International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Patient Condition-Aware Drug Recommendation and Clinical Review Sentiment Analysis Using NLP & Machine Learning

Dr .T. Syam Sundar Rao[1]
[1]Associate Professor, Department of CSE-Data Science
KKR & KSR Institute of Technology And Sciences,Guntur,India
Email: syamsundar.jes@gmail.com[1]
Sk.Kaleshavali[2] , Sk. Mohammad jani[3] , M.Hemanth Chowdary[4] , P. Hemanth[5]
[2,3,4,5]B.Tech Students,Department of CSE-Data Science
KKR & KSR Institute of Technology And Sciences,Guntur,India
Email: 22jr1a44c3genai@gmail.com [2] ,22jr1a44c4@gmail.com [3], 22jr1a44a1@gmail.com [4], 23jr5a4414@gmail.com [5]

Abstract:
The Health care sector generates large amounts of unstructured data in the form of reviews , observations and feedback about patient condition and medications . A manual analysis is very difficult and time consuming analysis of data by recommending suitable drug recommendations to the patient . In this project , we are using algorithms like Natural Language Processing (NLP) and Machine Learning algorithms .This process involves Analyzing the patient symptoms , diagnosis and opinions through clinical reviews for providing suitable drug recommendation. The main objective of this project is to design and develop a patient condition aware drug recommendation system integrated with clinical sentiment analysis using Natural Language Processing and Machine learning algorithms . This project focuses on structured data such as symptoms,medical history and diagnosis along with clinical data such as patient feedback , doctor notes , and medical reviews .
The proposed solution of this project involves preprocessing and feature extraction from healthcare data using NLP Techniques. The NLP techniques such as Tokenization,lemmatization and vectorization and machine learning models are to classify patient conditions and recommend suitable drugs based on patient condition data . Additionally , sentiment analysis is performed on clinical reviews to assess patient responses and effective treatment , enabling more informed and personalized drug recommendation . The project is implemented by using python based NLP libraries ,scikit-learn for machine learning model development and classification algorithms for predictive analytics. The final outcome of this project is to recommend suitable drugs to the patient.

Keywords: Healthcare, Natural Language Processing, Drug Recommendation System, Sentiment Analysis, Machine Learning, Patient-Centric Care

I. INTRODUCTION
The major goal of the project is to design the providing suitable drugs based on patient condition and clinical review. inappropriate drugs may cause human side effects , ineffective treatment leads to high risk . The patient's condition is being analyzed by Machine learning and Natural Language Processing algorithms .
The Healthcare system generates large amounts of data from electronic medical records , online healthcare platforms and patient reviews . The patients share their experience about medicines including their effective treatment , side effects and overall satisfaction . These reviews contain valuable information and these valuable information can be useful for both sectors and patients for better decisions . However , the Traditional drug recommendation system uses either past medical prescriptions or ratings and does not consider in depth of patient conditions or the opinions expressed in clinical reviews .
With the exponential growth of Natural Language Processing and Machine Learning , it is possible to analyze unstructured data such as patient reviews in an effective manner . The sentiment analysis is used to understand whether the patient experience with a drug is positive , negative or neutral . By applying Natural Language Processing techniques such as text cleaning , tokenization , feature extraction and Machine Learning algorithms for classification , meaningful insights can be extracted from a large volume of health care text data. Natural Language Processing and Machine learning algorithms and techniques can automatically process and analyze such health care data.
The proposed project focuses on building a project that recommends drugs based on patient condition and also analyzing the sentiment of clinical drug reviews . The Main objective of this project is to improve the effectiveness of patient treatment and prevent the side effects of medications and improve the patient satisfaction.

II. LITERATURE OVERVIEW
Several researches have explored drug recommendation systems and sentiment analysis in the healthcare domain to improve the patient condition and effective treatment outcomes . One of the earliest systems, GalenOWL , was focused on recommending systems by classifying structured clinical data to medical standards like ICD–10 and UNII [1] . While GalenOWL was effective only for handling structured medical records , it did not consider patient generated reviews , the reviews about drug effectiveness , side effects and patient satisfaction .

Leilei Sun et al . proposed a data-driven approach for treatment prescription based on clustering of treatment records [2] . This approach mainly focused on treatment history and clinical data , but did not consider patient feedback or sentiment .Similarly , Chandanal et al . (2021) use deep learning techniques and Long Short Term Memory models to analyze the patient reviews [3] . This system worked strongly in understanding the review sentiment and did not fully combine patient condition information with drug recommendation.

Zomorodi et al. introduced RECOMED , is a pharmaceutical recommendation system that combines both clinical and non-clinical data [4]. This system provided comprehensive drug recommendations and did not consider fully patient condition and satisfaction level . Recent research uses the Natural Language Processing techniques used for analyzing unstructured healthcare data , that shows patient reviews and drug effectiveness that do not capture the traditional clinical datasets [5][6] .

