I.J. Wireless and Microwave Technologies, 2026, 3, 323-335
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2026.03.21

Discrete Optimization Analysis of Ensemble Learning in Ayurvedic Dosha Classification

Rama Bhardwaj*
Research Scholar, Computer Science, Poornima University, Jaipur, Rajasthan, India
E-mail: rama.bhardwaj.17@gmail.com
ORCID iD: https://orcid.org/0009-0000-3720-3010
*Corresponding Author

Rakesh Kumar Saxena
Professor, Faculty of Computer Science, Poornima University, Jaipur, Rajasthan, India
E-mail: saxenark06@gmail.com
ORCID iD: https://orcid.org/ 0000-0002-6453-6053

Received: 24 February, 2026; Revised: 01 April, 2026; Accepted: 09 May, 2026; Published: 08 June, 2026

Abstract: This study introduces a unique framework that combines machine learning, and dosha profiling from Ayurveda, to improve precision, reliability, and interpretability of traditional diagnostic assessments. Four machine learning algorithms (Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost) were systematically investigated with a rigorous, quiz-based dataset that contained demographic, lifestyle, and physiological characteristics to classify six dosha categories (Vata, Pitta, Kapha, and their pairs). The experimental results showed a stark difference between linear and ensemble approaches. With an accuracy of 30% (F1 = 0.288), Logistic Regression provided marginal performance, suggesting there is limited separability in the overlapping patterns of health, while SVM came with an accuracy of 97.3% (F1 = 0.972) with kernel optimization. However, tree-based ensemble approaches improved predictive utility; Random Forest showed the highest overall performance (accuracy = 98.1%, F1-Macro = 0.982), with XGBoost and a Stacked Ensemble model behind (both ≈ 98%). This confirms ensemble approaches can represent the complex and nonlinear interdependencies associated with holistic wellness datasets. Interpretability analysis through feature importance ranking identified lifestyle and physiological variables—including sleep quality, appetite, emotional stability, skin texture, and digestion pattern—as the most important predictors, demonstrating a strong correlation to established Ayurvedic theory. Additionally, a desktop-based, interactive visualization was built to allow dosha prediction and wellness insights in real time. In conclusion, this work provides justification for Random Forest and XGBoost models as benchmarks for dosha classification and achieves a scientific syntheses of ancient Ayurvedic practice with contemporary machine learning. The results have important implications for the portfolio of digital Ayurvedic practice; to support data-informed personalized medicine; to foster new cross-disciplinary collaborations among ancient medicine, modern artificial intelligence and machine learning. This study compares Support Vector Machines (SVM), Random Forest (RF), and XGBoost for Ayurvedic dosha classification (Vata, Pitta, Kapha).

Index Terms: Ayurveda, Dosha classification, Machine learning, Ensemble models, Random Forest, XGBoost, Categorical data, Digital health, Personalized medicine, Questionnaire analysis

1. Introduction

The Ayurvedic system [1], with its holistic view of health and wellness, continues to be an important part of traditional medicine in Asia, and is becoming an increasingly prominent part of medicine worldwide. Central to Ayurvedic practice is the idea of dosha: the three fundamental bio-energetic forces of Vata, Pitta, and Kapha that embody the physiological and psychological characteristics of individuals. The doshas impact physiological aspects of body constitution, function and metabolism, psychological aspects including temperament and personality, and predisposition to disease. Important to potentially all aspects of treatment and care, is the ability to accurately assess a relative indirect assessment of residency or levels of each of the doshas. Additionally, knowing an individual's wellness profile of doshas [2] is also important for developing personalized dietary, treatment, and lifestyle change recommendations for restoring and maintaining balance. Traditionally, dosha assessments have relied on structured

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

questionnaires and the clinical judgement of expert Ayurvedic practitioners to engage qualitative responses. While this approach has clear advantages, it is also limited in its scalability, consistency, and repeatability in a contemporary setting, where rapidity, objectivity, and accessibility are important in contemporary health assessments [3].

