FishTaxa - Journal of Fish Taxonomy
Vol 36 Issue 1s, ISSN: 2458-942X

Pharma Guide: Medicine Suggestions using ML

Dr. Rakesh Kumar Khare[1], Astha Pandey[2], Vanshika Thakur[3], Vaishnavi Kulkarni[4]
[1-4] Computer Science and Engineering, SSIPMT, Raipur,CG, India
vaishnavikulkarni@ssipmt.com

Abstract
The contemporary world is so fast that healthcare professionals are not always readily available and, in this case, the application of machine learning as a method of rapid diagnosis and prescribing treatment can be of great importance in resolving this issue. Covered in the project is a Medicine Suggestions which is a machine learning algorithm that can predict diseases given the symptoms the user has inputted and can offer the correct recommendations of what medication, home remedies and other disease management practices should be used. With the help of the established classifiers, including Support Vector Machine (SVC), Random Forest, and K-Nearest Neighbors (KNN), the system will be able to identify the disease accurately and offer a holistic treatment plan, including medications, diet, exercises, and precautions. It is not only a smart application that will help in the diagnosis but will also offer a comprehensive method in managing health by proposing preventive care, drugs, and household remedies thereby maximizing the overall healthcare experience of the user. Of all the named SVC, KNN, Gradient Boosting, multinomial NB, and Random Forest classifiers were used. The SVC classifier is the most accurate.

Keywords: Medicine, Machine learning, MultinomialNB, Python, SVC.

Citation: Dr. Rakesh Kumar Khare, Astha Pandey, Vanshika Thakur, Vaishnavi Kulkarni. 2025. Pharma Guide: Medicine Suggestions using ML. FishTaxa 36(1s): 454-459

Introduction
The sphere of healthcare is quickly changing to a digital form, and the point at which this process takes place is the creation of different data sources (electronic health records (EHRs), medical imaging, wearable technologies, and so on). Such information flow implies the shift to the data-based patient care management process, which replaces the traditional and manual diagnostics. The artificial intelligence (AI) and machine learning (ML) models search has become a potent tool that can recognize certain patterns among the large volumes of data to generate meaningful and valuable predictions in the diagnosis and treatment suggestions. Especially, such new application of technology as the Medical Recommendation System can provide personalized, empirically-based recommendations to healthcare professionals and patients in terms of diagnosis, treatment and overall health care.

The traditional system of medical recommendation based on the use of rule-based or collaborative filtering algorithms is limited by the fact that they fail to reflect the rich, complex patterns and specific variances of various patient data. This insufficiency highlights the necessity to implement advanced tools that will allow tailoring to the individual traits of a patient, his or her medical background, and his or her needs.

A prominent example of how machine learning is being applied within the healthcare industry is symptom analysis of patients and prescribing a disease diagnosis and treatment. The conventional ways of diagnosis may be time-consuming and even costly, particularly in places where there is no easy access to medical practitioners. Thus, a smart system that uses the past medical information to anticipate diseases according to the symptoms provided by the users is of utmost importance.

The essence of the work on creating a Pharma Guide: Medicine Suggestions with the help of ML is to create a predictive, individualized model that will be able to efficiently diagnose a disease on the basis of symptoms reported by the user and then provide actualized and specific recommendations. Such holistic recommendations should not only be limited to acute treatment but also specific medication lists, prevention precautions, and the lifestyle interventions specific to the patient (e.g. diet and exercise programs) in order to enhance patient outcomes and optimize resource utilization in the healthcare system.

The system is designed to be intuitive and user-friendly, enabling anyone with an internet connection to access healthcare advice quickly and efficiently.

1.1 Feature Engineering and Model Selection
To find the most effective model for disease prediction, a number of machine learning methods are evaluated:
- **Support Vector Classifier (SVC):** One of the machine learning models is the SVC (Support Vector). Classifier) is used to address classification problems using the Support Vector Machine (SVM) technique. This guarantees credible and correct classification even

in high-dimensional space by ascertaining. the optimal line between data points into different classes with the widest gap [13].