In existing approaches , the proposed system focused on integrating patient condition classification with sentiment analysis of clinical reviews using Natural Language Processing and Machine Learning techniques . This proposed system uses multiple classifiers , feature engineering and voting techniques . The proposed system can improve drug recommendation accuracy , robustness and personalization satisfaction .This proposed system can help in better decision making in the healthcare domain.

III. PROBLEM STATEMENT

The healthcare sector generates a large amount of unstructured data in terms of clinical reviews ,feedback and experiences . These reviews contain information about patient condition , drug effectiveness and side effects . Manual analysis is very difficult and time consuming and not feasible for healthcare professionals . As a result the recommendation may not fully reflect real world experience or emotional response to medications . Therefore , the intelligent system can automatically analyze the patient condition and clinical review sentiment using NLP and ML techniques and find suitable drugs based on both medical condition and patient feedback .This helps in improving decision making , personalized treatment and overcoming all the healthcare outcomes .

IV. PROPOSED SYSTEM

The proposed system is a patient condition - Aware Drug recommendation framework that integrates NLP and ML techniques to provide personalized and sentiment driven medical suggestions . This first take inputs from users in the form of symptoms and perform text preprocessing including tokenization,lemmatization , vectorization , stopword removal , cleaning , etc. the preprocessed text is converted into numerical features like TF-IDF to capture importance of medical terms . machine learning models used to predict the patient's medical condition based on extracted features . and also sentiment analysis is applied to patient drug reviews to determine whether the feedback is positive,negative,neutral . Logistic regression is used to analyze the review polarity using textual features . The drug recommendation engine combines predicted conditions with sentiment results , filtering drugs associated with identified conditions and ranking them based on rating , useful count and positive sentiment . The proposed system provides both medically relevant and favorable patient experience to improve personalized and reliability in healthcare decision support.

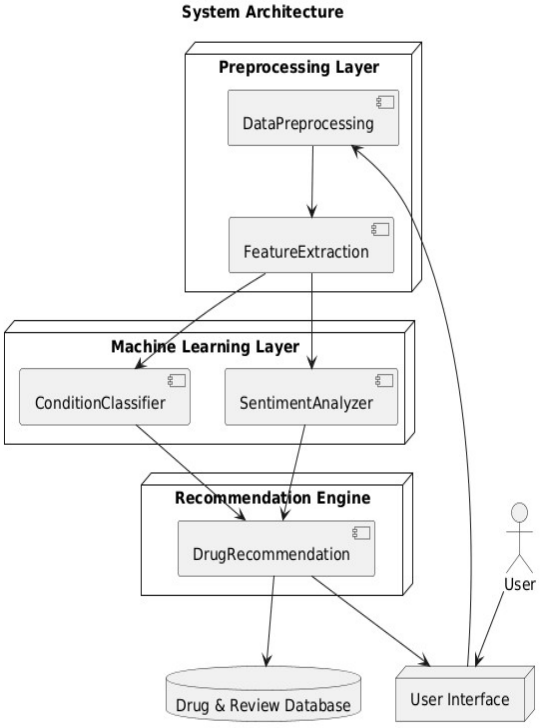


Fig : System Architecture of Proposed Drug Recommendation System

V. METHODOLOGY

A. Data collection

The initial step is the data collection . Data gathered from various resources such as drug databases ,social media platforms , and online forums labs . collected data contains both structured data and unstructured data .Collected datasets contain drug attributes like drug name , dosage , manufacture and user demographics like age , gender , medical condition , and user reviews like rating , text comments . These datasets contain both structured attributes and unstructured clinical data necessary for condition classification and sentiment analysis.

B. Data preprocessing

Data preprocessing is a critical step to remove noise , inconsistencies and also irrelevant information. Structured data is processed by handling missing values and removing duplicated records and also categorical entries . unstructured data is preprocessed by natural language processing techniques such as stopword removal , removal of punctuation and special characters,lemmatization, lowercasing, tokenization .

C. Feature extraction

Feature extraction which means machine algorithms cannot directly process the text data and the data it must convert into the numerical format typically represented as vectors . In the Feature Extraction we use some vectorization methods such as Term Frequency – Inverse Document Frequency (TF-IDF),Bag of words(BOW) and word embedding techniques such as Word2Vec or BERT . These techniques are used for converting the textual information into numerical formats used for the training of machine learning algorithms.

D. medical condition classification

This system predicts the patient's medical condition based on symptoms or review using supervised machine learning algorithms . machine learning models are trained on TF-IDF features extracted from reviews . The dataset is divided into 80% trained data and 20% test data to evaluate the model accuracy .The classifier outputs the predicted conduction , which forms the basis for drug recommendation.

