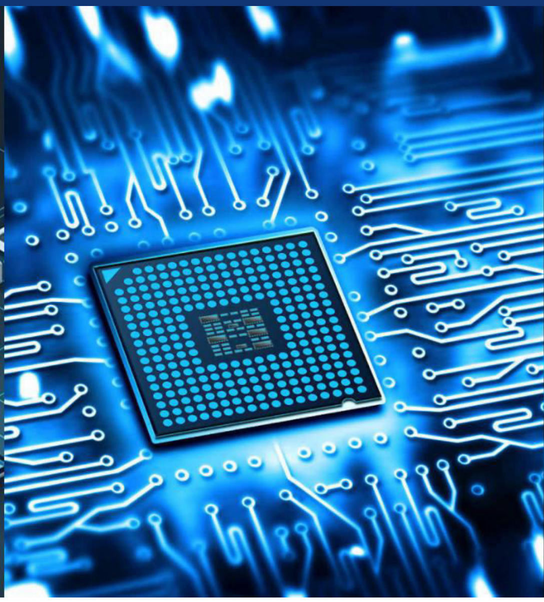
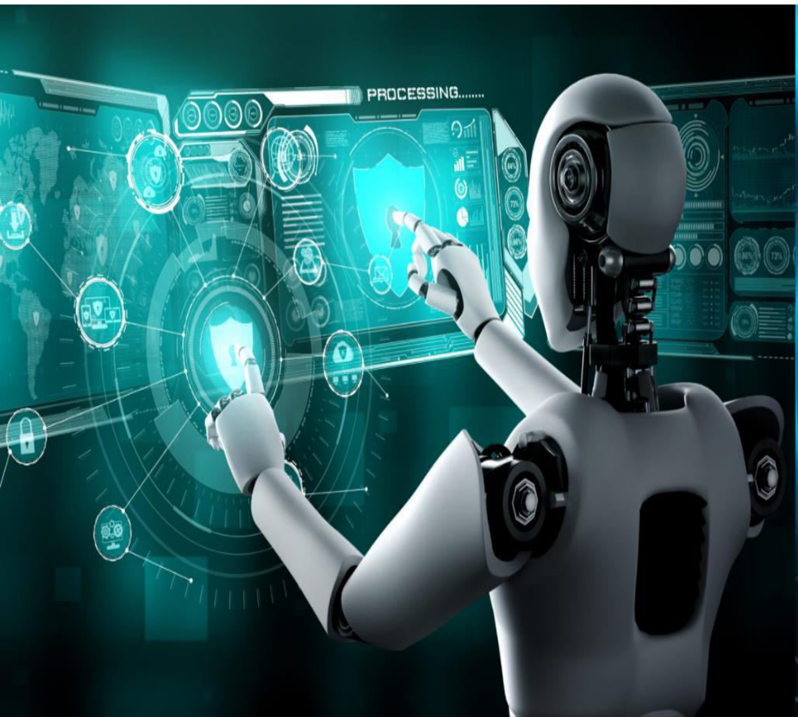
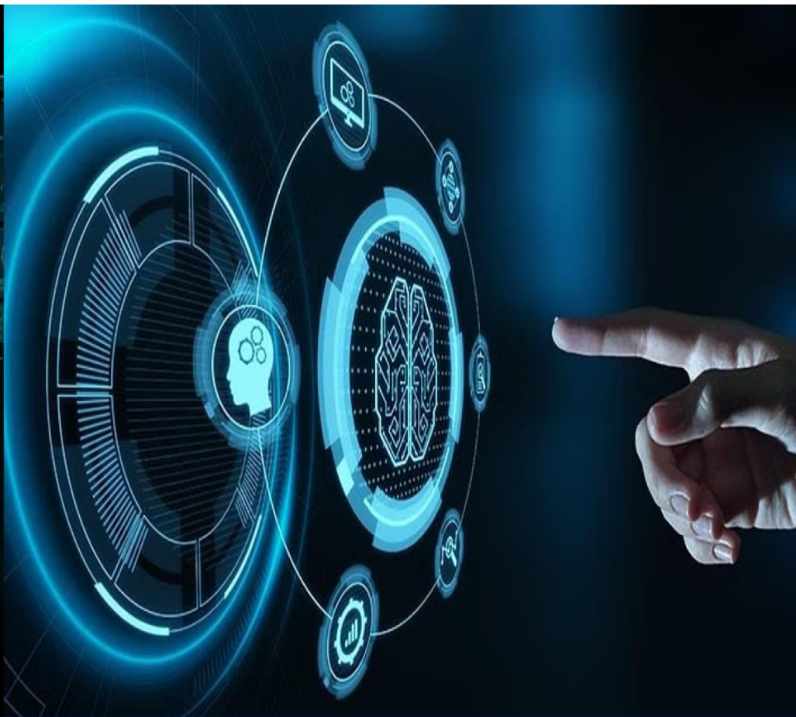
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304049