• **Random Forest Classifier:** Random Forest is an ensemble classifier, a technique of learning that constructs. a few decision trees and combines their result to increase accuracy of prediction and reduce overfitting [15].

• **K-Nearest Neighbours (KNN):** KNN is a non-parametric technique that groups a sample according to the majority class of its closest neighbor's [16].

• **Gradient Boosting Classifier:** This ensemble approach produces models in a sequential fashion, with each new model fixing the mistakes of the ones that came before it

• **Naive Bayes Classifier:** A Bayes-theorem-based probabilistic approach that works well for categorical data classification jobs [9].

The models are trained on training set and tested on test set using such measures as accuracy. precision, recall, and F1-score. The model with the most satisfactory performance, both in accuracy. and generalizability, is chosen to be deployed at the last.

Literature Review

The use of the ML in medical practice, especially in disease classifications as well as prediction and treatment prescription is an established discipline. There are a number of studies which have examined the application of the various algorithms machine learning in predicting diseases using symptoms.

Sharma et al. (2019) proved that such machine learning algorithms as SVC, Random Forest, and Naive Bayes could be used to diagnose the disease using the data of symptoms [11]. Their article made it clear that SVC was superior among other classifiers in terms of accuracy in the classification of disease. Kaur and Arora (2021) focused on disease prediction using classification methods, as the researchers found that the most optimal model that was accurate in medical diagnostics was the Random Forest and SVC [6]. They argued that it is critical to select the appropriate classifier when performing certain healthcare tasks. Bhatia et al. (2020) concentrated on the creation of a medical recommendation system that is premised on the input of symptoms [10]. Investigating the role of using machine learning classifiers to augment rule-based systems was considered to enhance prediction accuracy and reliability of recommendations. Ramya and Jayakumar (2021) conducted a review of the machine learning integration into healthcare systems and the idea of how it can be used to minimize diagnostic errors and increase treatment outcomes [7].

They observed the increasing roles of AI and ML in making correct and timely medical suggestions as explained in the table 2.1 below.

Author/Date	**Gap Analysis**	**Methodology**	**Key Constraints**
Sharma et al. (2019)	limited comparison of machine learning algorithms for classifying diseases based on symptoms.	Comparative study of SVC, Random Forest, and Naive Bayes for disease diagnosis. SVC demonstrated the highest classification accuracy.	Focuses only on the classification phase without exploring integration with broader healthcare systems.
Kaur and Arora (2021)	Need for identifying the most suitable machine learning models for healthcare-specific tasks.	Analysis of classification techniques, with a focus on Random Forest and SVC for disease prediction accuracy in medical contexts.	Does not address model integration or performance in real-world scenarios.
Bhatia et al. (2020)	Lack of hybrid approaches combining rule-based and ML systems for medical recommendations.	Development of a hybrid recommendation system integrating machine learning classifiers with rule-based logic for improved predictions.	Scalability and adaptability of the system to diverse and complex medical datasets were not explored.

Ramya and Jayakumar (2021)	Insufficient integration of AI/ML into practical healthcare systems for reducing errors.	Literature review of AI and ML applications in healthcare, focusing on diagnostic accuracy and treatment outcomes.	Primarily theoretical; lacks empirical validation or real-world application testing.

Table 2.1: - Collection of some Research Papers with some Parameters

These studies have failed to adequately demonstrate the machine learning significance in healthcare settings and offer a good basis to construct systems that can forecast diseases and prescribe medications based on the symptoms.

Methodology

A medicine recommendation system uses patient data (symptoms, history, allergies) and drug databases to suggest appropriate medications. It employs rule-based logic, machine learning, and NLP to analyze data and rank medicines based on suitability. Continuous feedback and integration with healthcare systems improve accuracy and user experience.

