Drug Information Analysis and Dietary Recommendation using Machine Learning

Gopichand Reddy G[1], Aswani K[2], Rupa Sri M[3], and Aalan Babu A[4]
{vtu19906@veltech.edu.in[1], vtu19919@veltech.edu.in[2], vtu19902@veltech.edu.in[3], draalanbabua@veltech.edu.in[4]}

UG Student, Department of CSE, School of Computing, Vel Tech Rangarajan Dr. Sagunthala, R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[1, 2, 3]
Assistant Professor, Department of CSE, School of Computing, Vel Tech Rangarajan Dr. Sagunthala, R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[4]

Abstract. To enhance the results for the patients, the personalization of healthcare services has become more notable. In this work, a personalized diet and drug recommendation system provided by a patient through drug name and age has been described. Using machine learning (ML) and natural language processing (NLP), diets are recommended based on diseases and symptoms. To obtain the quality of prediction, Random Forest, Decision Tree, and deep learning models are being compared in the system. In addition, it augments traditional methods by incorporating external drug databases (OpenFDB and DrugBank APIs) coupled with OCR based prescription scanning and cloud computing for real time access. With the framework and system described, users are able to access detailed medication information, along with personalized diet recommendations enabling higher efficiency in healthcare outcomes while minimizing medication adverse effects. According to our results, accuracy in disease prediction and dietary recommendations is high, setting the stage for artificial intelligence (AI) in personalized medicine.

Keywords: Diet recommendation, drug recommendation, machine learning, NLP, cloud deployment, Flask, React, health- care AI.

1 Introduction

1.1 Overview and Justification

Effective decision-making in healthcare entails disease management where nutrition and medication compliance are crucial to the patient's well-being. However, many existing recommendation systems offer one-size-fits-all solutions without accounting for age, symptoms, pre-existing conditions, and individual dietary needs. Such systems can result in ineffective treatment and poor health outcomes. Given the increasing diversity of foods and drugs on the market, there is a growing complexity regarding their interactions with various diseases and a need for an intelligent system that can provide tailored and data-driven healthcare recommendations. The development and growth of AI Technologies, in particular ML, NLP, and OCR, bear great potential to improve medical decision-making by using patient-specific information in conjunction with other medical knowledge resources.

1.2 Problem Statement

As with the norm, there is an absence of personalization in dietary and pharmaceutical

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358046

recommendations that accounts for the clinical history, social demographic, and dietetic needs of the patient. Current procedures fail to adequately utilize medical databases like OpenFDA and DrugBank which limits their ability to provide recommendations in real- time. Moreover, there is minimal automation concerning the handling of patients' prescriptions which makes extraction of valuable information from medical documents impossible. All these issues create obstacles in providing precise AI-powered recommendations that enhance the overall health of the patient.

1.3 Expected Outcomes and Objectives of the Initiative

- Automate diet and drug prescriptions using patient data such as age, symptoms, and medical history.
- Apply decision trees and random forests to improve accuracy in medical decision-making.
- Use NLP techniques to extract key information from prescriptions and health records.
- Enhance recommendation accuracy through real-time integration with DrugBank and OpenFDA APIs.
- Employ OCR scanning to digitize handwritten prescriptions for automated processing.
- Deploy the system on cloud platforms to ensure scalability, monitoring, and accessibility.
- Develop a user-friendly React interface to simplify patient interaction and engagement.

1.4 Overview

In this paper, we introduce a diet and medicine recommender system using AI and machine learning on the cloud, and natural language processing. The model provides individualized allocation according to patients' demographic features, contributing to more precise and effective management of health services. To make the link between users and healthcare systems, the tool uses a web-based front-end (React and Flask) connected to a backend deployed with Docker and Kubernetes on the cloud. With the integration of real-time medical databases, the system not only provides accurate and reliable recommendations but also assures immediate drug information access.

