iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.66122

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

AyurJanaKosh - Healthcare Chatbot Using LLM

Kshitij Meshram[1], Swanand Kulkarni[2], Akhilesh Kulkarni[3], Maneet Kour Sudan[4]
Department of Computer Engineering Vishwakarma Institute of Technology, Pune, 411037, Pune, Maharashtra, India

Abstract: ***AyurJanaKosh is an innovative software platform designed to revolutionize personalized Ayurvedic healthcare by Large Language Model. This platform provides personalized drug and formulation recommendations based on users' Prakriti (constitution) and Dosha, while also considering medical history. By using Langchain and a standardized data schema, AyurJanaKosh ensures efficient semantic search and personalized health insights. The results demonstrate significant advancements in personalized healthcare, merging ancient Ayurvedic wisdom with modern technology for holistic wellness management. This project addresses a critical gap in personalized Ayurvedic solutions, offering substantial improvements in holistic health practices.***

Keywords: Ayurveda Science, Chatbot, Healthcare, Natural Language Processing (NLP), LLM, Lang chain, Semantic Search, Vector Search.

I. INTRODUCTION

Ayurveda, the ancient system of medicine from India, emphasizes a holistic approach to health and wellness, focusing on the mind, body, and spirit. Central to Ayurvedic practice are the concepts of Prakriti and Dosha, which define an individual's unique constitution and imbalances, respectively. Personalized recommendations based on these principles are essential for effective health management, yet modern healthcare systems often fail to integrate these principles, resulting in generic and non-personalized solutions that do not fully cater to individual health needs.

This gap in personalized healthcare underscores the need for a comprehensive software platform that leverages advanced technologies to deliver tailored Ayurvedic care. AyurJanaKosh is proposed to address this need by integrating Machine Learning and Large Language Model to provide personalized health recommendations based on users Prakriti, Dosha, and medical history. The primary objectives of AyurJanaKosh are to offer drug and formulation recommendations according to user's specific Ayurvedic profiles, create a recommendation system informed by medical history, and develop an effective, user-friendly web application. Harnessing the power of ML and LLM AyurJanaKosh aims to revolutionize personalized Ayurvedic healthcare, bridging the gap between ancient wisdom and modern technology. This platform ensures accurate, relevant, and personalized health insights, making holistic wellness more accessible and effective. AyurJanaKosh not only integrates traditional Ayurvedic principles with cutting-edge technology but also sets a new standard for personalized health management, promoting deeper understanding and tailored wellness solutions for individuals.

II. LITERATURE REVIEW

The study in [1] A recent study validated a 28-characteristic questionnaire using machine learning and Ayurvedic experts' opinions. The questionnaire had high Cronbach's alpha coefficients for Vata, Pitta, and Kapha. Data from 807 healthy subjects aged 20-60 was used for machine learning experiments. Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), and Decision Trees were used as classifiers. The boosting algorithm CatBoost resulted in a maximum precision of 0.96, recall 0.95, F-score 0.95, and slightly lower accuracy.

In a study presented in the paper [2], the project aims to create an Ayurvedic drug recommendation website that integrates traditional Ayurvedic knowledge with technological advancements. The platform will cater to students and practitioners, offering drug prediction algorithms and herbal drug recommendations. The project involves analytical planning, research, and user input to ensure a user-friendly interface. The front-end development will focus on responsive design, while advanced algorithms will enhance drug prediction accuracy. The platform will undergo rigorous testing to ensure quality and functionality, with the beta version being launched.

The study in [3] project aims to create an Ayurvedic drug recommendation website that integrates traditional Ayurvedic knowledge with technological advancements. The platform will cater to students and practitioners, offering drug prediction algorithms and herbal drug recommendations. The project involves analytical planning, research, and user input to ensure a user-friendly interface. The front-end development will focus on responsive design, while advanced algorithms will enhance drug prediction accuracy.

©IJRASET: All Rights are Reserved

The platform will undergo rigorous testing to ensure quality and functionality, with the beta version being launched.

The study in [4] project aims to classify an individual's Prakriti based on Vata, Pitta, and Kapha (VPK) pulses using machine learning and hardware techniques. The system uses a front-end amplifier, high-speed programmable ADC, and a PIC microcontroller. Custom software captures pulse signals and encodes them for analysis. An Artificial Neural Network (ANN) is trained to classify features, blending current technologies with ancient Ayurvedic practices. This system enhances the efficacy of ayurvedic diagnostics.