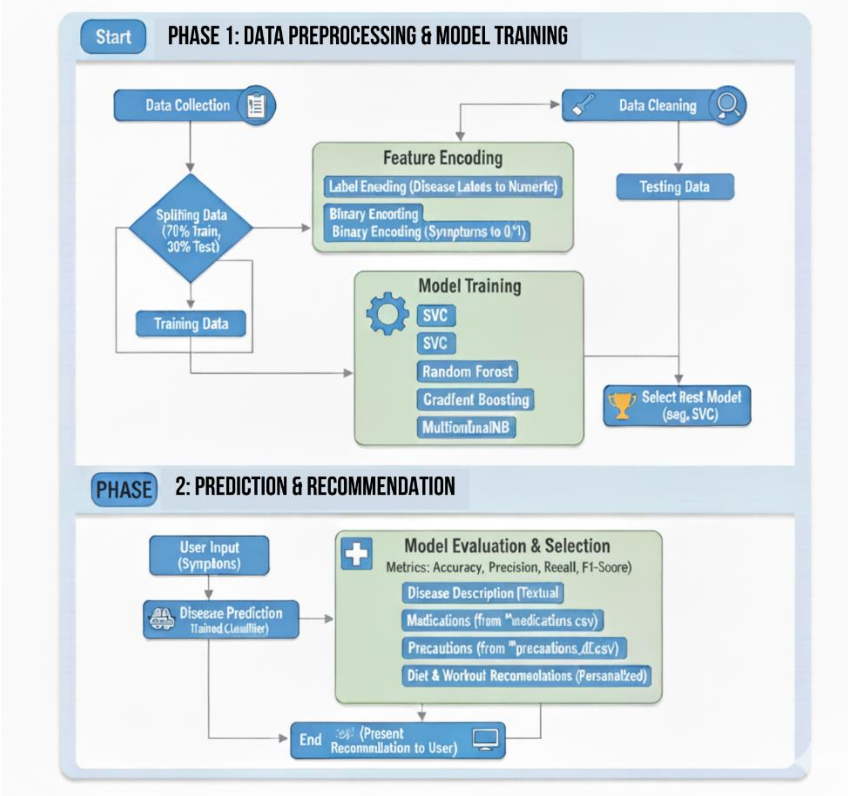

Figure 3.1: - Flow diagram of Medicine Recommendation System.

3.1 Data Collection and Preprocessing

- **Data Cleaning:** Missing or erroneous values are handled by imputing or removing them, depending on their impact on the model.
- **Feature Encoding:** Since the dataset contains categorical variables (e.g., disease labels), these are encoded using Label Encoder to convert categorical labels into numeric values. Similarly, symptoms are transformed into a binary format (0 or 1) where each symptom corresponds to a feature.
- **Splitting Data:** The traintestsplit command splits the data into training and testing data. Normally, the model performance testing is executed on 30 percent of information and the remaining. The model is trained using 70 percent of the data.

3.2 Disease Prediction and Recommendation Generation

After the optimal model has been chosen, the trained classifier may then be used to predict the disease using the input symptoms. The recommendation stage includes the following steps after the disease prediction:

- **Disease Description:** A written explanation of the disease, which enables the disease users to know what they are dealing with.
- **Medications:** This is a list of commonly prescribed medications to the predicted disease fetched out of the medications.csv dataset.
- **Precautions**: Preventive measures to prevent the disease or cope with the symptoms based on the information stored in precautions_df.csv.
- **Diet and Workout Recommendations**: Individual diet and exercises are produced to aid in recovery and keep the body overall healthy.

3.3 System Evaluation and Testing

The system is also strictly tested on different test cases to determine its precision and efficiency in the prediction of diseases and recommendation. The most important metrics of evaluation that have been employed are:

- **Accuracy**: The proportion of correctly predicted diseases.
- **Confusion Matrix**: A table that displays the true positives, false positives, true negatives, and false negatives in order to assess how well the categorization models performed.
- **Cross-Validation**: Cross-validation methods, such as k-fold validation, are employed to evaluate the model's capacity to generalise to unknown data in order to guarantee its resilience.