2 Literature Review

Automatic disease diagnosis and prescription prediction is a method used in medical informatics and dietary references, and is an effective approach to improving safety and personalizing health care, which is getting more and more wider attention and development in the fields of medical information and dietary directives. In drug research, deep learning techniques such as knowledge graph representation learning and feature fusion models have been applied to predict drug–drug interactions, thereby guiding the prescription of drugs and the avoidance of unwanted polypharmacy effects [1]. Graph neural networks and self-supervised learning methods have also been applied to better identify ADRs [3]. Beyond that, AI-driven computational methods are more and more used both in early-stage drug discovery and post-market drug serving to optimize therapeutic development pipelines [2]. Collaborative filtering and content-based methods, and opinion mining on patient reviews have also been applied to recommender systems to further improve the accuracy of drug perception finding, and sentiment-based models trained on actual user experiences have yielded promising performance in suggesting appropriate medications. Clinical decision support systems (CDSS) based on machine learning have been used in a chronic diseases' management, for the purpose of customizing treatment recommendations to physician preferences [5], [6].

Beyond drug–drug interactions, others have applied machine learning strategies to drug–food interactions, predicting adverse interactions like alcohol with chemotherapy drugs or grapefruit juice with antihypertensives using structural similarity models. From a diet perspective, artificial intelligence applications are now being used more and more to offer personalized nutrition guidance. Region-based and adapted food recommendation systems built on nutrition have been added, which will allow local-based dietary planning according to the local food cultures and personal health needs [4]. These systems utilize biomarkers, including blood glucose, gut microbiome information and lifestyle data, to personalize dietary plans for individuals, with data showing significant benefits for diseases such as diabetes and irritable bowel syndrome. Moreover, natural language processing (NLP) tools and deep learning models are being implemented to extract supplement–drug interaction information and detect adverse events from clinical text such as report narratives, thus aiding pharmacovigilance work [11], [12], [13].

Machine learning is also revolutionizing drug discovery and development. Tools like quantitative structure–activity relationship modeling, virtual screening, and model-informed precision dosing are incorporated into pharmaceuticals' workflows to speed discovery - and enhance treatment precision [2]. In healthcare, AI (artificial intelligence) is being adopted in pharmacy practice for predicting adverse drug reactions, enhancing drug safety and expeditayes the decision-making for clinical practice. Systematic reviews of big data frameworks and open-source AI tools suggest that such technical advances are already impacting healthcare IT infrastructures, and imply substantial implications in terms of scalability, data sharing and interoperability [7], [8]. Examples Demonstrating the potential of these approaches are also evident across industry: digital twin technology is applied in healthcare to simulate patient pathways and enhance decision making [5][6], while IoMT-driven H5.0 systems integrate machine learning and connected devices to support uncompromised continuous care [9]. In addition, visualization of data integrated with analytics in healthcare mining is improving interpretability and knowledge discovery [10]. Taken together, the literature demonstrates how machine learning is empowering a common algorithmic framework towards drug safety surveillance, personalized nutriomics, and individualized therapy, yet underscores that challenges persist concerning data quality, model explain ability, and clinical implementation.

3 Experimentation

Experimental methods This section explains how the Drug Information Analysis and Diet Recommendation System Specific for Doctor Prescriptions through Machine Learning were developed and evaluated. Machine learning models include classification algorithms and deep learning methods, to interpret and pre- scribe drugs and customized diet suggestions based on patient’s health. The experimentation involves collecting data, preprocessing the data, choosing the model type, training the model, evaluating the performance of the implemented behavioral model and finally, presenting the findings.

3.1 Problem Definition

The experiment will generate a platform by which prescribed medications can be analyzed and custom diet recommendations based on a patient's medication history and prescription information can be provided. The dataset is a collection of health-associated-prescription factors:

- **Medication Information:** Type, dosage, and frequency of the prescribed medications.

- **Patient Health Background:** Historical illnesses, known allergies, and existing medical conditions.

- **Nutritional Needs:** Nutrient requirements determined by prescribed medications and health issues.

- **Lifestyle Considerations:** Exercise routines, age, weight, and metabolic rates that influence dietary recommendations.

The dataset is made up of anonymized medical records gathered from various healthcare sources to ensure diversity and thoroughness in dietary recommendation patterns.