The study in [5] shown Large Language Models (LLMs) have shown potential in clinical practice, but have limitations in understanding medical images containing diagnostic information. This paper presents a framework connecting LLMs and Computer-Aided Diagnosis (CAD) networks for medical images. LLMs enhance CAD output, including diagnosis, lesion localization, and reportage, by putting CAD data into coherent sentence structure and language style. This approach combines logical and medical reasoning with vision understanding features of medical-image LLMs, aiming to improve patient approaches and enhance vision-based C-A-D models in medical imaging.

The study in [6] Paper analyzes churna datasets using two models: Pointwise Mutual Information (PMI)-Based Model and Graph-Based Louvain Method. The results show effective and reliable evaluations, highlighting synergistic and antagonistic relationships of churna ingredients and community structure frameworks. This method demonstrates how digital computing and Ayurveda can enhance traditional medicine's efficacy.

The research by [7] has shown that the use of AI Medical Chatbot can decrease overall healthcare costs through patients suggesting adequate use or overuse to the bots while they await the doctor. The Chatbot employs an AI that processes language and interacts with the patients in a more sophisticated way in terms of word combinations. N-grams, TF-IDF, and cosine distance measures are some of the techniques that help score the sentences and retrieve the best possible answers. This method increases access to healthcare and lessens the pressure on practitioners which in turn ensures that the overall expenses for quality healthcare are decreased.

According to the research conducted by [8], patients have the ability to use Artificial Intelligence medical Chatbot which can help decrease escalation of medical costs by recommending use and abuse of resources while waiting to see a physician. A natural language processing framework is integrated into the system which enables the machine to speak to the patient with a level of sophistication depending on how the patient speaks to it. Some of the techniques employed include n-grams, TF-IDF and cosine distance measures to rank the sentences and provide the best relevant answers. Such a system would reduce the cost of quality healthcare as it improves the affordability of the service and reduces the dependence on physicians.

The research in [9] The AI Health Care Chatbot system intends to assist users based on their understanding of some medical terminologies and in explaining for a particular disease. It encompasses an NLP aspect as it seeks to comprehend people's language around symptoms and health complications, their severity, among others. The bot can recommend conventional, Ayurvedic and Homeopathy medicine, depending on the situation of the user. It stores user information in a database which allows it to search for texts and provide optimal solutions.

The research in [10] explains Ayurveda, which is a holistic approach to medicine, suffers from being an intricate and lengthy process because of few practitioners to reach out to. The Prakriti Chatbot which is one of the most sought after AI solutions uses a straightforward method to determine a person's doshic balance. Its AI applications provide nutritional and life style recommendations, all targeted towards improving health with respect to one's Prakriti. The bot or the AI also aids users in answering their queries related to Ayurveda and this combination helps in better understanding of age-old techniques and their enhancements.

III. DATASET

Two key datasets are of significance for this project: the Ayurveda Texts and the Prakriti Analyzed Data which are essential to the creation of the AI Enabled Ayurvedic Chatbot.

1) *Ayurvedic Texts:* The first dataset has been developed in two phases. The first phase consisted of 15 Ayurvedic texts with more than 10,000 pages covering various topics, such as diagnostics, remedies and lifestyle recommendations. The majority of these Ayurvedic texts are in English language. This provides the required comprehensiveness necessary to pre-train the large language model (LLM) so that the company's Chatbot responds in an Ayurvedic manner. This in turn prevents the Chatbot from providing health advices which are not Ayurvedic focused.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

2) *Prakriti Analyzed Data:* This is the second dataset sourced from Kaggle, has 30 features capturing aspects of the anatomy, psychology and emotional well-being. These features capture an array of attributes such as body size and skin texture, hair color, protection and nourishment of the digestive system and that of the mind and consciousness. Such characteristics are useful in assessing an individual's Prakriti which is the body makeup of a person according to Ayurveda.

The final purpose of these datasets is to assist in the development of an artificial intelligence that will provide an all-encompassing Ayurvedic experience from the beginning to the end in a customized manner.

