



Personalized Medicine Recommendation System Using KNN for Cancer Treatment

Parmar Bhaumik Mehulbhai¹, Parmar Manav Sandipbhai², Tank Jayraj Dilipbhai³, Baraiya Nirav Ashvinbhai⁴, Baraiya Bhargav Himmatbhai⁵, Prof. Patel Janki Tejas⁶

¹ Computer Engineering, Sal College of Engineering

² Computer Engineering, Sal College of Engineering

³ Computer Engineering, Sal College of Engineering

⁴ Computer Engineering, Sal College of Engineering

⁵ Computer Engineering, Sal College of Engineering

⁶ Assistant Professor Computer Engineering, Sal College of Engineering

ABSTRACT :

Cancer continues to be one of the most complex diseases known to humanity, affecting millions of people every year. While traditional treatments such as chemotherapy, radiation, and surgery have saved many lives, they often fail to consider the genetic and biological uniqueness of each patient. Because of this, two people with the same type of cancer may respond very differently to the same medication. To solve this challenge, modern healthcare has started shifting toward personalized medicine — a treatment approach that tailors therapies based on individual characteristics such as genetics, age, medical history, and lifestyle. This research paper introduces a Personalized Medicine Recommendation System using the K-Nearest Neighbors (KNN) algorithm. The system aims to recommend the most suitable cancer treatment for a patient based on genetic mutation data and clinical profiles. By analyzing the similarities between a new patient and historical patient data, the KNN algorithm predicts which medication is most likely to be effective. The project involves several stages: collecting and preprocessing patient data, implementing the KNN model, validating its accuracy, and ensuring ethical and secure data handling. Our results indicate that KNN can produce accurate and explainable predictions, making it valuable for clinicians. Although challenges such as data imbalance and limited datasets exist, this system shows the potential to improve patient outcomes, reduce side effects, and support oncologists in making data-driven treatment decisions. Ultimately, the goal of this study is not to replace doctors but to assist them in providing customized, efficient, and effective cancer treatment plans that suit the genetic and clinical uniqueness of each patient.

Keywords: Personalized medicine, Cancer treatment, Machine learning, K-Nearest Neighbors (KNN), Medical data analysis, Predictive modeling, Precision oncology, Data-driven healthcare, Feature selection, Treatment optimization

Introduction

Cancer is a disease that affects people differently because every person's body, lifestyle, and genetic structure are unique. Even patients diagnosed with the same type of cancer often experience different responses to the same medication. This happens because cancer is not just one disease but a group of related diseases caused by changes in the DNA that control cell growth. These mutations can vary widely between individuals, making it difficult to design one treatment that works for all. Traditionally, oncologists have used standardized treatment protocols based on tumor type and stage. However, such generalized approaches may not yield the best results for everyone. In some cases, a drug that cures one patient may cause severe side effects or fail to help another. This realization has led to the rise of personalized medicine, an approach that customizes healthcare based on each patient's unique biological data.

Machine learning, a field of artificial intelligence (AI), plays a major role in making personalized medicine possible. It allows computers to analyze vast and complex datasets containing patient information, genetic codes, and treatment outcomes. One simple yet powerful machine learning algorithm suitable for this purpose is K-Nearest Neighbors (KNN). The KNN algorithm identifies patients with similar genetic and clinical features and recommends the treatment that proved successful for those similar patients.

In this research, we apply KNN to analyze genetic mutation data and clinical features of cancer patients. The system aims to classify patients and suggest treatments with a higher chance of success. The motivation behind this study is to improve the effectiveness of cancer therapy, reduce unnecessary side effects, and make treatment decisions more personalized and accurate.

The paper also discusses the methods used to collect, clean, and preprocess patient data, the design of the KNN model, its implementation using Python, and validation using metrics such as accuracy and precision. By combining medical knowledge and AI techniques, this study demonstrates how computational tools can support medical professionals in creating smarter and more personalized treatment plans for cancer care.

Understanding Personalized Medicine

Personalized medicine, also known as precision medicine, is an evolving approach in healthcare that focuses on tailoring treatment to the individual characteristics of each patient. Instead of applying a “one-size-fits-all” approach, it considers genetic information, environmental factors, and lifestyle patterns to determine the most effective therapy. In cancer treatment, this means that the medications and dosages given to a patient depend on how their cancer cells behave at the molecular level. For example, two patients might have the same type of lung cancer, but one may have a genetic mutation that makes the tumor sensitive to a certain drug, while the other does not. Personalized medicine uses this knowledge to prescribe the drug that best matches the genetic profile of the tumor.

