IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ai For Disease Prediction And Tailored Healthcare

[1]Isha Santosh Varade, [2]Sakshi Dasharath Shinde, [3]Sanika Naresh Butle, [4]Sawani Tarkeshwar Doshi

[1234]Third Year Student ,
[1234]Department Of Computer Engineering,
[1234]Vishwakarma Institute Of Information Technology , Pune , India

Abstract: This research paper present the design of web application based medical recommendation system. The system predicts diseases according to user input symptoms and recommends corresponding precaution ,diet and workout. Flask framework is used to create the web application. For prediction of disease, we have applied various machine learning models like Random Forest ,K-neighbours , Gradient Boost ,Support Vector Machine(SVM),Naïve Bayes .Then using data filtering in panda, system recommends precaution ,diet and workout for that predicted disease. Gemini integrated chatbot is also added to make user friendly platform. System also recommends nearby hospitals using Google maps.

***Index Terms* – Machine Learning , Disease Prediction , Health Recommendations.**

I. Introduction

1.1 Background

The World Health Organization and World Economic Forum reported that India had a huge loss of $236.6 billion by 2015 because of fatal diseases, caused by malnutrition and morbid lifestyles. This indicates how essential it is to detect the diseases early in order to reduce the fatality. In addition, early disease prediction can lessen the financial pressure on the economy and ensure better maintenance on the overall well-being of the community (Ferjani et al,2010) .

In recent years healthcare sector has witnessed a rapid growth in terms of data availability due to advancements in digital health technologies, electronic health records (EHRs), and the increasing prevalence of wearable devices (Dey et al ,2019).This vast amount of data has led to new opportunities in using machine learning for disease prediction.

Machine learning algorithms use various stastical, optimization and probabilistic methods to learn from the data collected from past experiences and apply it in decision making (Uddin et al,2019) .Generally supervised ml models are used for disease prediction. In this method, datasets with known labels are used to train the model, which is then predict outcomes for unlabeled data (Hasija et al ,2017).

1.2 Research problem

Existing systems primarily focus on disease prediction based on symptoms but integration of recommendations for corresponding health advices like diet, workout, precautions is still an emerging field. Furthermore, interactive features such as Gemini powered chatbot, which can guide users about diagnosis process are rarely included in these systems.

1.3 Objectives

This research aims to develop a personalized medical recommendation system that addresses the following objectives :

1.To predict the disease based on user input symptoms using machine learning models like Random Forest, SVM ,Naïve Bayes, KNN, Gradient Boost.
2.To recommend health advices like diet plans, workout routines, precautions corresponding to predicted disease.
3.To enhance user interaction by integrating Gemini-powered Chatbot.
4.To recommend nearby medical centers based on user's location.

II. RELATED WORK

The rapid development of machine learning(ml) has drastically transformed healthcare by improving accuracy and efficiency of medical recommendation systems. Many studies have explored how ml models can solve specific healthcare problems and improve patient care by personalized recommendations.

Monti et.al proposed a healthcare system named as Cross-Language Information Retrieval (CLIR) which uses big data and NLP. The system architecture includes an Intelligent Recommender for Healthy Diet (IR-HD) and a Hybrid Recommender System (HRS).This method uses data processing to make food recommendations, showing how ml and big data can be merged in healthcare applications (Lambay et al,2022).

Abhirup Dey et.al proposed merging of IoT data collected from wearable devices with machine learning to identify user behaviors and recommend health-related activities. For its implementation they have used various ml algorithms like Random Forest, KNN, Support Vector Classifier (SVC) and Logistic Regression (Dey et al,2019) .

For heart disease prediction , one study employed a Hybrid Random Forest with Linear Model (HRFLM) strategy. This hybrid strategy combines the neural networks and linear models. The results showed enhanced prediction accuracy in heart disease diagnoses by efficiently processing high-dimensional datasets and non-linear data patterns (Kiran et al,2022).