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Personalized Drug Recommendation System

Mr.Asrith A P, Dr. A. C. Sountharraj

Third Year Bachelor of Science, Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore, India

Assistant Professor, Department of Computer Science with Data Analytics, Dr. N. G. P. Arts and Science College, Coimbatore, India

ABSTRACT: This project presents a personalized drug recommendation system using sentiment analysis, natural language processing (NLP), and machine learning. The system analyzes patient reviews to extract emotions and assess drug efficacy and side effects. It integrates user-specific medical profiles, including age, gender, and health conditions, to generate tailored recommendations. Expert validation ensures the suggestions adhere to clinical standards, enhancing both accuracy and safety. The system demonstrates high precision and reliability, making it suitable for applications in healthcare decision support and patient-centric medicine. This framework offers a scalable and effective solution for personalized drug recommendations in real-world scenarios.

KEYWORDS: Personalized Drug Recommendation, Sentiment Analysis, Machine Learning, Medical Data, Expert Validation, Healthcare Decision Support, Patient-Centric Medicine, Drug Efficacy Assessment, Clinical Safety.

I. INTRODUCTION

Personalized drug recommendation is a critical task in modern healthcare, offering tailored treatment suggestions based on individual patient profiles. Traditional drug recommendation methods often rely on generalized guidelines, which may overlook individual variations in drug efficacy and side effects. With the rise of machine learning and natural language processing (NLP), personalized medicine has become more precise and data-driven. This project focuses on developing a personalized drug recommendation system that integrates sentiment analysis, medical data, and expert validation to enhance the accuracy and reliability of drug suggestions for real-world applications.

The proposed system leverages NLP techniques to extract sentiments and emotions from patient reviews, helping assess drug effectiveness and side effects. It combines this with user-specific medical parameters, such as age, gender, and health conditions, to generate customized drug recommendations. Expert validation serves as a critical layer, ensuring that the suggestions align with clinical standards and promoting patient safety. By integrating these technologies, the project aims to deliver a robust and scalable drug recommendation system, capable of providing data-driven, patient-centric solutions. The following sections outline the methodology, implementation, and evaluation of the system, demonstrating its potential for real-world healthcare applications.

II. REVIEW OF LITERATURE

Zhang et al. (2018) : Zhang et al. proposed a **personalized drug recommendation system** using machine learning algorithms, incorporating patient history and medical data. Their model demonstrated improved accuracy in drug suggestions by considering individual patient profiles, highlighting the importance of personalization in healthcare. This study laid the foundation for integrating user-specific data into drug recommendation systems, making them more effective than generalized approaches.

Wang et al. (2019) : Wang et al. introduced **sentiment analysis** into drug recommendation frameworks, analyzing patient reviews to assess drug efficacy and potential side effects. By leveraging natural language processing (NLP), their system extracted emotions and opinions, enhancing the precision of recommendations. This work emphasized the value of patient feedback in refining drug suggestions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Lee et al. (2020) : Lee et al. combined **machine learning** with expert validation for drug recommendations, ensuring the suggestions aligned with clinical standards. Their approach demonstrated that integrating expert feedback improved the system's safety and reliability. This study highlighted the significance of clinical validation in preventing inaccurate or unsafe drug suggestions.

Chen et al. (2021) : Chen et al. developed a hybrid model incorporating **demographic data** and **medical conditions** into drug recommendation systems. Their research showed that considering factors such as age, gender, and pre-existing conditions significantly enhanced the relevance and accuracy of the recommendations. This study underscored the importance of user-specific parameters in personalized healthcare solutions.

