Personalized Healthcare System Using Machine Learning

Aryan Beluse[1], Soham Dhavale[2], Atharva Jaid[3], Ayush Shinde[4], Prathamesh Mahajan[5], Noshir Tarapore[6]

[1,2,3,4,5] LY B. Tech Computer Engineering, Science & Technology, Vishwakarma University, Pune, India – 411048

[6]Assistant Professor, Dept. of Computer Engineering, Vishwakarma University, Pune, India – 411048

Abstract - *Health management has become increasingly complex with the rise of lifestyle-related conditions, chronic diseases, and the need for personalized care. Despite a growing interest in preventive healthcare, many individuals struggle to access reliable, actionable health information tailored to their unique needs. This paper presents a comprehensive, machine-learning-based healthcare application designed to empower users with personalized health recommendations, including diet and exercise routines, symptom-based disease suggestions, and preventive measures. Through distinct user portals for patients and doctors, the platform enables customized care, allowing doctors to analyze patient data and offer personalized treatment plans. Our goal with this application is to create a holistic, accessible health management system that bridges the gap between patients and healthcare providers and offer personalized health recommendations using machine learning.*

Key Words: *Disease prediction, SVC, Random Forest, Naive bayes, Personalized recommendation*

1.INTRODUCTION

The increasing focus on preventive healthcare and personalized treatment has driven a demand for accessible digital health solutions that empower individuals to take control of their wellness. Traditional healthcare systems, while essential, often lack the capacity to provide tailored guidance on a regular basis, leading patients to search for health advice on various online platforms that may not be reliable or personalized. This project aims to address these challenges by developing an integrated healthcare application that combines personalized health recommendations, symptom-based disease prediction, and continuous health monitoring.

Our application offers separate login options for doctors and patients, allowing both user groups to access tools suited to their needs. For patients, the application provides personalized diet and lifestyle recommendations, preventive measures, and symptom-based suggestions to help them maintain or improve their health. Through QR-enabled record-keeping, patients can easily update and access their health information, streamlining management of their health history.

Doctors, on the other hand, gain access to a range of tools, including the ability to conduct specific analyses and suggest tailored treatment plans. By empowering healthcare providers with such tools, the platform enables more effective and personalized patient care. Additionally, the application promotes community involvement by including information about fitness events, NGOs, and charitable fundraisers, thereby fostering a comprehensive approach to health and wellness that goes beyond individual treatment and encourages preventive and community-centered health initiatives.

1.2 Need of Study

This study addresses the growing need for personalized healthcare solutions, driven by the limitations of traditional, reactive healthcare models that often overlook preventive care. With rising awareness of chronic diseases and the importance of proactive health management, there is an urgent need for accessible, reliable sources of health information and tools that empower individuals to monitor and manage their well-being.

This study proposes a comprehensive web application that centralizes health information, symptom tracking, and tailored health recommendations. For individuals in remote or underserved areas, the app provides essential resources, bridging the gap in access to healthcare by enabling digital storage and sharing of health data. This platform also supports healthcare providers by giving them access to patient records, facilitating personalized care, and simplifying appointment scheduling. Through its secure, data-protective infrastructure, this study highlights a solution that not only meets today's healthcare needs but promotes preventive care and healthier lifestyles.

2. LITERATURE SURVEY

The integration of technology in healthcare has led to the development of various digital tools designed to enhance patient care, streamline medical record management, and improve health outcomes through personalized recommendations. Research has shown that personalized healthcare applications, which offer individualized guidance based on a patient's specific health profile, can significantly improve adherence to preventive measures and treatment plans.

In the area of personalized diet and lifestyle recommendations, studies have found that machine learning algorithms are capable of analyzing individual health data to generate specific recommendations for diet, exercise, and other lifestyle choices. Researchers such as D. Wang and J. Zhang has explored the use of these algorithms to predict health risks based on user inputs, showcasing the potential for early detection and prevention of diseases. These studies highlight the importance of creating digital health solutions that utilize advanced algorithms for customized health advice.

Recent advancements in machine learning have facilitated predictive healthcare applications, enabling early detection and management of diseases. Several studies have investigated the application of machine learning for disease prediction, each employing various algorithms and methodologies tailored to specific diseases or healthcare needs.

A study by K. Sharma explored the use of symptom-based machine learning models for disease prediction, focusing on improving diagnosis accuracy by analyzing symptom patterns. This approach aims to reduce the diagnostic burden on healthcare providers by automating initial assessments, which can be especially valuable in resource-limited settings.

