

AI-pharmaceutical Intelligence System: A Machine Learning Approach for Symptom-Based Drug Recommendation and Predictive Analytics

SAURAV SAID¹, SUJAL SAWARDEKAR², AAYUSH PADMAWAR³, ASST. PROF. YOJANA DEHANKAR⁴

^{1, 2, 3, 4}*Department of Artificial Intelligence and Data Science, New Horizon Institute of Technology & Management, Thane, India*

Abstract- The pharmaceutical industry today faces several challenges such as scattered information, lack of price transparency, and limited awareness of alternative medicines. In many situations, patients depend on branded drugs without knowing that similar and more affordable options exist. At the same time, healthcare providers also face difficulties due to changing drug availability and pricing patterns. Most existing digital tools focus only on specific tasks like symptom checking or maintaining drug databases, but they do not provide a complete solution. This paper presents an AI Pharma Intelligence System that combines machine learning, natural language processing, and data analytics into a single platform. The system allows users to enter symptoms in text form and predicts possible therapeutic categories using TF-IDF and Logistic Regression. Based on this prediction, it suggests relevant medicines and also provides alternative options that are more cost-effective. An additional feature of the system is the prediction of drug discontinuation. This module uses factors such as manufacturer frequency, side effects, and market competition to estimate whether a drug may be discontinued in the future. The system is developed using Streamlit and Scikit-learn, making it lightweight and easy to use.

Index Terms- Artificial Intelligence, Machine Learning, Drug Recommendation, Symptom Analysis, Pharmaceutical Analytics

I. INTRODUCTION

With the rapid growth of digital healthcare systems, a large amount of pharmaceutical data has become available. However, this information is often scattered across different sources such as manufacturers, pharmacies, and healthcare providers. Because of this fragmentation, users do not always get complete or clear information, especially regarding medicine pricing and alternative options.

Another important issue is that understanding drug-related data is not easy for common users. Medical terms, compositions, and side effects can be confusing. Even healthcare professionals may find it difficult to access all relevant information in one place. Most existing systems provide only basic functionality and do not offer predictive insights or intelligent recommendations.

Recent developments in Artificial Intelligence have shown that it is possible to analyze large datasets and extract useful patterns. Techniques such as machine learning and natural language processing can help in understanding user inputs and generating meaningful outputs. However, many existing solutions focus only on single tasks, such as disease prediction or drug recommendation, rather than combining multiple features.

In this paper, we propose an integrated system that brings together symptom analysis, medicine recommendation, cost-based alternatives, and predictive analytics. The aim is to create a practical tool that can be useful for both users and researchers.

II. RELATED WORK

A. Literature Study

Several studies have explored the use of AI in healthcare applications. Some research focuses on using natural language processing techniques to map symptoms to diseases. These methods show that techniques like TF-IDF and similarity measures can be effective, but they often lack deeper understanding of context.

Other studies have used machine learning models such as Logistic Regression and Naïve Bayes for

disease prediction. These models are simple and perform well when trained on good quality data.

There is also research on drug recommendation systems that use composition similarity to suggest alternatives. These systems are helpful in reducing cost but require regularly updated datasets.

Some papers highlight the issue of price transparency in the pharmaceutical market and suggest that intelligent systems can improve decision-making. In addition, AI-based assistants have been developed to improve accessibility, but they still require human supervision.

Recent work has also explored drug lifecycle prediction, including the possibility of discontinuation. Factors such as competition, side effects, and manufacturer presence have been found to influence these trends.

B. Workflow Structure

The workflow of the proposed system starts when the user enters symptoms in text form. This input is processed using natural language techniques and converted into numerical form using TF-IDF.

The processed data is then passed to a machine learning model, which predicts the therapeutic category. Based on this prediction, the system retrieves relevant medicines from the dataset.

To improve the quality of results, similarity matching is used. The system also suggests alternative medicines based on composition and cost. At the same time, a prediction module evaluates the possibility of drug discontinuation.

Finally, all results are displayed to the user through a simple interface.

III. METHODOLOGY

The system follows a step-by-step approach. First, the input data is preprocessed to remove unnecessary words and standardize the text. After preprocessing, TF-IDF is used to convert the text into numerical values.