IV. BACKGROUND

Lang Chain is a very robust framework aimed at creating applications that make use of large language models (LLMs) together with external tools, data sources or workflows. Here's a breakdown describing how the text is first extracted and then later on used to generate an answer:

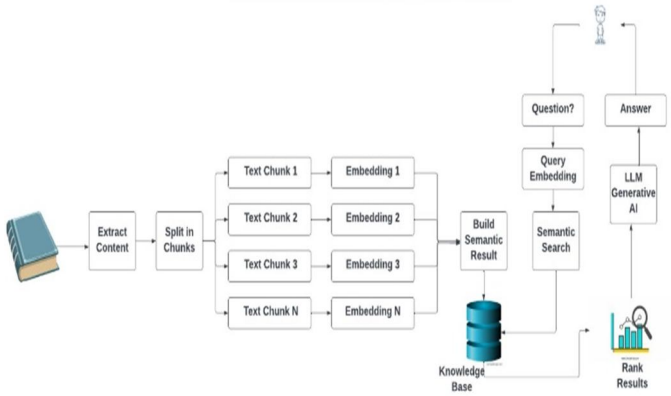


Fig.1 (Workflow of Lang Chain)

1) *Extracting Content*

The process starts with **extracting content** from a source, such as a document or a collection of texts (e.g., PDFs, articles, books). This is generally done using the text extraction tools, which convert the raw data (like a PDF) into machine-readable text. This content includes paragraphs, sentences, or other textual data, but it is not yet structured for use in a search or retrieval system.

2) *Splitting the Text into Chunks*

Once the text has been extracted, then we **split it into smaller, manageable chunks**. This is an important step for multiple reasons:

- Efficiency: By breaking a document into smaller chunks instead of processing a whole document at once, the system can process, search text and rank it more effectively.
- Context: Each chunk has sufficient context (sentences or small paragraphs) to potentially make sense on its own, improving search result relevance.
- Common techniques for splitting text into chunks include:
- Sentence or Paragraph Level: Split based on sentence boundaries or paragraph breaks.
- Fixed-Length Chunks: Split into fixed-sized chunks (e.g., 200-500 words).
- Topic-Based Chunking: Segmenting based on content similarity or topics, often achieved using more advanced NLP techniques.

Mathematically, this would be represented as:

$$T = \{T_1, T_2, T_3, \dots, T_N\}$$

Where T represents the original text, and T_i denotes the individual text chunk.

3) Embedding Text Chunks

For each of the text segments T_i, it is transformed into an embedding representing something like a probability in a hyper-dimensional vector space. Embedding's are the key since they retain the semantic meaning of the text; hence the model can compare chunks with respect to what it contains.

Embedding's are created with the help of pre-trained models such as Sentence-BERT, OpenAI's GPT, or other such models. This basically passes each chunk through an embedding function:

$$E_i = f_{embed}(T_i)$$

Where f_{embed} is the embedding model that transforms the text chunk T_i into a vector E_i.
These embedding's are representations of the meaning of the chunk, which enables semantic comparisons, rather than comparisons based on keywords. This is crucial for the system to grasp not merely precise words, but the correlations between them.

4) Building the Knowledge Base

These chunks of text are embedded and stored in a knowledge base or a vector database as the resulting vectors E1, E2, E3, …, EN . A knowledge base that supports efficient storage and retrieval of semantic data. The knowledge base can be queried during the search phase to find relevant chunks based on their semantic content.

Mathematically, the knowledge base can be represented as:

$$K = \{ E_1, E_2, E_3, \dots, E_N \}$$

Where K is the knowledge base consisting of the embedding of all text chunks.

5) Query Processing

When a user submits a question, the system encodes the query Q into the same vector space as the text chunks. You do this with the same embedding model:

$$E_Q = f_{embed}(Q)$$

Now, the query is expressed as a vector EQ in the same high-dimensional space as the text chunk embedding's. This allows for an efficient comparison between the query and knowledge base, such that relevant content can be retrieved promptly. This approach is crucial for building effective information retrieval systems that rely on semantic meaning rather than keyword matching

6) Semantic Search

The querying (or chunk matching) process consists of embedding the query into a Euclidean space and performing a semantic search between the embedded query EQ and the set of chunk embedding's in the knowledge. This is often done with cosine similarity, which measures the angle between two vectors. If we measure the cosine of the angle between these two vectors, then a smaller angle means that there is a high similarity. Cosine similarity is computed as:

$$Similarity(E_Q, E_i) = E_Q, E_i / (||E_Q|| ||E_i||)$$

Where · represents the dot product and ‖·‖ denotes the vector norm (magnitude). The higher the cosine similarity, the more semantically relevant the chunk is to the query.