Machine learning has proven effective in predicting chronic conditions, as demonstrated by a study on data preprocessing for chronic disease prediction. The study analyzed the performance of algorithms such as Decision Trees, Support Vector Machines (SVM), and Neural Networks, finding that preprocessing steps like data normalization significantly enhanced prediction accuracy. This highlights the importance of preparing clinical datasets to optimize the performance of machine learning models in chronic disease prediction.

Another study delved into disease prediction using general patient data, where algorithms like K-Nearest Neighbours (KNN), Naïve Bayes, and ensemble methods were employed to predict diseases based on demographic and health metrics. The study underscores the potential of using machine learning to analyze large-scale healthcare data and assist in early diagnostics, reducing healthcare costs and enabling timely intervention.

Integrating machine learning and deep learning, research by S. Rajetal. proposed a multi-disease prediction model that combines machine learning with web technologies for real-time analysis and prediction of diseases. This model allows healthcare providers and patients to access diagnostic predictions online, thereby enhancing accessibility and convenience in health monitoring.

3.IMPLEMENTATION

3.1.System Architecture

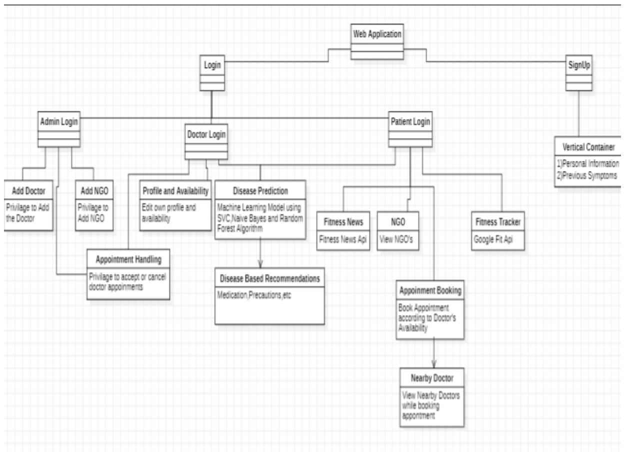

Fig 3.1-System Architecture

The proposed system architecture outlines a web-based healthcare application that provides various functionalities tailored to three user roles: Admin, Doctor, and Patient. The system has three primary modules: Login and Authentication, User-specific Functionalities, and Health & Fitness Support Services.

a. Login and Authentication:

The application supports login access for three user types: Admin, Doctor, and Patient. A dedicated Sign Up process allows new patients to register by providing personal information and recording any previous symptoms.

b. Admin Functionalities:

Upon successful login, Admin users gain access to privileged actions like adding doctors and NGOs to the platform. They also have control over Appointment Handling, which allows them to approve or cancel doctor appointments.

c. Doctor Functionalities:

Doctors, upon logging in, can manage their profiles and set availability for appointments. The system also integrates a Disease Prediction feature that uses machine learning algorithms (SVC, Naive Bayes, and Random Forest) to assist doctors in diagnosing patient conditions based on symptoms. Following a diagnosis, doctors can provide Disease-Based Recommendations for medication, precautions, and other necessary advice.

d. Patient Functionalities:

Patients can log in to access various features, including Appointment Booking, which allows them to schedule appointments according to doctor availability. Additionally, they can use the Nearby Doctor feature to find doctors within their vicinity while booking.

e. Health & Fitness Support Services:

The system provides health-related content through a Fitness News API, which supplies patients with up-to-date fitness and wellness information. A section for NGOs allows patients to view local healthcare and wellness organizations. The Fitness Tracker, powered by the Google Fit API, enables patients to monitor their health and fitness metrics.

3.2 Machine Learning Model

Dataset: In order to obtain some dataset and train our model we looked up some data online, and created an arrow dataset by merging all of that information. As a result, we now have a dataset.

Lastly, the disease class column. A few rows in the dataset that show diseases and the symptoms that go along with them.

	Disease	Symptoms
1	Malaria	{chills, vomiting, high_fever, sweating, heada...
2	Allergy	{continuous_sneezing, shivering, chills, water...
3	Fungal infection	{skin_rash, nodal_skin_eruptions, dishromic_...
4	Gastroenteritis	{vomiting,sunken_eyes, dehydration, diarrhoea
5	arthritis	{muscle_weekness, stiff_neck, swelling_joints,...
6	Typhoid	{chills,vomiting, fatirgue, high_fever, headac...
7	Hypertension	{muscle_weakness, stiff_neck, swelling_joint,....

