomas Blaschke, The rise of deep learning in drug discovery, Drug Discovery Today,Volume 23, Issue 6,2018,Pages 1241-1250, doi.org/10.1016/j.drudis.2018.01.039.

[8] Z. Qiao, T. Chai, Q. Zhang, X. ZhouandZ.Chu,"Predictingpotential drug abusers using machine learning techniques," 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Shanghai, China, 2019, pp. 283-286, doi: 10.1109/ICIIBMS46890.2019.8991550. /x

[9] XiaoQi Duan,Qin Zhou,Qi Zhang et al. Application and research progress of digital health care in post-epidemic era [J]. Data mining, 2021, 11(4):7.DOI:10.12677/HJDM.2021.114018.

[10] Pangetal,“IntroductiontoDatamining”,PearsonIndiaEducationPvt Ltd, (2016).