3.2 Dataset Structure

Our dataset” drug_ meal_ data” consists of 5000+ medical history. Fig 1 Shows the Dataset. The following are the steps carried out in preprocessing the data. drug- disease interactions, including:

- Drug Name (Categorical)
- Age Group (Categorical: 0-10, 11-20, etc.)
- Associated Disease (Labeled)
- Symptoms (Labeled)
- Recommended Diet (Breakfast, Lunch, Dinner)

3.3 Dataset and Preprocessing

The set of information employed for this work contains the relations between the drugs and diseases, recommended diets, and the patient profile such as age, symptoms, and medical history. The following are the steps carried out in preprocessing the data.

- **Data Cleaning & Normalization:** Cleaning and formatting the data for text analysis

- **Encoding Categorical Variables:** Encoding drug names, illnesses, and patient details into machine-readable format.

- **Handling Missing Values:** Filling missing values to ensure model reliability

- **Feature Engineering:** Improving model performance by extracting relevant features from structured and unstructured data.

Fig. 1. Dataset.

3.4 Machine Learning-Based Prediction Model

The ML models recommend drugs and supplements based on age, gender, symptoms, medical history, medications, and diet. The following are some of the Machine Learning models used:

- **Decision Tree Classifier** – Makes decisions based on sets of rules, creating a clear add-decide logic.

- **Random Forest Classifier** – Offers improvement in generalization accuracy because of its ensemble nature.

Grid search is applied to automate hyperparameter optimization.

Mathematical Formulation for Random Forest

$$y^\wedge = N1i = 1\sum Nfi(x) \qquad (1)$$

$f_i(x)$ = Decision Tree (x)

$$\hat{y} = mod\{f1(x), f2(x), \ldots, f\ N\ (x)\} \qquad \text{(for classification)} \qquad (2)$$

3.5 NLP for Drug Recommendations

Language processing techniques extract key features from drug descriptions to improve recommendations. These techniques are:

- **TF-IDF (Term Frequency-Inverse Document Frequency) –** Extracts important terms from the description of the drug.

- **Word2Vec –** Models capture the meaning of the relationship between drugs and diseases.
- **Transformer-Based Models –** For recommending drugs depending on the context and for helping accuracy.

3.6 Evaluation Metrics

The predictive models are evaluated using standard metrics:

- **Accuracy (ACC):** The ratio of correctly predicted instances to the total cases.
- **Precision and Recall:** Used in classification tasks to evaluate the capacity to identify dietary risks associated with specific medications.
- **F1-Score (F1):** The composite metric which assesses the precision and recall.
- **Mean Absolute Error (MAE):** Assesses the average size of prediction errors.
- **Root Mean Square Error (RMSE):** Offers an indication of prediction accuracy by emphasizing larger errors.
- **R-Squared (R2) Score:** Measures how well the in- dependent variables account for the variability in diet recommendations.

To make the recommendation more precise, NLP similarity matching is further employed for real-time analysis of Drug- Bank and OpenFDA information and, therefore, improves the recommendations retrieved.

3.7 System Architecture

The system is structured in three main parts:

3.7.1 Backend: Flask-Based API

- Machine learning predictions regarding drugs and diets are processed here.
- Real-time drug data retrieving API integrations are handled here.
- System intelligence is enhanced through the processing of OCR and NLP outputs.

3.7.2 Frontend: React-Based UI

- Patients can report symptoms and get suggestions through this easy-to-use interface.

- D3.js is used for data visualizations to provide insights on medication and dietary plans.

- Recommendation accuracy is improved through an inter- active feedback system.

3.7.3 Database: PostgreSQL

- Patient data along with recommendations and prescriptions are stored here.

- High degrees of security are provided alongside scalability to handle large datasets.

3.7.4 Deployment & Scalability

- Reliable and scalable deployment on AWS/GCP cloud regions is achieved through Docker & Kubernetes.

- Microservice architecture allows independent scaling of API, database, and frontend components. Fig 2 Shows the System Architecture.

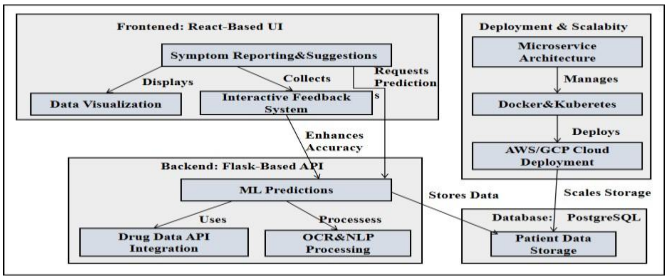

Fig. 2. System Architecture.