7) Ranking the Results

The similarity scores for all the chunks are calculated, leading to ranking those in descending order of similarity. The top k most salient chunks are then retrieved from the system as relevant answers to the user's query:

$$\{T_{i1}, T_{i2}, T_{i3}, \dots, T_{iN}\}$$

where $T_{i1}, T_{i2}, T_{i3}, \dots, T_{iN}$ are the chunks that have the highest similarity to the query embedding E_Q.

8) Answer Generation (LLM)

Finally the text chunks are retrieved and passed to a large language model (LLM) together with the query, and a detailed, contextually accurate response is produced based on the relevant chunks:

$$A = f_{LLM}(Q, \{ T_{i1}, T_{i2}, T_{i3,} \ldots, T_{ik} \})$$

The LLM takes the relevant chunks T_{i1}, T_{i2}, $T_{i3, \ldots,}$ T_{ik} as input to generate a cohesive response that is semantically consistent with both the input query and retrieved context. This Allows the response to not only be relevant to the query, but also accurate as to what needs to be said, this combines both the query and the context provided by the most similar text chunks

V. METHODOLOGY

The structure of Ayurvedic Chatbot *AyurJanaKosh* covers structured means for engaging users, getting the assessment, and making recommendations to the users. The following describes the details of the process as per the flowchart:

1) Prakriti Quiz*:* Users engaging with the Ayurveda Chatbot start Prakriti quiz, which is an important factor in determining their Ayurvedic constitution. However, if we notice that a user has not attempted the quiz, then the chatbot asks him to do the quiz for answering important questions about their constitutions. Each of the three types weighs a set of questions that precisely correlate with physiological, psychological, and behavioral characteristics. The only requirement for users who completed the quiz is to restate the questions they answered the first time, and click submit.

2) Analyzing Responses: The responses submitted by users are processed in the toolkit through learning or a logic contained in the chatbot. This analysis predicts the Prakriti type of Vata, Pitta, Kapha or combinations based on the answer interpretation. Also this classification is supported with the basic principles of Ayurveda reasoning.

3) Storage of Prakriti Type: It could be seen that as soon as the Prakriti type is predicted a corresponding record is made which is retrievable anytime. This step assures that the insights provided by the chatbot will be engaging and specific for each user at the next interaction with the AI.

4) Bot Raising Question: After that, the user engages with the automated assistant to ask more health-related concerns, lifestyle, or questions on Ayurvedic practices. The chatbot presents a straightforward interface for raising any of the concerns mentioned above.

5) Queries Resolution and Searching of Information: As soon as the question is issued by the user, the chatbot searches its internal database for the appropriate information. This could be as simple as searching through the Ayurvedic texts or journals or any other database. The retrieval of the data is done through natural language processing and semantic search.

6) Explanations and Treatment Suggestions: Each of these responses is tailored according to the user's Prakriti type and each of these responses contains:

- *Explanations*: Communication regarding the user's questions is done along with requisite additional information in a very elaborate but also in a non-complicated way.
- *Treatment and Formulations*: The automated assistant recommends solutions consistent with Ayurveda. Such remedies are all natural, consisting chiefly of herbs, diet, and lifestyle modifications.

7) Doses and Precautions Guidelines: In order to avoid repercussions and inconsistencies in the implementation of the use case recommendations, the AI powered chatbot provides dosage levels, adverse effects and safety measures to minimize risks to users. This step also integrates the overall wellbeing and safety focus expected of Ayurveda practitioners.

8) Personalized Response: Responses are uniquely prepared in accordance with the user's Prakriti type thus providing the users with an enhanced experience and privacy. This customization helps users in comprehending their body composition in relation to the usefulness of the advice provided.

9) Concluding Remarks and User Satisfaction: This is followed by the AI powered chatbot sending a message recommending the user to have an active life with the use of specified measures. The user friendly conversational tone creates the desired satisfaction with the target audience while fostering interaction and trust.

