WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Symptoms-Based Drug Prediction: A Machine Learning Approach

***M. Gagan Aditya***

UG student, GMR Institute of Technology, Rajam

**ABSTRACT:**

The healthcare centers are killing lot of time of the patients, even if it is a very common health problem. Treatment in hospitals for general health issues like cold, cough, fever, etc., is very time consuming and cost is also high. When we go to the doctor with a health issue, they will give Prescription for that problem, but when the patient do not have the availability of the doctor immediately, they may use the online prescription which is generated by AI or Machine Learning Model. By using the online prescription, they can save their time and money. In the previous models Set-to-set comparison method is used to predict the drug or medicine, but Machine Learning models may help to produce the prescription more accurately, for the patient by looking at their symptoms. For this model of drug prediction, decision trees will be used to serve as an important resource and assist patients when the doctor is not available.

Keywords: Drug, Prescription, Machine Learning, Symptoms, Multinomial Naïve Bayes, Set-to-set.

## Introduction:

Healthcare and pharmaceuticals are becoming increasingly important in today's rapidly evolving society The technology is used to combat almost all known diseases. Due to the current status Due to their environment and lifestyle, people today are exposed to a variety of diseases. Second It is estimated that over 70% of Indians are at risk of common illnesses such as flu, cold, cough, and viral infections every two months. Because many people ignore common health problems could be a more dangerous byproduct of reality. Lack of early detection of many diseases, including cancer, diabetes and others, are the leading cause of increasing causes of death worldwide. Therefore, it is important to detect the disease as early as possible in order to maintain a strategic distance from it. Machine learning plays an important role in detecting diseases early. Machine Learning is the understanding of a computer system in which a machine learning model learns from data and experience. Training and testing are two iterations of a machine learning algorithm. M.L. model with we can build models to cleanse, process data and deliver results faster. machine learning isn't just about finding diseases early; it's also about personalized medicine. This means creating treatment plans that fit each person based on their genes, lifestyle, and specific health issues. By doing this, we can make treatments more precise and effective while reducing unwanted side effects. It's like having a medical plan that's customized just for you. In this machine learning there are many methods to construct the model like SVM, Decision Trees, Linear Regression, Random Forest, Gaussian Navie Bayes, etc. These algorithms can identify subtle patterns and associations that may not be apparent through traditional analysis methods. Symptom-based drug prediction, as previously mentioned, is a notable application of machine learning in healthcare. By analysing symptoms and patient data, machine learning models can predict which drugs are likely to be most effective for a particular individual. Early disease detection is crucial, as it allows for timely intervention and treatment, significantly improving patient outcomes. When using machine learning in healthcare, it's crucial to think about ethics and privacy. This means being careful with how we handle patients' information and making sure that the decisions made by algorithms are clear and fair. Protecting patient privacy and being open about how these technologies work is very important. As technology gets more advanced in healthcare, we need to make sure we're using it in a way that's safe, respectful of people's privacy, and easy for everyone to understand.

What is Drug Prediction?

Drug prediction utilizes computer-based methods and data analysis to anticipate the effects of a drug within the body. This involves determining the specific body parts the drug will interact with, assessing potential harmful effects, and forecasting how the body will process the drug. Scientists employ various techniques, including machine learning, to make these anticipations. They also examine the chemical structure of a drug to establish connections with its effectiveness and potential side effects. This accelerates the discovery and development of new drugs by making informed estimates about their behavior in the body. However, it's crucial to acknowledge that these predictions are akin to intelligent estimations, and actual experiments are necessary to validate their accuracy.

## Methodology:

We can use so many methods for the prediction of the drugs based on the symptoms of the user in the machine learning. We can select the method based on the features, which are present in the datasets that we consider to implement. So. Based on the developer they consider the different types of methods

like Support Vector Machine (SVM), decision Trees, Random Forest, Logistic Regression, Multinomial Naïve Bayes, K-Nearest Neighbor (KNN), Linear Regression, K-Means Clustering, Multilinear Regression, etc.so, there is a lot of methods to implement the drug prediction system.

***Support Vector Machine (SVM):***

A Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. It aims to find a hyperplane that best separates data points into distinct classes while maximizing the margin between them. The data points closest to this hyperplane are called support vectors. SVM is effective in high-dimensional spaces and can handle both linear and non-linear classification tasks using kernel functions. It's known for its ability to handle complex decision boundaries and its resistance to overfitting. SVM has applications in image classification, text classification, and bioinformatics, among others. The primary goal of SVM is to find the optimal hyperplane that minimizes classification errors, making it a robust choice for various real-world problems.