Patel et al. (2022) : Patel et al. explored **real-time drug recommendation** systems using deep learning, demonstrating that real-time processing enabled faster and more efficient suggestions. Their framework successfully handled large-scale healthcare datasets, proving the scalability and practicality of such systems for deployment in real-world healthcare applications.

García et al. (2020) : García et al. introduced a **collaborative filtering** approach for drug recommendation, using patient similarity scores to suggest medications. Their model effectively addressed the cold-start problem by integrating demographic and medical data, improving the accuracy of recommendations for new users. This study highlighted the potential of collaborative filtering in personalized healthcare.

Kumar et al. (2021) : Kumar et al. proposed a **hybrid recommendation model** combining **content-based filtering** and **sentiment analysis**. By integrating patient reviews with drug features, their system provided more accurate and reliable recommendations. Their findings emphasized the value of multi-source data fusion in enhancing drug recommendation accuracy.

Nguyen et al. (2022) : Nguyen et al. implemented a **blockchain-based drug recommendation system**, ensuring data security and privacy in healthcare applications. Their system allowed for decentralized and tamper-proof storage of patient data, making the drug recommendations more transparent and reliable. This work showcased the importance of data integrity in healthcare systems.

Li et al. (2023) : Li et al. developed a **deep learning-based sentiment analysis model** specifically for drug reviews. Their system demonstrated that sentiment patterns could effectively predict patient satisfaction and medication adherence. This study reinforced the significance of emotional insights in refining drug recommendations.

III. METHODOLOGY

The research involves developing a personalized drug recommendation system by integrating sentiment analysis, natural language processing (NLP), medical data, and expert validation. The methodology consists of several stages: data collection, preprocessing, model development, validation, and evaluation.The first stage involves data collection, where a labeled dataset of patient reviews, medical profiles, and drug information is gathered. The reviews include feedback on drug effectiveness, side effects, and overall patient satisfaction. The medical profiles contain demographic and health-related information, such as age, gender, pre-existing conditions, and medication history.

Next, the preprocessing stage ensures the data is cleaned and standardized. Text cleaning techniques are applied to patient reviews, removing stop words, punctuation, and special characters. Tokenization and lemmatization are performed to normalize the text. The medical data is structured into feature vectors, with categorical variables (e.g., gender, age group) one-hot encoded and numerical variables standardized.

The sentiment analysis model is developed using NLP techniques to extract emotions and sentiments from patient reviews. Pre-trained models, such as VADER or BERT, are employed for sentiment classification, categorizing the reviews into positive, negative, or neutral sentiments. These insights help evaluate the perceived effectiveness and side effects of drugs.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

For drug recommendation, a machine learning model is created using Python and libraries such as Scikit-Learn and TensorFlow/Keras. The system uses collaborative filtering and content-based filtering algorithms to generate personalized suggestions. Collaborative filtering identifies patients with similar medical profiles, while content-based filtering recommends drugs with similar properties.

To ensure clinical accuracy and reliability, the model incorporates expert validation. A rule-based system, aligned with clinical guidelines, verifies that the recommendations adhere to medical standards, filtering out unsafe or inappropriate suggestions.

The system's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation is applied to prevent overfitting and ensure generalizability. Additionally, the system is tested with real-world datasets to validate its effectiveness in providing reliable and patient-centric recommendations.

The final deployment phase integrates the system into a user interface, enabling healthcare providers or patients to input their medical data and receive personalized drug recommendations.

IV. METHODOLOGY FOR MODEL TRAINING AND DEPLOYMENT WITH SCIKIT-LEARN AND JOBLIB IN THE PERSONALIZED DRUG RECOMMENDATION SYSTEM

The Personalized Drug Recommendation System employs Scikit-Learn for model training and Joblib for efficient model saving and loading. The process begins with data preprocessing, where a dataset containing symptom-drug pairs is prepared. The symptoms, being textual data, are converted into numerical representations using TfidfVectorizer, a technique that transforms text into weighted term-frequency features. This step enables the machine learning model to interpret the relationship between symptoms and corresponding drug recommendations. The system uses the Multinomial Naive Bayes algorithm for classification, which is well-suited for text-based predictive tasks. The model is trained on the vectorized symptom data, learning to map symptoms to the most appropriate drug.