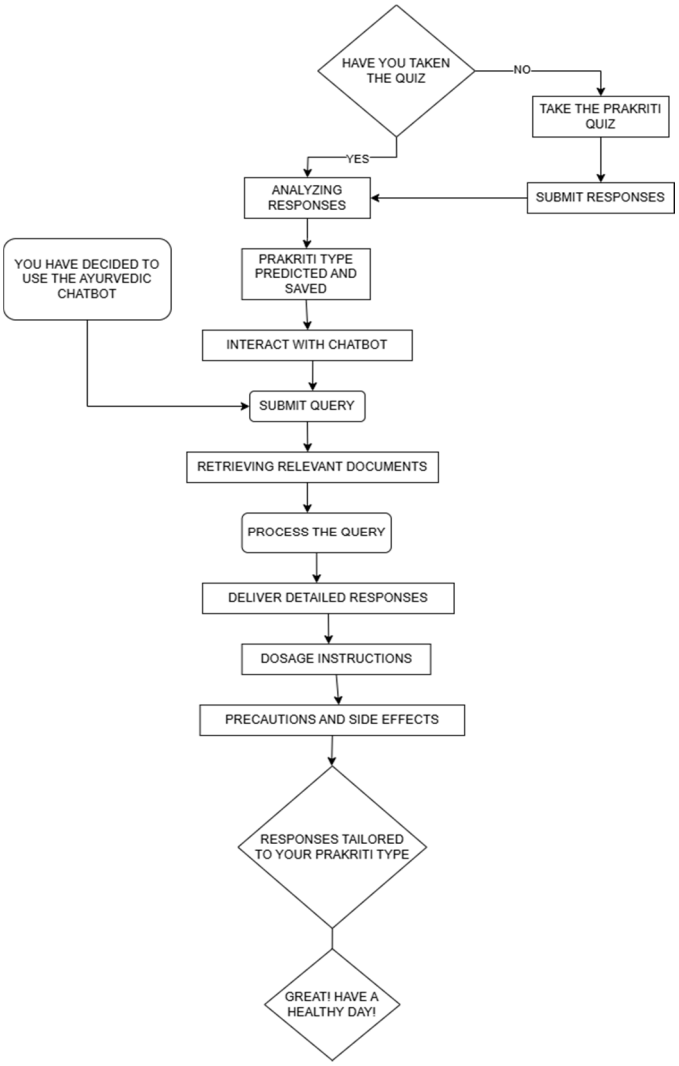


Fig.2 (Flowchart of the project)

VI. RESULT

After completing the prakriti quiz where questions are asked to determine the dosha of users' body. After the quiz, a chatbot interface opens up where user can ask for query and will get desired output according to the users' nature.

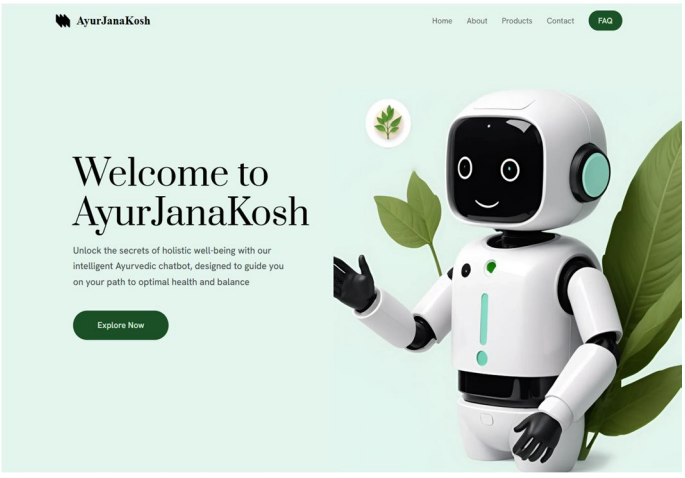


Fig.3 (Landing page)

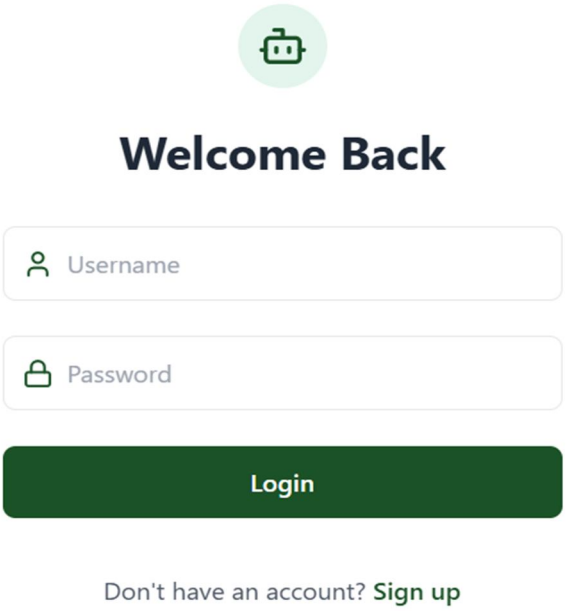


Fig.4 (Login page)