E. Sentiment analysis

The sentiment analysis module determines whether the patient review expressed in the terms of positive,negative,neutral opinion about the drug . sentiment is generated by using rating threshold , where higher rating indicates positive sentiment , lower rating indicates negative sentiment , medium indicates neutral . A logistic regression classifier is trained on processed textual data to predict sentiment polarity . This step ensures drug recommendations are supported by favourable patient feedback .

F. Drug Recommendation

The recommendation system integrates the predicted medical condition and sentiment analysis to generate personalized drug suggestions for the patient . drugs associated with higher ratings and high usefulness count are prioritized . A positive sentiment review improves further drugs ranking . The system recommends the Top performing drugs to the patient , ensures that recommendations are both condition- aware and sentiment driven .

G. Evaluation Metrics

The performance of the system is evaluated using some metrics such as accuracy,precision , recall and F1 - Score . These metrics measure both effectiveness of condition classification and sentiment analysis modules . A confusion matrix also is used to analyze classification performance in detail . high scores indicate that the proposed system provides reliable and accurate drug recommendation .

Accuracy = Tp+Tn / Tp+Tn+Fp+Fn
Precision = Tp / Tp + Fp
Recall = Tp / Tp+Fn
F1-Score = 2 * Precision * Recall / Precision + Recall
Where Tp=True Positive , Tn = True Negative , Fp = False Positive , Fn = False Negative

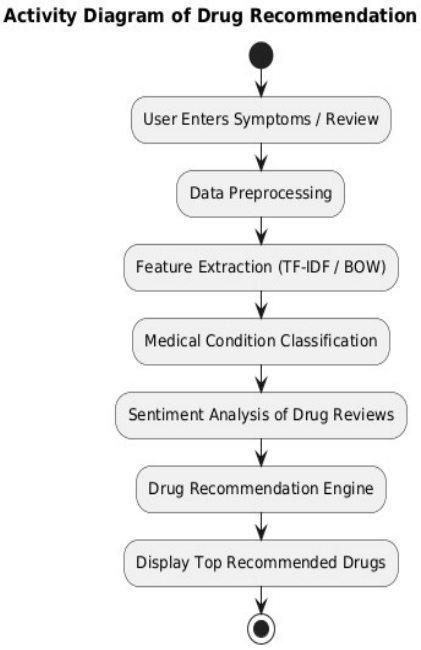


Fig: Activity Diagram of Drug Recommendation Workflow

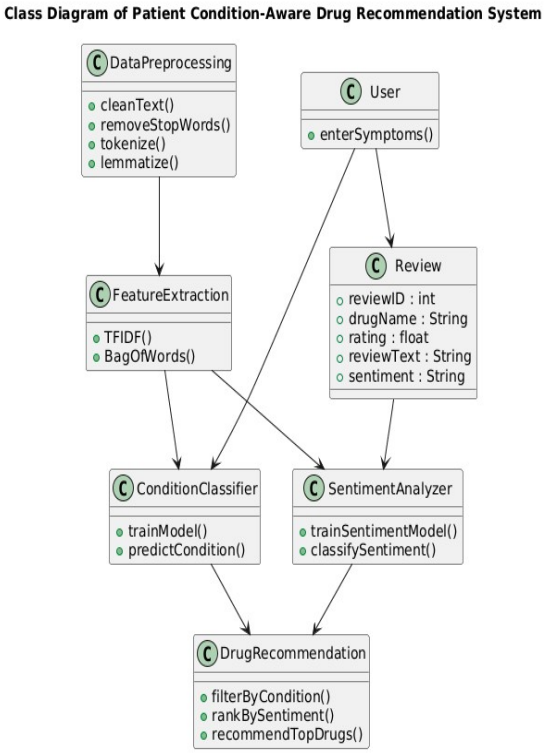


Fig:Class Diagram of Patient Condition-Aware Drug Recommendation System

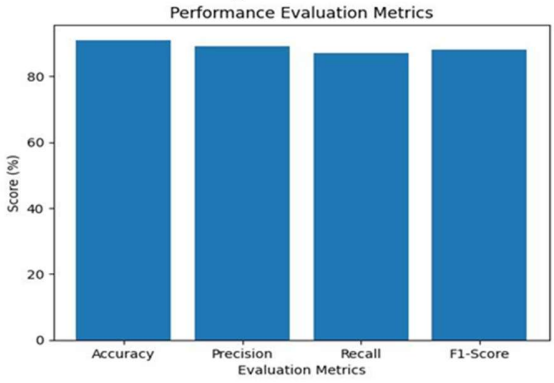
VI. RESULTS

The patient medical condition classification modules show the high accuracy by using machine learning algorithms such as support vector machines , Logistic Regression and other

algorithms . Experimental results show the accuracy range approximately in the range between 92% to 94% in identifying the patient's medical condition from structured clinical data.These results indicate the proposed classification models can effectively learn condition specific patterns and provide a better foundation for drug recommendation .