Islam et.al conducted the study regarding how ml models can be used in early-stage chronic disease prediction. The result showed that SVM model performed well in predication of diseases such as diabetes, cancer, and neurological problems while Random Forest excelled at kidney disease and breast cancer prediction. The study emphasized the relevance of combining feature selection approaches such as LASSO and RELIEF to improve model performance (Islam et al,2024).

The Disease-Diagnosis and Disease-Treatment Recommendation System (DDTRS) demonstrates further developments in clustering and association rules, since it uses the Apriori method for disease-symptom clustering and Apache Spark for parallel processing. This method enhances data processing for large-scale medical datasets, lowering latency and improving treatment recommendations (Chen et al,2018) .

Finally, a comprehensive analysis examined the evolution of machine learning in healthcare recommender systems, comparing Collaborative Filtering, Content-Based Filtering, and Hybrid Systems. The paper highlighted the constraints of cold-start issues and data sparsity, particularly in collaborative filtering models, and examined neural networks' potential to improve recommendation system performance in healthcare (Ko et al,2022) .

These works collectively highlight the progress and continued problems of applying machine learning for medical recommender systems, paving the way for future developments that include big data, IoT, and complex ML algorithms to improve customized healthcare.

III. SYSTEM ARCHITECTURE OF WEB APPLICATION

The system consists of several modules, including:

1.Symptom Input Module: Takes symptoms as input from the user in text or voice format.
2.Disease Prediction Module: Utilizes a pre-trained ML model to predict diseases.
3.Recommendation Module: Recommends a description, diet, workout, and precautions for the predicted disease using filtered datasets.
4.Chatbot Integration: Gemini chatbot facilitates natural language interaction with the user.
5.NearBy Medical Centers Recommendation: Recommends medical centers nearer to user's location using google map.

Figure 3.1 shows the system architecture overview :