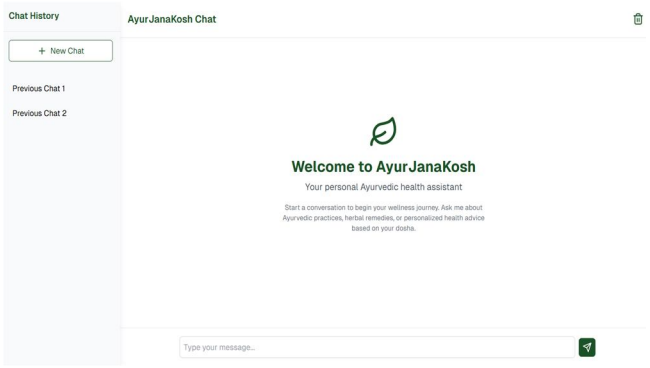


Fig.5 (Chatbot interface)

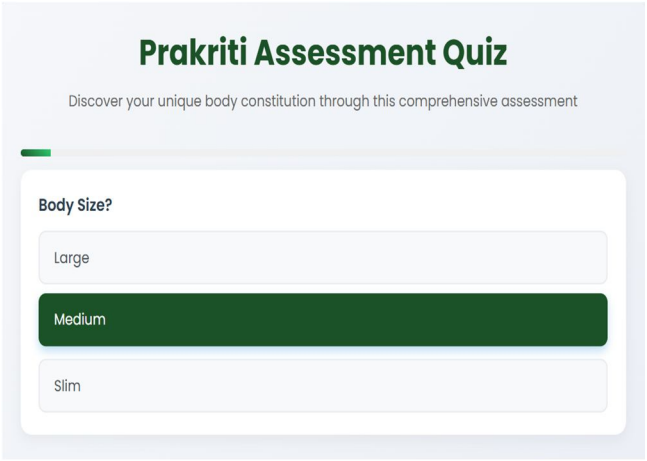


Fig.6 (Prakriti Quiz)

VII. SYSTEM EVALUATION

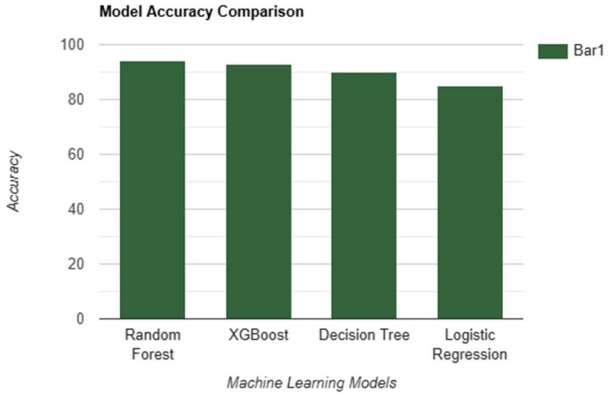


Fig.8 (Algorithm Comparison)

Model	Accuracy (%)	Precision	Recall	F1-Score
Random Forest	94	0.93	0.92	0.93
XGBoost	93	0.91	0.91	0.91
Decision Tree	90	0.89	0.88	0.88
Logistic Regression	85	0.83	0.82	0.83

Fig.9` (Performance Comparison of Machine Learning Models)

These figures compares the performance of four machine learning models—Random Forest, XGBoost, Decision Tree, and Logistic Regression—using key metrics: Accuracy, Precision, Recall, and F1-Score. The Random Forest model achieved the highest accuracy (94%) and performed well across all metrics. XGBoost followed closely with 93% accuracy, while Decision Tree and Logistic Regression showed slightly lower performance, with accuracies of 90% and 85%, respectively. This comparison highlights the strengths and weaknesses of each model in predicting Prakriti, demonstrating the robustness of ensemble methods like Random Forest and XGBoost for this task.