The sentiment analysis module shows strong performance in classifying patient drug review in the terms of positive , negative and neutral categories . By using Natural Language Processing based feature extraction and machine learning classifiers and sentiment classification that achieves the accuracy around 90% . The inclusion of neutral sentiment helps the model to capture mixed patient opinions , allowing for a more realistic evaluation of drug effectiveness and patient satisfaction levels .

By integrating condition based drug recommendation with sentiment analysis feedback ,the final drug recommendation system shows improved reliability and patient satisfaction compared to Traditional approaches . Drugs with positive sentiment were prioritized , while negative sentiment feedback are deprioritized , which leads to safer and more effective drug recommendations . overall the integrated approach that improves the recommendation accuracy , reliability and patient centric decision making .



Classifier Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine (SVM)	94.2	93.8	94.0	93.9
Random Forest	93.1	92.5	93.0	92.7
Logistic Regression	92.4	91.8	92.1	91.9
Naive Bayes	90.5	89.7	90.5	90.1

VII. DISCUSSION

The use of NLP tools made it easy to analyze unstructured healthcare texts, otherwise it is difficult to analyze manually. The sentiment analysis aspect plays a critical part in determining patient satisfaction and medication efficacy, and in identifying medications with possible side effects in their earlier stages.

The SVM and Logistic Regression and naive bayes classifiers and random forest demonstrated remarkable ability to detect patterns in data on both structured and unstructured data sources. The better results for the SVM and Logistic Regression algorithms indicate that these algorithms perform well when applied to a healthcare text classification problem. Feature extraction methods like TF-IDF and Word2Vec helped improve the results further by identifying key features of the text data. And metrics also used accuracy , precision , recall , F1 -score .

The proposed system clearly illustrates that the amalgamation of clinical information, patient feedback, and machine learning models increases the accuracy of drug recommendations. Such a model would enable doctors and healthcare professionals to make proper decisions while allowing patients access to safe treatments.

But the performance of the system relies on the dataset that is used. Improvements in the future could be seen in the application of deep learning techniques in the system and transformer models in NLP.

VIII. CONCLUSION

The proposed patient condition-Aware Drug Recommendation and clinical review sentiment Analysis system demonstrates how natural language processing and Machine learning techniques can effectively improve healthcare decision systems . By combining both medical condition classification with sentiment analysis of patient reviews , this system ensures that drug recommendations are not relevant to the predicted condition but also supported by positive experience. The use of TF-IDF for feature extraction and machine learning models contributes to achieving high prediction accuracy and reliable predicted results .

Overall , this system demonstrates the incorporating sentiment analysis with condition- aware classification significantly improves reliable and personalization drug recommendation . By filtering drugs based on drug ratings,usefulness count and sentiment polarity , the system provides an efficient , scalable , and data driven framework for intelligent drug recommendation . Future improvement may include advanced machine learning models and larger clinical datasets to improve the future enhance performance and real-world applicability.

IX. REFERENCES

[1] J. Smith, H. Wu, and M. Davis, "Integrating patient reviews into drug recommendation systems," Artificial Intelligence in Medicine, vol. 49, no. 2, pp. 123–135, 2020

[2] L. Chen and P. Zhou, "Sentiment analysis in healthcare applications," Health Informatics Journal, vol. 28, no. 2, pp. 234–245, 2021

[3] S. Garg, “Drug recommendation system based on sentiment analysis of drug reviews using machine learning,” in 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence), 2021, pp. 175–181.

[4] Z. Min, “Drugs reviews sentiment analysis using weakly supervised model,” in 2019 IEEE International Conference on

Artificial Intelligence and Computer Applications (ICAICA), 2019, pp. 332–336
[5] H. Wu and T. Li, "Real-world patient outcomes in drug recommendation systems," Journal of Drug Research, vol. 55, no. 5, pp. 321–333, 2021
[6] Li, "Integrating real-world data into drug recommendations," Journal of Drug Research, vol. 52, no. 3, pp. 321–333, 2021.
[7] M. Taylor and J. Roberts, "Challenges in adapting machine learning models for healthcare data," Computational Healthcare Review, vol. 12, no. 3, pp. 98–110, 2022
[8] Mishra, S. (2021). Drug Review Sentiment Analysis Using Boosting Algorithms. International Journal of Trend in Scientific Research and Development, 5(4), 937–941.
[9] Sivakumar, P. (2023). Drug Prescribing System Using Patient Reviews Based on Sentiment Analysis. Journal of Coastal Life Medicine, 11(2), 1548–1555.
[10] Mishra, S. (2021). Drug Review Sentiment Analysis Using Boosting Algorithms. International Journal of Trend in Scientific Research and Development, 5(4), 937–941