Table 3.1-Symptoms Dataset

Data Pre-processing -Since the collected data is raw, we must prepare it for use in training our machine learning model. To make the data acceptable for machine learning models, we used Python packages like NumPy and pandas.

Our data is now prepared for output prediction using machine learning methods. We have employed the Naive Bayes algorithm since our problem falls within the category of unsupervised machine learning.

The goal column, or prognosis, is a string type that is converted to numerical form using a label encoder. All of the other columns in the dataset are numerical.

Data Splitting- The data is later split into training, validation, and test sets to ensure that the model's performance is evaluated accurately. The following steps were taken:

1. Train-Test Split: Initially, the data was divided into a training set (80%) and a test set (20%) to evaluate final model performance on unseen data.

2. Validation Set (Cross-Validation): To enhance robustness and avoid overfitting, we applied K-Fold Cross-Validation on the training set with k=5 (for example). This allowed us to iteratively train the model on k-1 folds and validate on the remaining fold, providing a better estimate of generalization error.

Model Selection- After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We used this cleaned data to train the Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier. We used a confusion matrix to determine the quality of the models.

Support Vector Classifier- SVC is a supervised learning model that classifies data points by finding a hyperplane that best separates classes in the feature space. It's particularly effective for binary classification and excels in high-dimensional spaces. For disease prediction, SVC can classify based on symptoms and other relevant health features by maximizing the margin between categories, helping to distinguish between diseases with high precision.

Random Forest Classifier-Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs for more robust predictions. Each tree is trained on random subsets of data and features, making it highly resistant to overfitting and noise. Random Forests are well-suited for handling complex datasets in disease prediction, as they aggregate multiple "weak" learners to form a strong predictor, enhancing both accuracy and interpretability.

Naive Bayes-Field monitoring for machine learning and the different model choices it makes Numerous methods, including tree, naïve Bayes, and random forests, are available to forecast different diseases. Three distinct naïve Bayes models exist: the Gaussian, Multinomial, and Bernoulli models. The accuracy varies from model to model.

All three models essentially use the same application and data fitting to evaluate disease outcome. This project work was completed utilizing the Gaussian naïve Bayes because it is considerably simpler and easier to understand than the other two.

. Code:

```
from sklearn.naive_bayes import MultinomialNB
```

mnb = MultinomialNB()

Naïve Bayes classifier depends on Bayes Theorem.

Bayes theorem:

P(Y/X1,X2,…….,Xn)= P(Y) P(X1,X2,…….,Xn)

P(X1,Xn)

Where, Y is the class Variable

X1,X2,…..,Xn are the dependent features.

Model Tuning-

1.SVM: Tuned kernel type and regularization parameter (C).

2.Random Forest: Tuned number of trees (n_estimators), maximum depth, and minimum samples per leaf.

3.Naive Bayes: Default parameters were used as they typically perform well without significant tuning.

Fig 3.2- Machine learning model architecture

Classification Report-Classification report visualizes the precision recall and F1 score of a model.

Models	SVC	Naive Bayes	Random Forest Classifier
Accuracy	98.7	98.9	99
Precision	98.88	99	99
Recall	98.78	98.8	99
F1 Score	98.77	98.9	99

Table 3.2-Classification Report

We can see from the output above that all machine learning algorithms are operating at extremely high performance levels, and the mean scores following k fold cross-validation are likewise quite high. We can combine, that is, take the mean of the forecasts of all three models, to create a resilient model. This way, even if one model makes a mistaken prediction and the other two do, the final result will be accurate.

This method will enable us to maintain considerably more accurate forecasts on data that has never been seen before. The code below will train all three models on the train data, use a confusion matrix to assess the models' quality, and then aggregate the predictions from all three models.

Personalized Recommendations

After predicting diseases, this system goes a step further by offering individualized health recommendations tailored to each user's specific condition. These recommendations address five key areas of wellness, each aimed at promoting recovery, maintaining health, and preventing future ailments. These aspects are:

Medicine: The system suggests over-the-counter medicines, supplements, or other pharmacological interventions that are appropriate for the predicted disease. This serves as a first step toward treatment while advising users to seek medical consultation.

Diet: Based on the predicted disease, the system provides dietary suggestions tailored to support healing, reduce symptoms, and bolster immunity. This dietary plan includes recommended foods, portion sizes, and frequency of meals, aiming to supply essential nutrients and minimize triggers for specific health issues.