VIII. FUTURE SCOPE

In the implementation of this Chatbot in the future, the use of formulation recipe videos that have been AI generated through the use of Stable Diffusion Technology can be useful. Through this, the Chatbot will provide more than just text instructions but will also create video demonstrations taking users through instructions on how to prepare, create an Ayurvedic remedy, a personalized recipe or even a pharmaceutical compound. As for the **Stable Diffusion** which aims for visual purposes, the Chatbot will be able to draw video that are active to a users' health requirements as per formulation laws and recent studies.

This feature would be quite useful in medical education and patient care since a user would be able to know how to accurately prepare the medication or the treatment by simply watching it. When the procedure is complicated, videos can bridge this issue by depicting real time activities while giving out steps the user has to take to accurately follow the video.

With the incorporation of AI's video generation tools, the Chatbot could integrate videos as a full adjunct and provide practical or information provision. It would make users' learning easier as well as them engaging in a variety of activities hence widening their chances to learn. Such innovations would greatly shift user experience allowing for much more than written text prompts, thus evolving the Chatbot into an irreplaceable resource for visual healthcare personalization.

IX. CONCLUSION

In this project, we constructed an AI-based Ayurvedic Chatbot that employs machine learning and the principles of Ayurveda to provide primary health advice. In this regard, the system possesses two data sets:

The Ayurvedic Books and The Body Analysis Data. The Ayurvedic Books data set consists of more than 10,000 pages from 15 Ayurvedic books and enables the Chatbot to respond accurately to queries based on traditional Ayurvedic practices. The Body Analysis Data set assists the Chatbot in determining Prakriti of the individual by analyzing 30 physical, psychological, and emotional traits thereby enabling the design of tailored Ayurvedic therapies for the individual.

We used several machine learning models such as Random Forest, XGBoost, Decision Tree, and Logistic Regression which help to predict Prakriti and in classifying the user's inputs. Most of the models produced high accuracy level with Random forest 94% accuracy, XGBoost 93%, Decision Tree 90%, and Logistic Regression 85%. These results again illustrates the strength of the system in classifying user's data.

Overall, this project serves as an inspiration in the area of using AI and Ayurveda together to develop new Startups in the healthcare sector which can provide unique, personalized well-being advice to the user.

REFERENCES

[1] Madaan, Vishu & goyal, Anjali. (2020). Predicting Ayurveda Based Constituent Balancing in Human Body Using Machine Learning Methods. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2985717.

[2] Shevantikar, Pradnesh & Udata, Avinash & Korachagao, Abhishek. (2024). Web Application for Recommendation of Ayurvedic Drugs and Medicine using ML. Journal of Data Processing and Business Analytics. 1. 13-18. 10.48001/jodpba.2024.1213-18.

[3] H. Pogadadanda, U. Shwetha Shankar and K. R. Jansi, "Disease Diagnosis Using Ayurvedic Pulse and Treatment Recommendation Engine," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1254-1258, doi: 10.1109/ICACCS51430.2021.9441843.

[4] S. Joshi and P. Bajaj, "Design & Development of Portable Vata, Pitta & Kapha [VPK] Pulse Detector to Find Prakriti of an Individual using Artificial Neural Network," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6, doi: 10.1109/I2CT51068.2021.9418155.

[5] ChatCAD: interactive computer-aided diagnosis on medical image using large language models. (2023). https://arxiv.org/abs/2302.07257.