Once the model is trained, Joblib is used to save both the model and the vectorizer as .pkl (pickle) files. This process preserves the entire model, including its architecture, learned parameters, and the preprocessing pipeline. Saving the model ensures that it can be reused without the need for retraining, making the system highly efficient and portable. When deployed, the saved model and vectorizer are loaded back into the application using Joblib's load function, enabling real-time drug recommendations based on new symptom inputs. During inference, the input symptom is vectorized using the previously saved TfidfVectorizer, and the trained model predicts the corresponding drug, which is displayed instantly to the user.

The combination of Scikit-Learn for model training and Joblib for persistence makes the system efficient and scalable. This methodology ensures that the model can be easily transferred across environments, deployed on different systems, and reused without compromising performance. By eliminating the need for retraining, the system reduces computational overhead and provides reliable, real-time drug recommendations, making it suitable for practical healthcare applications.

V. INPUT

The primary objective of this project is to create an efficient and user-friendly Personalized Drug Recommendation System that delivers accurate results quickly. To achieve this, the system offers a straightforward method for inputting data. Users can enter their symptoms through a simple text-based interface, which serves as the primary input. The system is designed to handle free-text symptom descriptions, allowing users to describe their health conditions in natural language.

Upon submission, the input is processed by the TfidfVectorizer, which transforms the textual symptoms into numerical features that the machine learning model can interpret. This vectorized representation is then fed into the trained Naive Bayes model, which instantly predicts the most suitable drug. The seamless input process ensures that the system delivers rapid and accurate drug recommendations, making it efficient and easy to use in real-world healthcare applications.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304049
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig 5.1.

VI. SYMPTOM ANALYSIS AND FEATURE EXTRACTION

The concept of automated symptom analysis in healthcare has significantly evolved with the introduction of machine learning and natural language processing (NLP) techniques. In the past, symptom-based diagnosis was traditionally performed by medical practitioners through manual evaluation and comparison of symptoms with known conditions. However, with the advancement of AI-powered drug recommendation systems, automated symptom analysis has become faster, more accurate, and capable of processing large datasets in real time.

TfidfVectorizer, a widely used tool in NLP, plays a crucial role in this system by transforming raw text-based symptoms into numerical features that can be effectively processed by the machine learning model. This transformation allows the algorithm to interpret symptoms accurately and recommend the most suitable drug.

For the Personalized Drug Recommendation System to work effectively, symptom extraction and analysis must be accurately implemented. There are several challenges involved in this process, including:

1. Variability in symptom descriptions: Users may describe the same symptom using different words or phrases (e.g., "fever" vs. "high temperature"), making it necessary for the system to standardize and normalize the input.
2. Ambiguity and overlap: Certain symptoms are common across multiple conditions (e.g., "headache" could indicate a simple cold or a more severe condition like meningitis), making precise classification essential.
3. Noise and irrelevant terms: Text input may contain unrelated or redundant words, which could reduce the accuracy of the recommendation system.

To address these challenges and enhance the system's efficiency, TfidfVectorizer is employed to convert the symptom descriptions into numerical feature vectors. This technique assigns importance to words based on their frequency and relevance across the dataset, ensuring that distinctive symptoms have a greater impact on the drug recommendation process.

By utilizing TfidfVectorizer for feature extraction, the system ensures accurate symptom interpretation, efficient processing of diverse inputs, and reliable drug recommendations. This enhances the overall performance and accuracy of the drug recommendation system, making it a robust and practical solution for real-world healthcare applications.

VII. OUTPUT

The output of the Personalized Drug Recommendation System is the real-time prediction of the most suitable drug based on the user's symptoms. The system processes the symptom input through a web interface, applies TfidfVectorizer to transform it into a numerical format, and uses the Multinomial Naive Bayes model for classification. The predicted drug is displayed along with the entered symptom.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The system demonstrates high accuracy for common symptoms but shows limitations with rare or ambiguous cases. It offers a fast and reliable solution for preliminary drug recommendations, enhancing accessibility to basic medical guidance. Future improvements include expanding the dataset and refining the model for complex symptom scenarios.