Workout: The system recommends exercises or physical activities that are safe and beneficial for the user’s predicted condition. For instance, for a disease that may impact joint health, the system may suggest low-impact activities, while for metabolic issues, it may advise cardiovascular exercises.

Precautions: Precautionary measures help mitigate symptoms, prevent exacerbation, and reduce the likelihood of disease recurrence. This may include lifestyle modifications, avoiding certain substances or activities, and suggestions for maintaining optimal hygiene and safety.

Description: A brief overview of the predicted disease, covering its primary symptoms, potential causes, and health impacts. This information helps the user understand the disease's effects on the body and motivates informed decision-making in their treatment and lifestyle choices.

Implementation of Personalized Recommendations

The recommendations are generated by mapping each predicted disease to its optimal treatment and wellness plan, drawn from medical literature and expert health resources. A recommendation engine links each disease to specific advice across the five aspects, ensuring that users receive a holistic, integrative approach to their health management.

4.CONCLUSION

This study presents a comprehensive approach to predicting diseases and providing personalized recommendations by leveraging machine learning models. By utilizing a multi-model approach, we integrate the outputs of three models: Support Vector Classifier (SVC), Multinomial Naive Bayes (MNB), and Random Forest (RF).

These models were trained on a dataset containing a wide array of symptoms and their corresponding diseases, allowing the system to make accurate predictions when given a set of input symptoms.

The model's predictions are not just limited to disease identification but are also extended to provide personalized health recommendations in five key areas:

1. Medicines
2. Diet
3. Workout
4. Precautions
5. Description

The accuracy of the models was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. By using K-fold cross-validation, the models' performance was optimized and validated for generalization. The integration of multiple models (SVC, MNB, and RF) ensures that predictions are made more robustly by taking the mode of the results from each model, which reduces the likelihood of error and improves the reliability of the recommendations.

By offering personalized, actionable health insights through machine learning-based predictions, this system provides a valuable tool for individuals seeking to manage their health in a more informed and proactive manner. With continuous refinement of the model and expansion of the dataset, this system has the potential to offer even more accurate predictions and more customized recommendations, supporting a wide range of users across various health conditions.

In conclusion, this research demonstrates the feasibility and potential impact of using machine learning models in healthcare, with an emphasis on personalized recommendations in medicine, diet, exercise, precautions, and disease description. The system represents a promising step forward in applying AI to personalized health management, providing a scalable solution that can be expanded with more data and additional disease classifications for broader applicability.

5. REFERENCES

1. K.M., Bakar, A.A. and Othman, Z.: Medical data classification with Naïve Bayes approach. Information Technology Journal. 11(9), 1166 (2012).

2. Bhadra, S., & Borah, S. (2018). A Comprehensive Survey on. Personalized Medicine Recommendation Techniques. International Journal of Advanced Research in Computer Science, 9(1), 1-8.

3. Liu, Z., Guo, F., Gu, J., & Wang, Y. (2018). Integration of Omics Data and Machine Learning for Personalized Medicine: A Review of Current Approaches and Future Directions. IEEE Reviews in Biomedical Engineering.

4. Pattekari, S.A. and Parveen, A.: Prediction system for heart disease using Naïve Bayes. International Journal of Advanced Computer and Mathematical Sciences. 3(3), 290-294 (2012).

5. Masethe, H.D. and Masethe, M.A.: Prediction of heart disease using classification algorithms. In Proceedings of the World Congress on Engineering and Computer Science. 2, 22- 24 International Association of Engineers, Francisco (2014).

6. Shouman, M., Turner, T. and Stocker, R.: Using decision trees for diagnosing heart disease patients. In: Proceedings of the Ninth Australasian Data Mining Conference, Volume. 121, 23-30. Association of Computing Machinery, Victoria (2011).

7. Khraisat, A., et al. (2019). Machine Learning Models for Personalized Medicine Recommendation: A Review. International Journal of Advanced Computer Science and Applications, 10(9), 7-14.

8. Vembandasamy, K., Sasipriya, R. and Deepa, E., 2015. Heart disease detection using the Naive Bayes algorithm. International Journal of Innovative Science, Engineering & Technology, 2(9), pp. 441-444.

9. Srinivas, A., et al. (2019). A Review on Personalized Medicine Recommendation Techniques in Online Transactions. International Journal of Scientific Research in Computer Science, Engineering, and Information Technology (IJSRCSEIT), 4(4), 467-472.