***Decision Trees:***

A decision tree is a machine learning algorithm used for both classification and regression tasks. It resembles an upside-down tree with nodes and branches. Each node represents a decision or a feature, and each branch corresponds to the possible outcomes of that decision. Starting from the root node, the algorithm makes binary decisions to traverse the tree until reaching a leaf node, which provides the final prediction or outcome. Decision trees are interpretable and easy to visualize. They can handle both categorical and numerical data and are prone to overfitting, which can be mitigated through techniques like pruning. Random Forests and Gradient Boosting are ensemble methods that use multiple decision trees to improve predictive accuracy. Decision trees are widely used in various fields, including healthcare, finance, and marketing.

***Random forest:***

A Random Forest is like a team of decision trees working together to make predictions. They prevent making guesses that are too specific, which could be wrong. Instead, they each focus on different parts of the problem. Each tree suggests an answer, and the final decision is made by what most of the trees think. This teamwork helps to predict things better. Random Forest can handle different types of problems, like sorting things into groups or predicting numbers. It's also good at dealing with confusing or strange data, and it can tell you which parts of the data are most important for making predictions. People use Random Forest in many areas, like deciding what's in a picture, giving recommendations, and understanding DNA. It's like a smart and reliable helper in machine learning.

***Logistic Regression:***

logistic regression is applied to identify the new recommendation from this dataset is valid or not. Basically, it checks the accuracy of the dataset. First, we import the Logistic Regression module and create a classifier object using the function Logistic Regression(). Then, using fit() function, we fit the model on train dataset. After that using predict() function, we perform prediction on test dataset . Then, we use confusion matrix for model evaluation. The model was evaluated using model evaluation metrics such as accuracy, precision, F1 score, and recall.

***KNN:***

K-Nearest Neighbors (KNN) is a popular and easy machine learning method for sorting things or predicting values. It works by looking at nearby examples in a bunch of features and deciding on a category based on what most of its close neighbors are. The 'k' you choose tells the algorithm how many neighbors to check, affecting how sensitive it is to local details. KNN doesn't assume anything specific about how the data is spread out, which is called being non-parametric. It works well when there are clear groups in the data and not too much noise. However, it might not do as well with lots of features or noisy data. KNN is straightforward and good for quickly trying out ideas with data. People often use it in recommending things, recognizing images, and areas where finding similar stuff is key for accurate predictions.

***Multinomial Naïve Bayes:***

Multinomial Naive Bayes is a helpful tool for sorting and understanding text. It's like a smarter version of the Naive Bayes algorithm, specially built for things like counting words in documents. The "multinomial" part just means it's good at dealing with lots of different categories. Even though it keeps things simple and assumes the words in a text are independent, Multinomial Naive Bayes usually works well, especially in jobs like figuring out what type of document you have or filtering out spam emails. It figures out the chances of each word showing up in different types of documents and uses that to make decisions. While it might not be perfect for really complicated situations, it's still a popular and easy choice for tasks involving text because it works well and isn't too hard to use.

## Objective:

i. Efficient Time Savings

ii. Cost-effective Healthcare

iii. Enhanced Prescription Accuracy

## Results

| SN O | AUTHOR | DATAS ET | METHODS | ACCURACY | PRECISI ON | SENSTIVIT Y | SPECIFI TY |
|---|---|---|---|---|---|---|---|
| 1 | Usharani Bhimavarapu | They collect the datasets from WHO, drug banks, news reports | Random forest | 0.841 | 0.840 | 0.841 | 0.920 |
| | | | KNN | 0.840 | 0.823 | 0.824 | 0.915 |
| | | | SVM | 0.719 | 0.714 | 0.719 | 0.860 |
| | | | Logistic Regressi on | 0.534 | 0.522 | 0.534 | 0.767 |
| 2 | Mamata Nayak | Dataset from Kaggle over Google Colab platform | Logistic Regressi on | 0.805 | 0.823 | - | - |
| 3 | Suvendu Kumar Nayak | data.wor ld , sympto ms dataset from Newyor k-presbyte rian Hospital , side effects dataset from druglib.c om and the drug review dataset | Multino mial Naïve Bayes | 0.8993 | - | - | - |
| | | | SVM | 0.8993 | - | - | - |
| | | | Decision Tree | 0.8684 | - | - | - |

This is the results of some papers which is taken as the reference. From these papers I understand that some of the methods are mostly repeated ones and some of the methods are not implemented before like multinomial naïve bayes, even though it is also having the accuracy like the repeated methods. So, if we implement the different methods rather than the repeated ones, it can get the more knowledge on the newly implemented methods. Even though the multinomial naive bayes got the same accuracy like the remaining methods used in the same paper. By after seeing all the papers multinomial naïve bayes got the more accuracy like 0.8993. In some other papers some methods got the low accuracy because dataset of same features is not suitable for every model.