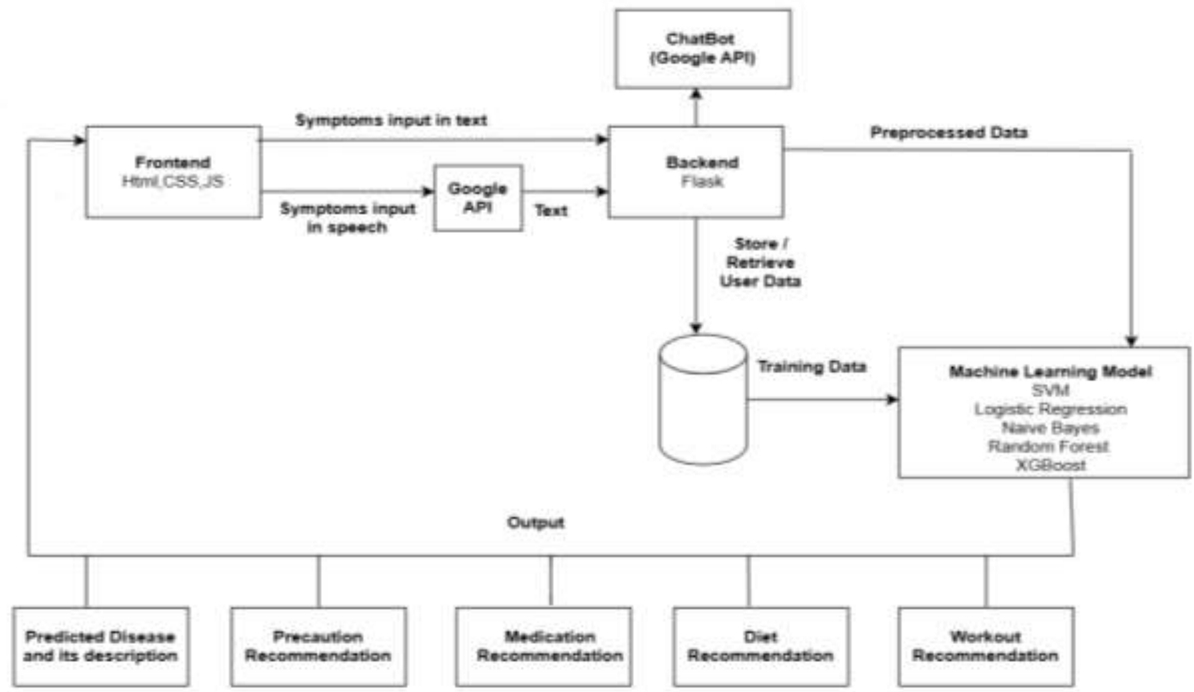

Figure 3.1:System Architecture of web application for ml driven disease prediction and tailored health recommendations

3.1 Machine Learning Model

We trained our system on a dataset containing symptoms and prognosis to predict diseases. We tested several machine learning models, including:

- Gradient Boost
- Naïve Bayes
- K-Nearest Neighbors (KNN)
- Random Forest
- Naïve Bayes
- Support Vector Machines (SVM)

The models were evaluated based on accuracy.

3.1.1 Dataset

Our training dataset contains records of symptoms with corresponding disease labels. Separate datasets for each disease include descriptions, diets, medications, workouts, and precautions. The system uses these datasets to offer personalized health recommendations to the user.

3.1.2 Model Evaluation

After training the model ,we evaluated their performance on the dataset using various algorithms including Support Vector Classifier (SVC), Random Forest, Gradient Boosting, K-Nearest Neighbors (KNN), and Multinomial Naive Bayes. Each model achieved an accuracy of 100% on the test set, with confusion matrices confirming these results as shown in figure 3.1.4, showcasing no misclassifications across different categories.

Here is a summary of accuracy for each model:

1. SVC: 100% accuracy
2. Random Forest: 100% accuracy
3. Gradient Boosting: 100% accuracy
4. K-Nearest Neighbors: 100% accuracy
5. Multinomial Naive Bayes: 100% accuracy

These findings indicate all models perform exceptionally well and highly suitable for disease prediction based on symptom input.

3.1.3 Recommendation System

Based on disease predicted,the recommendation system provides users a comprehensive set of health insights, which includes a description, precautions, medications, dietary suggestions, and workout routines. This is done by loading databases for symptoms, precautions, workouts, medications, and diets. The helper function fetches data specific to each disease from these sources.

- **Symptoms Input**: A dictionary used to map each symptom to an index for easier processing and retrieval.
- **Disease Matching**: Based on the predicted disease, the helper function retrieves:
  - **Description**: Overview of the disease.
  - **Precautions**: Steps to avoid or manage the disease.
  - **Medications**: Recommended treatments or medications.
  - **Diet**: Suggested dietary changes.
  - **Workouts**: Exercises that may help in managing symptoms.

3.1.4 Results

Our models achieved outstanding accuracy, effectively predicting diseases based on input symptoms given by users. The recommendation system further improves this by giving tailored health advice based on predictions, helping users better understand and manage their conditions. This platform acts as a comprehensive tool for personalized health insights, combining accurate prediction with actionable recommendations, thus enhancing user engagement and supporting proactive health management.

```
SVC Accuracy: 1.0
SVC Confusion Matrix:
[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]

========================================

RandomForest Accuracy: 1.0
RandomForest Confusion Matrix:
[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]

========================================

GradientBoosting Accuracy: 1.0
GradientBoosting Confusion Matrix:
[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]

========================================
```

```
KNeighbors Accuracy: 1.0
KNeighbors Confusion Matrix:
[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]

========================================

MultinomialNB Accuracy: 1.0
MultinomialNB Confusion Matrix:
[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]

========================================
```

Figure 3.1.4:Confusion matrix and accuracy of ml models

3.2 Chatbot Integration

For enhancing user interaction the system is integrated with the Gemini chatbot.The chatbot allows users to inquire about symptoms and get real-time responses about their condition.