After feature extraction, Logistic Regression is used for classification. This model calculates the probability of each class and selects the most likely one.

The system also includes a module for suggesting alternative medicines. These alternatives are selected based on similar composition and lower price.

Another important feature is the discontinuation prediction module. This uses features such as number of manufacturers, side effects, and competition to estimate risk.

IV. SYSTEM ARCHITECTURE

Here The system is designed using a layered approach.

Data Layer: Stores pharmaceutical data such as medicine name, composition, price, and side effects

Processing Layer: Handles text preprocessing and feature extraction

Machine Learning Layer: Performs prediction and similarity matching

Application Layer: Manages recommendation and filtering logic

Presentation Layer: Provides user interface using Streamlit

Analytics Layer: Supports visualization using tools like Power BI

This structure makes the system flexible and easy to extend in the future.

IV. SEQUENCE FLOW AND DESIGN ANALYSIS

The system is designed in a modular way so that each component can be updated independently. First user enters symptoms. Then system processes text and converts it. After that model predicts category. Then medicines are shown. Alternatives are also suggested.

At last, risk prediction is done and everything is shown to user.

Design

The design is kept simple. Each part works separately. It is easy to add new features later.

Advantages:

Easy to use. Can be improved later. Fast working.

Limitations:

Depends on data. Not fully accurate always. Not a replacement for doctor

VI. DISCUSSION AND FUTURE WORK

The system works well for basic prediction and recommendation. It is helpful for users to get idea about medicines and alternatives.

The cost-based suggestion is useful because many people look for cheaper options. Discontinuation feature is also different from other systems.

But system accuracy depends on dataset. Also, it does not understand deep medical context.

In future, we can improve system by:

- Using deep learning models
- Adding real-time data
- Connecting with hospital systems
- Making chatbot interface
- Deploying on cloud

V. CONCLUSION

In this paper, we created a system that helps in symptom-based drug recommendation. It combines machine learning and simple tools to give useful output. It also provides cheaper alternatives and predicts future risk. This makes it more useful than basic systems. The system can be improved further, but it gives a good starting point for AI in healthcare.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide, Asst. Prof. Yojana Dehankar, for her

guidance and support throughout this work. We also thank the Department of Artificial Intelligence and Data Science, New Horizon Institute of Technology & Management, Thane, for providing the necessary resources and environment to complete this project. We are also grateful to our friends and colleagues for their valuable suggestions and support.

REFERENCES

- [1] J. Devlin, M. Chang, K. Lee and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of NAACL-HLT, 2021.
- [2] T. Chen, H. Yin, H. Chen and Z. Wang, Disease prediction using machine learning over big healthcare data, IEEE Access, vol. 9, 2021, pp. 123–135.
- [3] S. Shrestha and S. Pokhrel, A comparative study of machine learning algorithms for healthcare prediction, Applied Sciences, vol. 11, no. 5, 2021.
- [4] M. S. Uddin, A. A. Moni, S. Ahamed and M. S. Islam, Machine learning based drug recommendation system: a review, Journal of Healthcare Engineering, 2021.
- [5] R. Miotto, F. Wang, S. Wang, X. Jiang and J. T. Dudley, Deep learning for healthcare: review, opportunities and challenges, Briefings in Bioinformatics, vol. 22, no. 6, 2021.
- [6] S. K. Sharma and X. Wang, Live data analytics for healthcare systems using machine learning, IEEE Access, vol. 9, 2021, pp. 456–470.
- [7] Y. Koren, R. Bell and C. Volinsky, Advances in recommender systems, IEEE Computer, vol. 54, no. 6, 2022, pp. 42–49.
- [8] A. Rajkomar, J. Dean and I. Kohane, Machine learning in medicine, New England Journal of Medicine, vol. 386, 2022.
- [9] H. Liang, X. Tsui, H. Ni et al., Evaluation and accurate diagnoses of pediatric

diseases using artificial intelligence,
Nature Medicine, vol. 28, 2022.

- [10] S. Raschka, V. Mirjalili and others,
Machine Learning with PyTorch and Scikit-
Learn, Packt Publishing, 2023.