## Conclusion

By using smart computer programs like AI and Machine Learning to create online prescriptions seems like a good way to fix the problems and high costs of regular healthcare, especially for common health issues. Multinomial Naïve Bayes, a part of these programs, help make more accurate predictions about which medicines a person might need based on their symptoms, especially when a doctor isn't available right away. As healthcare and technology keep getting more connected, finding diseases early on is very important. Machine learning tools, like s Naive Bayes, Random Forest, K Nearest Neighbor, Gaussian Naive Bayes, Logistic Regression, Support Vector Machine, and Multinomial Naïve Bayes are like helpful assistants that learn from information and experiences to quickly provide better and easier healthcare solutions. It is widely assumed that the suggested method can reduce illness risk by identifying them early and lower the cost of diagnosis and treatment. However, the choosing of symptoms has a significant impact on drug prediction accuracy. We can also build the drug prediction system by using the Deep learning methods.

**References:**

List all the material used from various sources for making this project proposal

Research Papers:


1. Nayak, S. K., Garanayak, M., Swain, S. K., Panda, S. K., & Godavarthi, D. (2023). An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms. *IEEE Access*
2. Palanivinayagam, A., & Sasikumar, D. (2020). Drug recommendation with minimal side effects based on direct and temporal symptoms. *Neural Computing and Applications*, *32*, 10971-10978.
3. Ponraj, T. E., & Charles, J. Detection and Prediction of Adverse Drug Reactions Using Data Driven Modelling with Machine Learning Models.

4. Tan, Y., Kong, C., Yu, L., Li, P., Chen, C., Zheng, X., ... & Yang, C. (2022, August). 4sdrug: Symptom-based set-to-set small and safe drug recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3970-3980).

*5*. Kuang, X., Wang, F., Hernandez, K. M., Zhang, Z., & Grossman, R. L. (2022). Accurate and rapid prediction of tuberculosis drug resistance from genome sequence data using traditional machine learning algorithms and CNN. *Scientific reports*, *12*(1), 2427.

6. Granda Morales, L. F., Valdiviezo-Diaz, P., Reátegui, R., & Barba-Guaman, L. (2022). Drug recommendation system for diabetes using a collaborative filtering and clustering approach: development and performance evaluation. *Journal of Medical Internet Research*, *24*(7), e37233.

7. Bhimavarapu, U., Chintalapudi, N., & Battineni, G. (2022). A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN. *Algorithms*, *15*(6), 186.

8. Mohapatra, M., Nayak, M., & Mahapatra, S. (2022). A Machine Learning based Drug Recommendation System for Health Care. *Graduate Research in Engineering and Technology (GRET)*, *1*(6).

9. Gupta, J. P., Singh, A., & Kumar, R. K. (2021). A computer-based disease prediction and medicine recommendation system using machine learning approach. *Int J Adv Res Eng Technol (IJARET)*, *12*(3), 673-683.

10. Rashid, M., Goyal, V., Parah, S. A., & Singh, H. (2022). Drug prediction in healthcare using big data and machine learning. In *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 1062-1071). IGI Global

11. Mudaliar, V., Savaridaasan, P., & Garg, S. (2019). Disease prediction and drug recommendation android application using data mining (virtual doctor). *International Journal of Recent Technology and Engineering*, *8*.

12. Rashid, M., Yousuf, M. M., Ram, B., & Goyal, V. (2019, April). Novel big data approach for drug prediction in health care systems. In *2019 International conference on automation, computational and technology management (ICACTM)* (pp. 325-329). IEEE.

13. Lio, P., & Dimitri, G. M. (2017). DrugClust: A machine learning approach for drugs side effects prediction.

14. Huang, C., Mezencev, R., McDonald, J. F., & Vannberg, F. (2017). Open-source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PLoS One*, *12*(10), e0186906.

15. Bajor, J. M., & Lasko, T. A. (2016, November). Predicting medications from diagnostic codes with recurrent neural networks. In *International conference on learning representations*