System Architecture

User Interface: Users interact with the chatbot through a command-line interface (CLI), but it can be easily adapted to a web or mobile interface in production.
Medical Chatbot (TF-IDF-Based): User input is first processed by a pre-trained TF-IDF vectorizer. This vectorizer then compares the question to a local database of healthcare data. The system identifies the best matching question-answer pair using cosine similarity. If the similarity score exceeds a predefined threshold(e.g., 0.3), a response is generated.
Fallback to Gemini API: The system queries the Gemini API if the medical chatbot unable to find a high-confidence response. Gemini API uses a large language model (LLM) which can generate responses based on a broader range of medical information. The Gemini API helps in giving responses for complex or unstructured questions that are not covered in the local database.
Response Integration: The system gives appropriate outputs which is either TF-IDF-based or Gemini-generated to the user, providing seamless integration between both kinds of responses.
Environment & Security: The system is secured using environment variables for API keys and employs lemmatization and tokenization for consistent text processing which improves both performance and security.

VI. WEBSITE INSTANCES

Figure 6.1:Homepage With Medical Chatbot

Figure 6.2:Symptoms input

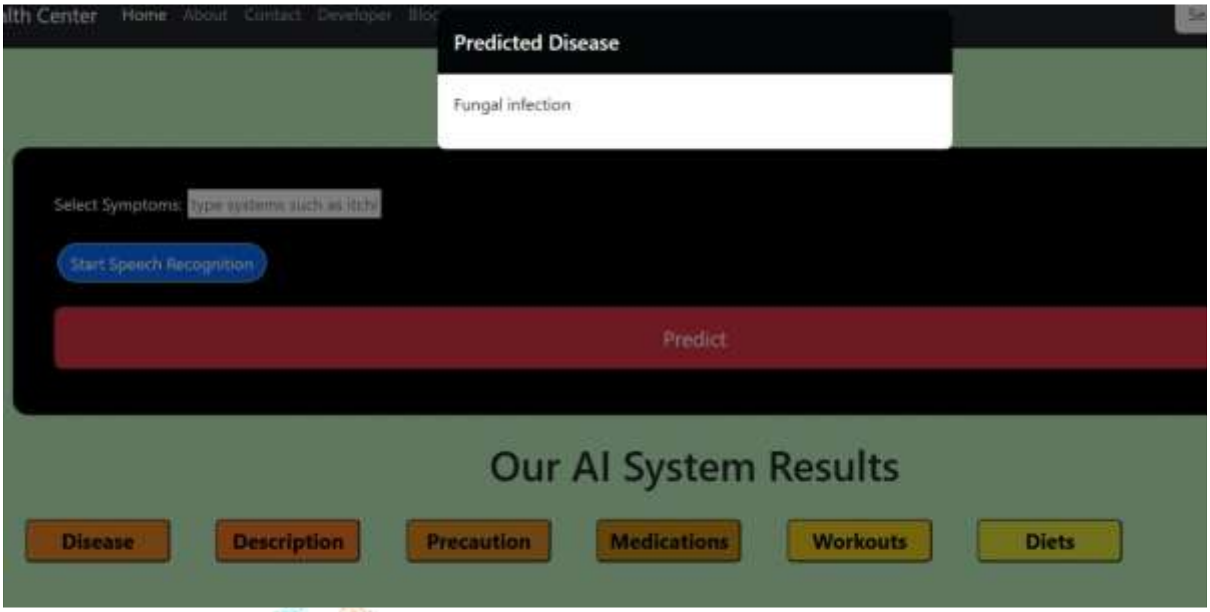

Figure 6.3:Disease Prediction

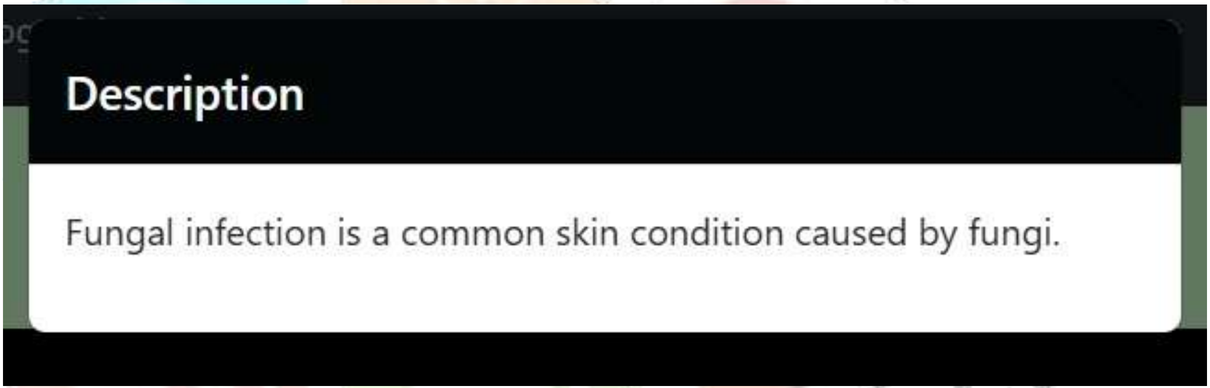

Figure 6.4:Disease Description

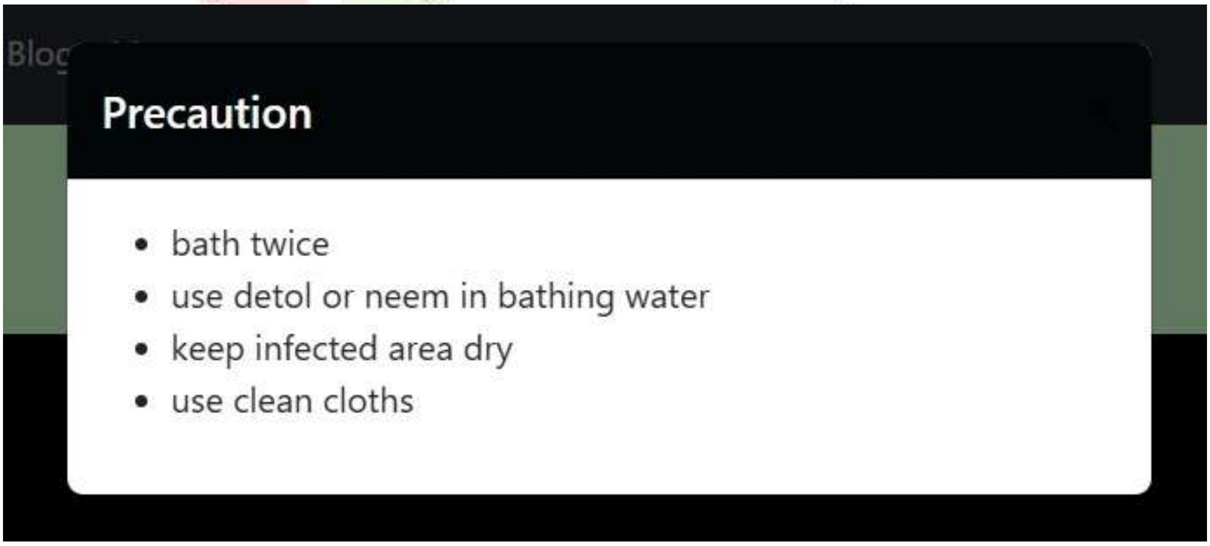

Figure 6.5:Precaution Recommendation

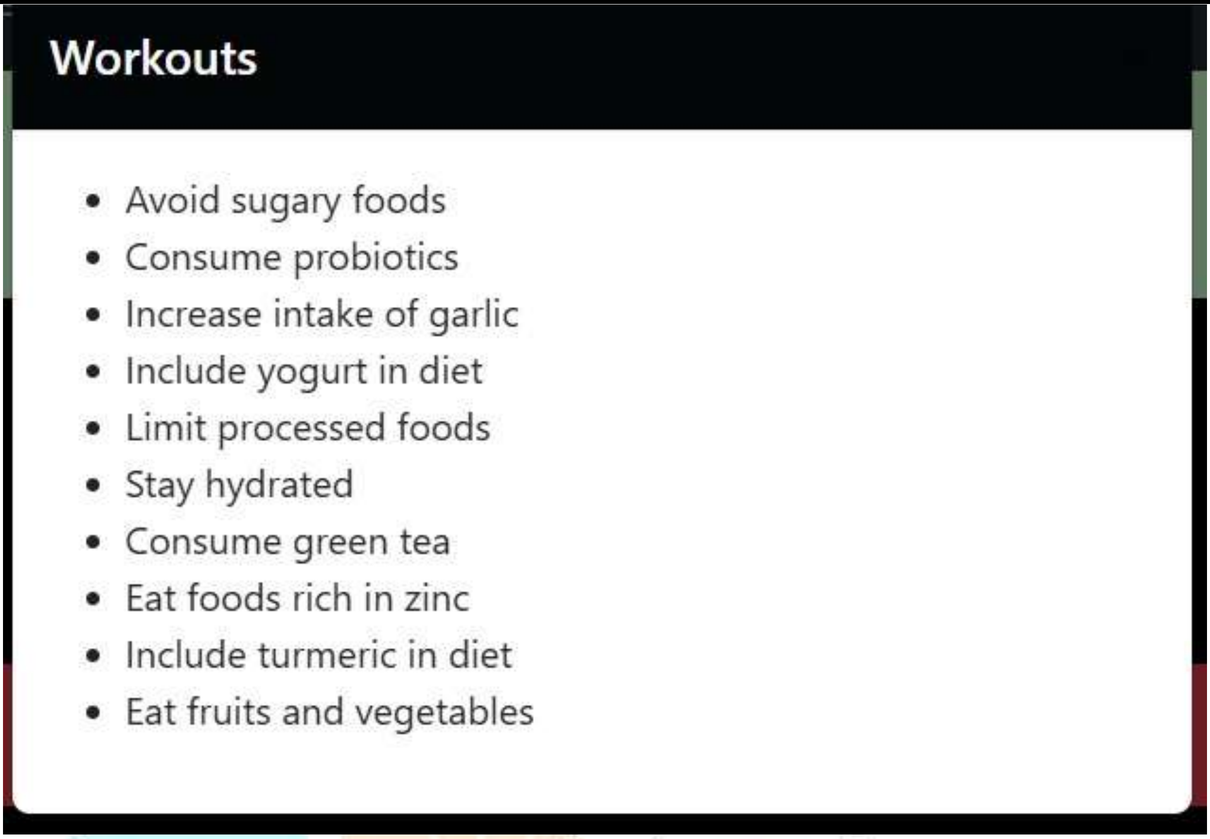

Figure 6.6:Workout Recommendation

Figure 6.7:Diet Recommendation

Figure 6.8:Nearby Hospital Recommendation

VII. CONCLUSION

The Personalized Medical Recommender Web Application developed in this study is user-friendly and effective platform. It integrates the ai-ml and web technologies to provide efficient early disease prediction and tailored health recommendations to the user. For accurate disease prediction, system uses various ml models such as SVM, Random Forest, K-Nearest Neighbors, Naïve Bayes, and Gradient Boost. Based on predicted diseases, System provides health insights like precaution, diet, medications and workout routines to help patient manage predicted condition. The integration of Gemini-powered chatbot enhances user interaction by providing assistance throughout the diagnostic process. System also gives information about nearby medical centers based on user's location.