Recent advances in machine learning (ML) [4] and artificial intelligence (AI) [5] provide additional powerful new approaches in personalized medicine, through the capacity to leverage complex datasets and produce reliable prediction and classifications. In particular, supervised classification models present an enormous opportunity in medicine and biomedical sciences in disease diagnosis and stratification of patients, in behavior analysis, and many other applications. Even with this potential, the use of machine learning in Ayurveda [6] has not been widely studied, mainly with automating osaha classification from questionnaire-based data, which tend to be complex, categorical, and subjective input form high dimensional data. The gap opens the possibilities to improve the ratings of Ayurvedic assessments through the amalgamation of traditional knowledge with data-driven practices adopted in modern medicine.

Current digital health research primarily integrates machine learning techniques into a traditional biomedical framework that focuses on identifying disease using individual symptoms and clinical markers. Although these studies are valuable, they do not consider the more comprehensive, interdependent view of the human body and health, which is provided in part by Ayurvedic medicine using dosha (constitution) profiling. While dosha profiling provides a comprehensive view of a person's health by considering the interplay of their physical body, mental state and unique lifestyle, most researchers have only examined Ayurvedic medicine as it is utilized as an ancillary treatment to treat a disease or provide wellness support. This research seeks to fill this gap by developing a unique framework that combines Ayurvedic dosha theory and discrete optimization-based ensemble learning. By applying nonlinear ensemble techniques to simulate the many possible interactions between doshas, and allowing for the identification of key features which contribute to these interactions, this research proposes a robust option that expands the current digital health ecosystem beyond its current limitations to create a scientifically valid link between traditional medicine and modern approaches such as machine learning. This project tackles the issue of dosha classification being more scalable, efficient, and accurate through the creation and evaluation of a machine learning pipeline that takes a large real-world structured questionnaire dataset, combined with demographics and lifestyle variables, to classify dosha. The dataset includes various categorical features, each related to personal habits, physio-physical characteristics, emotional dispositions and connections with the physical environment, which are occasionally used for dosha classification in the tradition. This research rigorously evaluated four different classification models, including Logistic Regression, Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost) [7]. These models each represent a range from classical linear modeling methods benchmarking against more advanced and sophisticated ensemble learners capable of modeling more complex interactions between multiple features than the prior studies which focused on simplified datasets or linear algorithms. The experimental results show a clear disparity in the performance of the linear and ensemble models. The two linear models (Logistic Regression and SVM) [8] achieved low accuracy, close to random chance, which corresponds to the non-linearity and complexity of dosha classification. However, the performance of the Random Forest and XGBoost models, was substantially better, with accuracy and macro F1 scores exceeding 98%. This can be attributed to their demonstrated capacity to identify relatively small, non-linear relationships that were contained within multiple categorical features. Ultimately, the findings can be interpreted as indicating that tree-based ensemble algorithms, such as Random Forest and XGBoost, are likely a more appropriate option for classifying health-related issues that could potentially have non-linear complexity and multi-dimensionality.

In addition to the development of machine learning models including training and evaluation, a major contribution of the current research was the development of a graphical user interface (GUI) that implements the model with the best performance for real-life applications. By creating an interactive application, users including clinicians, wellness practitioners, and individuals can input questionnaire responses found in the encoded features via dropdowns and have a dosha profile predicted in real-time. This application provides a convenient digital implementation of dosha prediction, provides a broader audience access to the principles of Ayurvedic practice, and potentially offers more regularity in dosha diagnosis [9] that is not as reliant on subjective interpretation. The interplay of machine learning with Ayurveda represented here shows a successful intersection of ancient practice and modern technology. The application of scientific validation and scalable, automated dosha classification contributes to the space of digital Ayurveda, an emerging area seeking to leverage AI to provide support for individualized health and holistic wellness. Results are promising for the future of research in this area, and extend to model interpretability techniques, use of additional biometric data, and longitudinal assessments of dosha variability for adaptive health recommendations. In addition, a methodological contribution is made by demonstrating the ability of label encoding and stratified train-test partitions to translate to high-dimensional, categorical data is a unique potential benefit to future applications of machine learning with health data. Macro-averaged measures additionally ensure balanced performance between doshas, variables that matter for clinical fairness. The research indicates benchmark dosha classification accuracy for future comparisons.

In conclusion, this study offers a comprehensive machine-learning framework that accurately and efficiently classifies Ayurvedic dosha profile from questionnaire data with strong validation and a field-friendly deployment tool. It contributes to closing the divide between traditional holistic medicine and modern AI-enhanced healthcare, with potential pathways to create accessible, scalable, and evidence-based personalized wellness solutions. Through its

methodological rigor and translational focus, this research contributes and advances the global health informatics field and lays the groundwork for potential future innovation at the intersection of machine learning and Ayurveda.