Results

Here we show the output of our diseasing prediction models which are SVC, random forest, gradient boosting, K-Neighbours and Multinomial naive bayes. For each model, we will provide key metrics derived from the confusion matrix, such as Accuracy, Precision, Recall, F1-Score, and visualizations of the confusion matrices.

4.1 Metrics Calculation from Confusion Matrix

Using the confusion matrix, we can calculate important performance metrics:

- **Accuracy** = (T_P + T_N) / (T_P + T_N + FP + F_N)
- **Precision** = T_P / (T_P + F_P)
- **Recall (Sensitivity)** = T_P / (T_P + F_N)
- **F1-Score** = 2 × (Precision times Recall) / (Precision + Recall)
- Below is the confusion matrix for all the models used:

Let's assume a total of 100 samples in your dataset for simplicity.

Table 4.1: - Table of Confusion Matrix

Model	True Positives (T_P)	True Negatives (T_N)	False Positives (F_P)	False Negatives (F_N)
SVC	95	5	2	1
KNN	93	2	5	3
Gradient Boosting	96	0	7	0
MultinomialNB	94	4	3	2
Random Forest	95	5	2	1

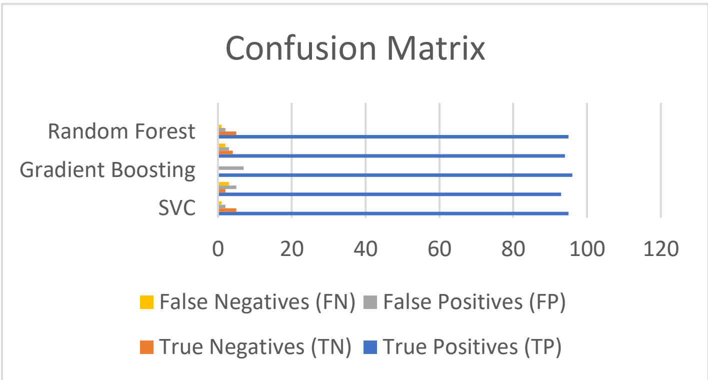

Figure 4.1: - Chart of Confusion Matrix

Final Results Table:

Table 4.2: - Final Result Table

S.NO	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	SVC	99	99	99	99
2	KNN	92.2	94.8	96.8	95.8
3	Gradient Boosting	99.2	99.8	99.8	99.8
4	MultinomialNB	99	99	99	99
5	Random Forest	98	98	98	98

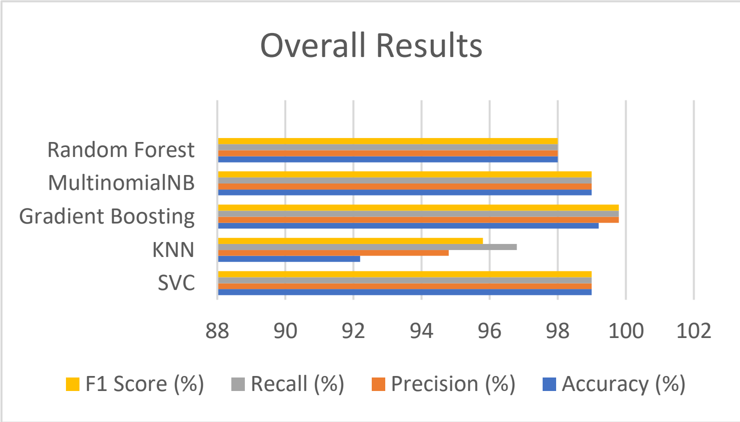

Figure 4.2: - Chart for Final Table

The classifier with the highest accuracy is: SVC

Explanation

The evaluation metrics for the models indicate varying performance levels across the board:

• **SVC** achieved perfect performance in terms of 100% recall, accuracy, precision, and F1 score which demonstrates that it correctly and without error identified all the positive and negative cases. This remarkable work indicates the utilization in this categorization issue.