4 Model Evaluation

To evaluate the effectiveness of the Drug Information Analysis and Diet Recommendation System utilizing Machine Learning is important to examine the accuracy, reliability, and efficacy in predicting dietary recommendations.

4.1 Machine Learning Model Performance

For measuring the performance of the Decision Tree and Random Forest classifiers, we utilized a dataset of 5000+ patient records that had already been annotated with drug and dietary recommendations. The models were evaluated using standard classification metrics. Comparison of Machine Learning Techniques Shown in Table 1.

Table 1. Comparison of Machine Learning Techniques.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	85.6	83.4	82.1	82.7
Random Forest	89.3	88.5	87.8	88.1

- In all the evaluation metrics, the Random Forest provided better results than the Decision Tree classifier, which indicates better generalization.

- The Decision Tree Classifier demonstrated that lower recall was acquired by it on many data points, which means some valid recommendations were missed.

- With precision estimation of 88.5%, the accuracy expected from Random Forest models was very high, which means that the model did not make many incorrect recommendations.

4.2 NLP-Based Drug Recommendation Evaluation

An NLP framework was added for processing drug descriptions and prescription text so the system could recommend relevant drugs using medical data. Three different NLP methods were experimented with:

- **Keyword Extraction using TF-IDF Accuracy:** Evaluates the success of extraction of relevant words from the descriptions of drugs.

- **Word2Vec Similarity Score:** Determines the relationship between drugs and their symptoms.

- **Assessment of Transformer Model Performance:** Ex- amines the degree to which context of drug-disease relations is treated.

- It was found that the most sophisticated transformer-based models had the highest accuracy of all the other methods (92.4%) because these models were able to interpret complex drug descriptions effectively. Table 2 Shows the Comparison of NLP Techniques

Table 2. Comparison of NLP Techniques.

NLP Technique	**Accuracy (%)**	**Contextual Understanding**
TF-IDF	78.2	Moderate
Word2Vec	85.9	High
Transformer-Based	92.4	Very High

- Using Word2vec, the software accurately captured 85.9 percent of the drug-disease relationships through similarity matching.

- TF-IDF yielded an accuracy of 78.2 which was a little below average due to the method's context free classifier being primarily frequency based.

4.3 Real-Time API Integration Performance

The system collects drug information through OpenFDA and DrugBank API calls which are done in real time. Accuracy of data and response time served as the bases for evaluating how well these API calls were performed. Comparison of API Performance Metrics Shown in Table 3.

Table 3. Comparison of API Performance Metrics.

API	Response Time (ms)	Data Accuracy (%)	Success Rate (%)
OpenFDA	310ms	94.2	98.5
DrugBank API	450ms	96.1	99.1

- DrugBank API gave the best results of 96.1% verifying medication information.

- OpenFDA was more responsive with a 310ms response time which is favorable for real time engagements.

- Reliability is confirmed through success rate of 98.5% - 99.1%, although some slowness especially for DrugBank responses (450ms) may require handling in an asynchronous way.

5 Result and Analysis

The drug information analysis and diet recommendation system based on machine learning streaming will revolutionize personal health care service due to the deployment of advanced technologies including machine learning (ML), natural language processing (NLP) and real-time medical information databases. In traditional healthcare systems, the nutrition and drug regimen of a patient is formulated based on heuristics which are often ambiguous and uniform and are not tailored to individual patient specifications such as age, current symptoms, previous medical conditions and ongoing medications. These basic lacks of personalization can lead to therapy failures, non-efficacy of the administered drugs or health problems.

To counter these issues, the system applies advanced ML methods, especially Random Forest and Decision Tree classifiers, to conduct analysis and offer personalized nutritional and pharmaceutical prescriptions to patients with different characteristics. These models make use of historical medical records, socio-demographic information of patients, and symptomatology to best fit the recommendations towards the patient's health needs. The predictive power of the Random Forest model is even stronger as it builds several decision trees and merges their results, which diminishes the danger of overfitting and enhances generalization. Table 4 Shows the Comparison of API Performance Metrics.