The major advantage of personalized medicine is that it improves treatment accuracy while minimizing harmful side effects. By analyzing DNA sequences and identifying mutations in specific genes, doctors can predict how a tumor will respond to a particular therapy. This reduces trial-and-error prescribing and ensures patients receive the most suitable treatment from the start. However, processing genetic data manually is complex and time-consuming. The human genome contains billions of data points, and no doctor can analyze this information alone. This is where machine learning models like KNN become essential. They can process large genetic datasets, find patterns, and make predictions faster than humans, all while continuously improving as more data is added.

Personalized medicine represents the future of healthcare — where treatments are no longer generalized but are made uniquely for each patient based on science and data.

The Role of Machine Learning

The K-Nearest Neighbors (KNN) algorithm is a simple yet effective method used in both classification and regression problems. In personalized medicine, it helps predict the best treatment for a new patient by comparing their data to previously treated patients.

KNN works by identifying the “k” closest data points (neighbors) to the patient’s genetic and clinical data in a multidimensional space. Each data point represents a past patient and contains information such as gene mutations, cancer type, and treatment outcomes. The algorithm calculates the distance between patients — usually using Euclidean distance — and then identifies the nearest neighbors. The treatment most common among those neighbors is recommended for the new patient. The success of KNN depends on selecting the right value for “k.” If “k” is too small, the model might overfit, meaning it will be too sensitive to minor variations in the data. If “k” is too large, it may underfit, ignoring useful local details. Typically, cross-validation is used to determine the best value of “k.”

KNN is non-parametric, meaning it does not assume any specific structure or distribution of the data. This flexibility makes it ideal for complex and high-dimensional medical data. Furthermore, the algorithm is easy to interpret — doctors can visualize which patients were considered neighbors and how the decision was made. This transparency is a major advantage in medical applications where explainability is crucial.

KNN Algorithm Basics

The proposed system consists of three core components:

- Database Layer – Stores patient data such as genetic mutations, demographics, and clinical outcomes.
- Machine Learning Layer – Implements the KNN algorithm to find similar patients and predict the most suitable medication.
- User Interface Layer – Allows doctors or researchers to input new patient information and receive treatment recommendations.

Data Collection and Preprocessing

The system uses genetic data obtained from next-generation sequencing (NGS) and clinical data from hospital databases. The dataset includes patient age, gender, cancer stage, type, genetic mutations, and previous treatment responses.

Preprocessing involves cleaning the dataset by removing errors, handling missing values through imputation, and normalizing numerical data. Feature selection ensures that only relevant genes and clinical features are used to train the model. The dataset is then divided into training and testing sets to validate the model’s performance.

Model Implementation

The model is built in Python using libraries such as pandas, scikit-learn, and Flask. The user interface is web-based and enables clinicians to input patient details and view the recommended treatments along with accuracy scores. Data security and privacy are maintained by encryption and role-based access control. This design ensures that the system can handle sensitive medical information safely while remaining user-friendly for doctors.

Conclusion

This research successfully developed and evaluated a Personalized Medicine Recommendation System using the K-Nearest Neighbors algorithm for cancer treatment. The study demonstrates how machine learning can bridge the gap between large-scale genetic data and clinical decision-making. By

analyzing patient similarities, the system provides personalized drug recommendations that can improve treatment success rates and patient satisfaction. The KNN model's strength lies in its simplicity, transparency, and adaptability. It allows healthcare providers to understand and trust the recommendations generated. Although there are limitations related to dataset quality, computation time, and data imbalance, these can be addressed through better data collection, feature optimization, and hybrid algorithms. As the future of medicine becomes more data-driven, systems like this can play a crucial role in making healthcare more accurate, efficient, and personalized. The combination of artificial intelligence and medical expertise has the power to transform cancer treatment and improve countless lives.

REFERENCES

1. Gu, D. et al. (2021). A Personalized Medical Decision Support System Based on CBR and KNN. PMC.
2. Siddalingappa, R. et al. (2023). KNN Algorithm to Predict Survival Time of Oral Cancer Patients. PMC.
3. Hassan, B.M. et al. (2025). Personalized Medical Recommendation System with Machine Learning.
4. Xie, X. et al. (2024). Personalized Anti-Tumor Drug Efficacy Prediction Based on Clinical Data.
5. Shahzad, M. et al. (2024). Classification of Clinically Actionable Genetic Mutations in Cancer. PMC.
6. Patel, S. et al. (2025). Integrating Machine Learning to Customize Chemotherapy Regimens.
7. UNLV Graduate College Manual. (2021).