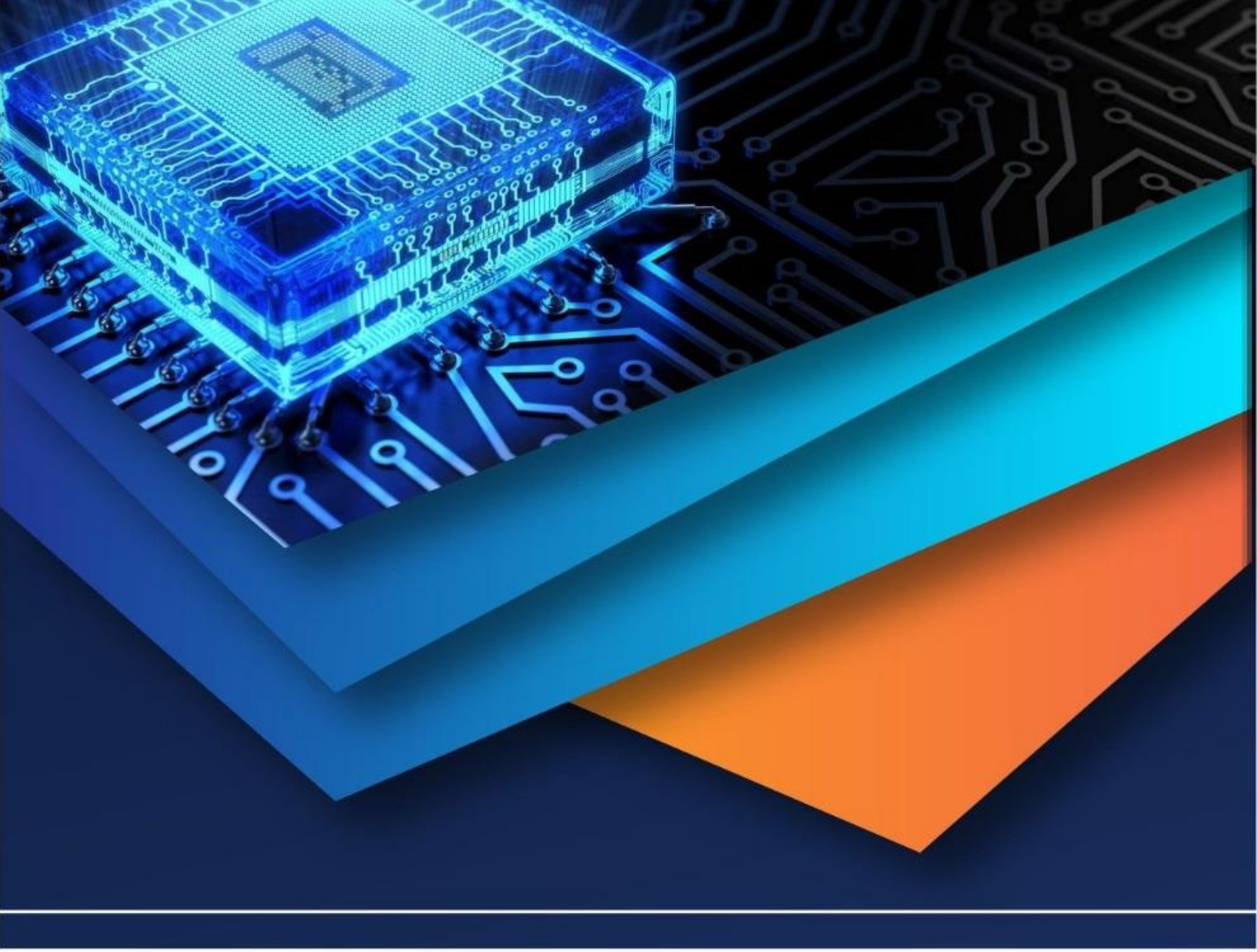
[6] H. Joshi, "Towards a recommender system for the ingredients of traditional ayurveda churna," 2013 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, 2013, pp. 1-3, doi: 10.1109/NUiCONE.2013.6780088.

[7] Athota, Lekha & Shukla, Vinod & Pandey, Nitin & Rana, Ajay. (2020). Chatbot for Healthcare System Using Artificial Intelligence. 619-622. 10.1109/ICRITO48877.2020.9197833.

[8] Arun, Asane & Balasaheb, Pangavhane & Babasaheb, Jori & Kailas, Jadhav & Adinath, Kale & Dadasaheb, Nalawade. (2024). Artificial Intelligence and Challenges in Ayurveda Pharmaceutics: A Review. Research Journal of Science and Technology. 237-244. 10.52711/2349-2988.2024.00034

[9] Ramalingam, Jegadeesan & Srinivas, Dava & Nagappan, Umapathi & Ganesan, Karthick & Venkateswaran, Natesan. (2023). Section A-Research paper Personal Healthcare Chatbot for Medical Suggestions Using Artificial Intelligence and Machine Learning Eur. 12. 6004-6012. 10.31838/ecb/2023.12.s3.670.

[10] CHATBOT TO KNOWN INDIVIDUAL PRAKRITI (PHENOTYPE)", IJNRD - INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (www.IJNRD.org),ISSN:2456-4184.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)