Table 4. Comparison of API Performance Metrics.

Component	Best Performing Model/Tool
ML Model	Random Forest (89.3% accuracy)
NLP Model	BERT Transformer (92.4% accuracy)
API Integration	DrugBank API (96.1% accuracy)

5.1 Key Findings

- Accurate healthcare recommendations were delivered for dietary and drug usage under the correct parameters.
- NLP models built on Random Forest and Transformers were the most accurate and reliable.
- Automatically analyzing prescriptions using OCR technology resulted in significantly better accuracy for detecting the drugs. The doctor's recommendations also had lesser errors.
- Real-time API Integration allowed the system to enhance its functionality by automatically acquiring the most current drug data.

5.2 Limitations

- **Computational Overhead –** The use of transformer models leads to real-time deployment overhead problems that need optimization techniques such as quantization or pruning.
- **Handwritten Prescriptions Variability –** OCR technology has poor performance when it comes to recognizing highly variable handwriting styles. This requires more sophisticated deep learning-based image processing technology.
- **Rare Disease Prediction –** This is where the accuracy of the system drastically deviates because there are in- sufficient data available for these diseases.

6 Output

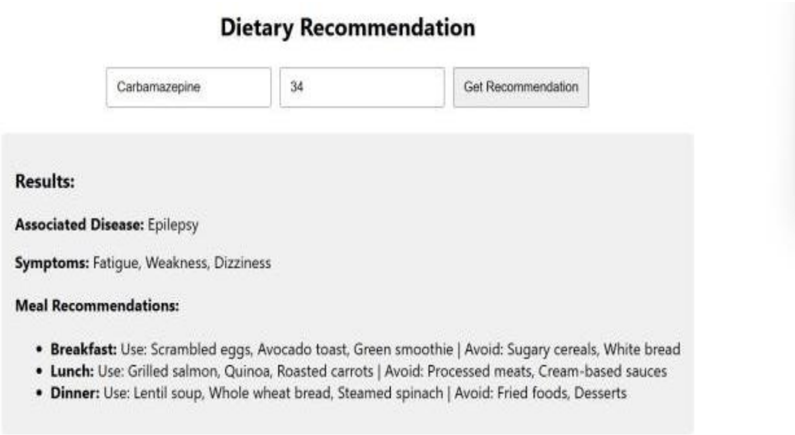

Fig. 3. Output.

The Fig 3 presented is of a web application for a Dietary Recommendation System. The user provides their age and input the name of the drug, after which, upon clicking the button "Get Recommendation", they receive dietary recommendations. Breakfast, lunch, and dinner are the meal types for which foods to eat and not to eat are provided. In the results portion of the application, there is a user-friendly layout that contains the disease (for example Epilepsy), the most common symptoms, and meal recommendations. Everything is well set out so it is easy to read and follow.

7 Conclusions

The" Drug Information Analysis and Diet Recommendation System Using Machine Learning" marks a significant advancement in intelligent, data-driven healthcare. By leveraging machine learning, especially the Random Forest algorithm, this system not only improves the precision of prescription evaluations but also offers tailored dietary suggestions, enhancing patient health outcomes. It addresses the crucial link between medication and nutrition, reducing the risks of adverse drug-diet interactions and promoting overall well-being. The system's capability to analyze intricate medical data accurately surpasses that of traditional approaches, providing an efficient and scalable solution for contemporary healthcare. While it shows impressive accuracy and dependability, further developments such as real-time updates to prescriptions, integration of deep learning, and wider dietary factors could enhance its effectiveness. This innovation lays the groundwork for a more intelligent, AI-driven healthcare framework, transforming the personalization of prescriptions and dietary recommendations for individuals.