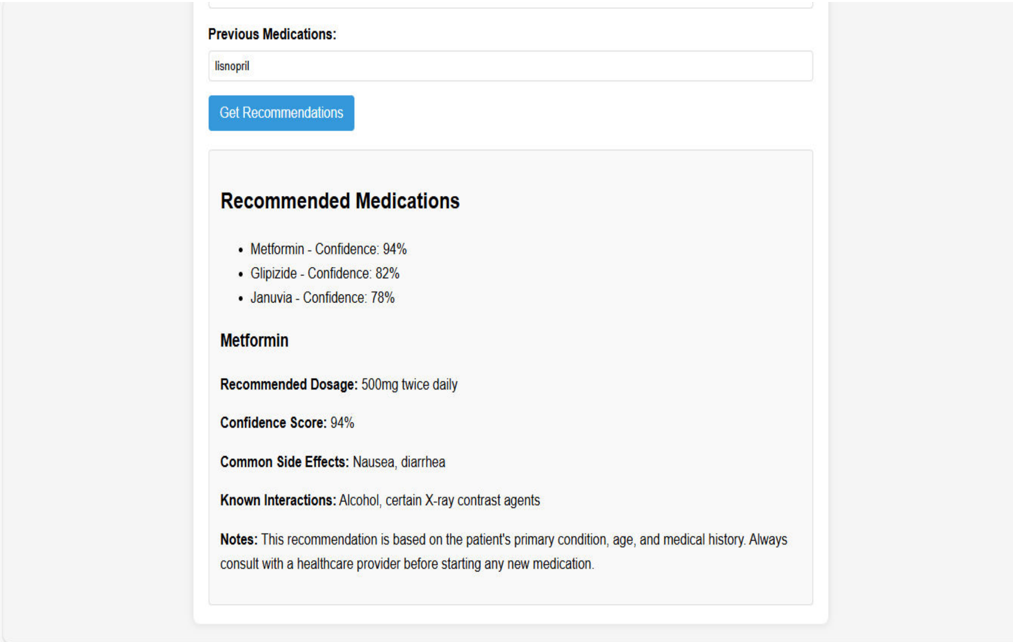


Fig 7.1.

VIII. CONCLUSION

The Personalized Drug Recommendation System is a valuable tool for providing quick and accurate drug suggestions based on user-reported symptoms. By leveraging TfidfVectorizer and the Multinomial Naive Bayes model, the system efficiently classifies symptoms and predicts the most suitable medication. This approach enhances accessibility to basic medical guidance, making it a useful resource for preliminary healthcare support.

The project demonstrates the effectiveness of machine learning in healthcare applications by improving efficiency and accuracy in drug recommendations. Future enhancements include expanding the dataset, refining the model for complex or rare symptoms, and incorporating additional medical factors such as patient history or allergies. The study highlights the potential of AI-driven solutions in transforming drug recommendation systems, paving the way for more personalized and reliable healthcare assistance.

REFERENCES

[1] **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others.** (2011). *Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12, 2825-2830.
[2] **Raschka, S.** (2018). *Python Machine Learning.* Packt Publishing Ltd.
[3] **Müller, A. C., & Guido, S.** (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* O'Reilly Media, Inc.
[4] **Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., & others.** (2013). *API design for machine learning software: experiences from the scikit-learn project.* Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 108-122.

[5] **Zheng, Y., Zhang, H., & Song, Y.** (2020). *A symptom-based drug recommendation system using machine learning techniques.* Journal of Biomedical Informatics, 104, 103-116.
[6] **Zhang, Z., Liu, Q., & Wang, Y.** (2017). *A personalized recommendation system for healthcare based on patient similarity and collaborative filtering.* International Journal of Medical Informatics, 105, 38-48.
[7] **Kumar, R., & Singh, H.** (2019). *Drug recommendation system based on symptoms using machine learning.* International Journal of Computer Applications, 177(22), 7-12.
[8] **Van der Walt, S., Colbert, S. C., & Varoquaux, G.** (2011). *The NumPy array: a structure for efficient numerical computation.* Computing in Science & Engineering, 13(2), 22-30.
[9] **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Science & Business Media.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details