The result shows that ml models performed well giving higher accuracy and correctly predicting the diseases. This system's unique combination of prediction and tailored recommendations not only helps users better understand and manage their health, but it also demonstrates machine learning's promise in preventive healthcare. The use of Flask as a backend framework ensures scalability and flexibility, making the system suitable for larger use and future growth.

In future, several improvements can be added in the system to enhance its impact. Expanding the dataset to include a broader range of diseases and symptoms, implementing NLP for dynamic user input, and incorporating real-time data from wearable IoT devices could all increase the system's effectiveness in modern healthcare. Furthermore, advances in the chatbot's capabilities and feature selection approaches, inspired by recent research in machine learning and health advice, would allow for a more customized, responsive experience for consumers. Addressing data sparsity and cold-start concerns, particularly in collaborative filtering, may improve recommendation accuracy and user happiness. Collectively, these advancements open the way for a truly tailored and scalable medical recommendation system, moving us closer to an era of easily available, AI-powered healthcare solutions.

REFERENCES

[1] Subramani S, Varshney N, Anand MV, Soudagar MM, Al-keridis LA, Upadhyay TK, Alshammari N, Saeed M, Subramanian K, Anbarasu K and Rohini K (2023) Cardiovascular diseases prediction by machine learning incorporation with deep learning. Front. Med. 10:1150933.). https://doi.org /10.3389/fmed.2023.1150933

[2] Mondal, Aishwarya & Mondal, Banhishikha & Chakraborty, Amaresh & Kar, Agnita & Biswas, Ayanty & Majumder, Annwesha. (2023). Heart Disease Prediction Using Support Vector Machine and Artificial Neural Network. Artificial Intelligence and Applications. https://doi.org/2. 10.47852/bonviewAIA3202823.

[3] Vellameeran, F. A., & Brindha, T. (2021). An integrated review on machine learning approaches for heart disease prediction: Direction towards future research gaps. Bio-Algorithms and Med-Systems, 0(0). https://doi.org/10.1515/bams-2020-0069

[4] Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8

[5] dey, A., Rajkumar, R., & Masih, J. (2019, April 29). Analytical study and visualization of human activity and content based recommendation system by applying ML Automation. https://www.researchgate.net/ publication/332082239 _ Analytical _ study_ and_visualization_of_human_activity_and_content_based_recommendation_system_by_applying_ML_ Automation

[6] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1 16, 2019.

[7]Y. Hasija, N. Garg, and S. Sourav, "Automated detection of dermatolog ical disorders through image-processing and machine learning," in 2017 International Conference on Intelligent Sustainable Systems (ICISS), 2017, pp. 1047–1051.

[8]Marouane Fethi Ferjani. (2020, December 16). Disease Prediction Using Machine Learning. https://doi.org/10.13140/RG.2.2.18279.47521

[9] Lambay, M. A., & Mohideen, S. Pakkir. (2022). A Hybrid Approach Based Diet Recommendation System using ML and Big Data Analytics. https://doi.org/10.21203/rs.3.rs-1044422/v1

[10]Kiran P, Swathi A, Sindhu M, Manikanta Y, & Mahesh Babu K. (2022). Effective heart disease prediction using hybrid machine learning technique. South Asian Journal of Engineering and Technology, 12(3), 123–130. https://doi.org/10.26524/sajet.2022.12.49

[11] Islam, R., Sultana, A., & Mohammad Rashedul Islam. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. Journal of Electrical Systems and Information Technology, 11(1). https://doi.org/10.1186/s43067-024-00150-4

[12]Chen, J., Li, K., Rong, H., Bilal, K., Yang, N., & Li, K. (2018). A disease diagnosis and treatment recommendation system based on big data mining and cloud computing. Information Sciences, 435, 124–149.https://doi.org/10.1016/j.ins.2018.01.001

[13] Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*, *11*(1), 141. https://doi.org/10.3390/electronics11010141