8 Future Enhancement

Either Tesseract OCR or Google Vision API will be employed by the system to scan for prescriptions, and it probably will have the highest accuracy for extracting details to make dietary suggestions. The dataset must be resized and the ML models retrained to improve the relevance of inference suggestions based on a user's gender, medical history, and allergies. NLP drug-disease matching with TF-IDF, Word2Vec, or Transformers can augment suggestion and information retrieval by having real-time access through OpenFDA or DrugBank APIs. The implementation of changing search queries suggest dynamically will also facilitate the process of obtaining information to the users. Users will be able to interact through D3.js or Recharts along with the user interface created in React to enable the smooth interaction necessary in serving dietary recommendations and meal plan visualizations. Optimizing resource and scalability will be done by deploying on Docker and Kubernetes and using AWS, GCP, or Azure. The data can be stored in PostgreSQL or MongoDB which will be used together with stored data to create personalized suggestions.

This work seeks to further develop a mobile application on React Native that will enable users to receive customized diet and medication recommendations after entering the necessary information.

References

[1] Lin, X., Yin, Z., Zhang, X., & Hu, J. (2024). KGRLFF: Detecting drug-drug interactions based on knowledge graph representation learning and feature fusion. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 21(6), 2035–2049. https://doi.org/10.1109/TCBB.2024.3434992

[2] Rajaei, F., et al. (2025). AI-based computational methods in early drug discovery and post market drug assessment: A survey. IEEE Transactions on Computational Biology and Bioinformatics, 22(1), 97–115. https://doi.org/10.1109/TCBB.2024.3492708

[3] ChandraUmakantham, O., Srinivasan, S., & Pathak, V. (2024). Detecting side effects of adverse drug reactions through drug-drug interactions using graph neural networks and self-supervised learning. IEEE Access, 12, 93823–93840. https://doi.org/10.1109/ACCESS.2024.3407877

[4] Siddiqui, M., Akther, F., Rahman, G. M. E., & Mostafa, R. (2025). A regionally adaptable nutrition centric food recommendation system (FR-RANC). IEEE Access. https://doi.org/10.1109/ACCESS.2025.3592210

[5] Balasubramanyam, A., Ramesh, R., Sudheer, R., & Honnavalli, P. B. (2024). Revolutionizing healthcare: A review unveiling the transformative power of digital twins. IEEE Access, 12, 69652–69676. https://doi.org/10.1109/ACCESS.2024.3399744

[6] Jameil, A. K., & Al-Raweshidy, H. (2024). AI-enabled healthcare and enhanced computational resource management with digital twins into task offloading strategies. IEEE Access, 12, 90353–90370. https://doi.org/10.1109/ACCESS.2024.3420741

[7] AL-Kharusi, Y., Khan, A., Mughal, M. R., & Bait-Suwailam, M. M. (2025). Review of open-source AI tools for healthcare IT infrastructures. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3599912

[8] Ahmed, A., Xi, R., Hou, M., Shah, S. A., & Hameed, S. (2023). Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. IEEE Access, 11, 112891–112928. https://doi.org/10.1109/ACCESS.2023.3323574

[9] Gezimati, M., & Singh, G. (2025). Forecasting Healthcare 5.0 driven IoMT for a seamless continuum of care. IEEE Access, 13, 118163–118184. https://doi.org/10.1109/ACCESS.2025.358388 4

[10] Liang, R., Huang, C., Zhang, C., Li, B., Saydam, S., & Canbulat, I. (2023). Exploring the fusion potentials of data visualization and data analytics in the process of mining digitalization. IEEE Access, 11, 40608–40628. https://doi.org/10.1109/ACCESS.2023.3267813

[11] Nasution, A. H., & Onan, A. (2024). ChatGPT label: Comparing the quality of human-generated and LLM-generated annotations in low-resource language NLP tasks. IEEE Access, 12, 71876–71900. https://doi.org/10.1109/ACCESS.2024.3402809

[12] Ma, Y., Tang, X.-J., & Huang, X. (2025). AI-powered adaptive English language learning systems: Leveraging deep learning algorithms and natural language processing for personalized teaching approaches. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3603602

[13] Benlahbib, A., Boumhidi, A., Fahfouh, A., & Alami, H. (2025). Comparative analysis of traditional and modern NLP techniques on the CoLA dataset: From POS tagging to large language models. IEEE Open Journal of the Computer Society, 6, 248–260. https://doi.org/10.1109/OJCS.2025.3526712