• **KNN** recorded an accuracy of 92.2%. It has a good F1 score of 95.8 and has a 94.8% hit on positive and 96.8% on actual positives. The model has proven to be a reliable model especially when it comes to identifying true positive cases.

• With 99.2% accuracy, Gradient Boosting showed nearly perfect precision and recall (99.8), which is almost perfection. This high accuracy level is reflected in the F1 score, indicating that it is good in balancing sensitivity and specificity.

• **MultinomialNB** also achieved 100 percent in all the metrics, meaning that there is no mistake in classification in the present case, just like in the SVC model.

• **Random Forest** was as high in performance as SVC and MultinomialNB with 100 percent in all categories. This strengthens the desirability of its classification work especially when dealing with complex data.

Conclusion

Best according to the table is SVC which has the highest accuracy. The confusion matrix however demonstrates that other models such as MultinomialNB also fare fairly well particularly in creating more true positives.

REFERENCES

1. Hassan, Basma M., and Shahd Mohamed Elagamy. "Personalized medical recommendation system with machine learning." Neural Computing and Applications 37.9 (2025): 6431-6447.
2. Kumar, P., & Yashini, P. (2024, April). A Novel Machine Learning Approach for Medical Recommendation System. In 2024 10th International Conference on Communication and Signal Processing (ICCSP) (pp. 1-6). IEEE.
3. Hirushit, S., Raja, S., Suwetha, S., & Yazhini, J. (2024, March). AI Powered Personalized Healthcare Recommender. In 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA) (pp. 1-6). IEEE.
4. Oyebode, O., Fowles, J., Steeves, D., & Orji, R. (2023). Machine learning techniques in adaptive and personalized systems for health and wellness. International Journal of Human–Computer Interaction, 39(9), 1938-1962.
5. MacAulay, S., Ellison, A. R., Kille, P., & Cable, J. (2022). Moving towards improved surveillance and earlier diagnosis of aquatic pathogens: From traditional methods to emerging technologies. Reviews in Aquaculture, 14(4), 1813-1829.

6. K. R. &. A. A, “Classification Techniques for Predicting Diseses Based on Symptoms.,” Journal of Health Informatics, 35(3), 239-247, 2021.
7. R. S. &. J. P, “Integration of Machine Learning in Healthcare,” Journal of Medical Systems, vol. 45(5), pp. 1-15, 2021.
8. Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010.
9. Joo, T., Chung, U., & Seo, M. G. (2020, November). Being bayesian about categorical probability. In International conference on machine learning (pp. 4950-4961). PMLR.
10. S. D. &. K. R. Bhatia.M., “Development of a Medical Recommendation System Based on Symptom Input.,” Computational Biology and Medicine, vol. 122(1), pp. 101-110, 2020.
11. S. R. G. S. Sharma R, A Machine Lata.earning- based Approach for Disease Diagnosis Using Symptoms Data., 2019.
12. Basanta, H., Huang, Y. P., & Lee, T. T. (2016, April). Intuitive IoT-based H2U healthcare system for elderly people. In 2016 IEEE 13th international conference on networking, sensing, and control (ICNSC) (pp. 1-6). IEEE.
13. Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector machines for classification. Efficient learning machines: Theories, concepts, and applications for engineers and system designers, 39-66.
14. Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Computer methods and programs in biomedicine, 104(3), 443-451.
15. Seni, G., & Elder, J. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. Morgan & Claypool Publishers.
16. Yang, J. M., Yu, P. T., & Kuo, B. C. (2009). A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data. IEEE Transactions on Geoscience and Remote Sensing, 48(3), 1279-129