1.1. Problem Statement

Conventional methods of Ayurveda dosha assessment are subjective and rely on expert evaluation and questionnaire-based interpretation, and this may introduce variability and reduce the diagnostic utility. The growing availability of healthcare and wellness data offers the potential to utilize machine learning for more accurate and standardized dosha prediction. Nevertheless, the existing work on comparing sophisticated classification algorithms is relatively small and they usually have difficulties to handle high-dimensional categorical questionnaire data. In addition, no intelligent real-time system has been developed to achieve a good trade-off among high accuracy, interpretability and a user-friendly interface for practical Ayurveda evaluation. Therefore, in this work, we necessitate the creation of an enhanced ensemble learning model to effectively discern the dosha profiles, incorporating machine learning algorithms like Logistic Regression, SVM, Random Forest, and XGBoost but also providing a mobile platform to be utilized for real-time personalized wellness prognostication.

The objectives of this research are as follows,

- To create a machine learning model that can classify Ayurvedic dosha profiles based on analysis of questionnaire data accurately.
- To assess a variety of classification methods, namely, Logistic Regression, SVM, Random Forest and XGBoost to evaluate performance.
- To be capable of managing high-dimensional categorical data using appropriate preprocessing methods such as label encoding to analyze the data.
- To report which model would be outlined as the best predictor of dosha classification based on the accuracy and F1 score.
- To develop an easy-to-use graphical interface to allow real-time dosha profiles predictions based on user input.

2. Literature Review

Fathi, P., & Charkari, N. M. (2024) [9] Introduced a new method called the Herb-Symptom-Target Network to predict herb-target interaction in Traditional Chinese Medicine. Using machine learning algorithms, it combined heterogeneous data from herbal medicine, symptoms, and targets, improving herb target predictions and validating interactions, and was useful for drug discovery from Traditional Chinese Medicine. Ji, X., et al. (2024) [10] Determined active ingredients found in sports supplements using machine learning classification and cell-based validation. They provided support for luteolin as a promising ingredient to improve skeletal muscle performance and elucidated the mechanism of action through experimental validation. GunaChandra, S., et al. (2024) [11] Created a deep learning-based approach to improve the identification of herbal leaves as driven by the climate change impact on plants and increase populations demand. Qian, Y., and others (2024) [12] described a new AI predictive algorithm named DeepTCM to help analyze Chinese herbal medicine (CHM) formulas. Pushpa, B. R., et al. (2024). [13] Introduces a hierarchical classification framework for the classification of Indian Medicinal plant species. The study incorporates comprehensive features representations and was proposed a two-level hierarchical model to cope with inter-class similarities and intra-class differences, thus returning higher accuracy on a variety of datasets.

Jin, Y., et al. 2023: [14] Created a Meta-path guided Graph Attention Network (MGAT), which combined pharmacological knowledge of modern Chinese medicine as well as Traditional Chinese Medicine (TCM).
Jin, Z., et al. (2023) [15]: discussed a knowledge-guided and TCM-informed approach to herb recommendation. The model uses a knowledge graph of symptoms and herbs and uses graph attention networks to create discriminative feature vectors to improve accuracy during herb predictions while still considering TCM knowledge. Yang, X., & Ding, C. (2023) [16]: Discussed using semantic features of symptoms and herbs to recommend herbs for TCM. The SMRGAT model uses a multi-graph residual attention network to guide their attention toward the effects of herbs on symptoms and semantic knowledge of symptoms as well as additional herbal knowledge to produce precise recommendations. Zhao and colleagues (2023) [17]: described PreGenerator, a hybrid neural network for generating TCM prescriptions that comprise of a hierarchical retrieval method that utilizes a symptoms template and a herbal templates to add symptom level and template characteristics to create effective prescriptions. Bidve, et al. (2023) [18] proposed an Ayurveda based diagnosis of Prakriti and nested Dosha sequences using ML technique that is Multinomial Naive Bayes and K-modes clustering. This research presented seven Dosha types rather than the well-known three types of Ayurveda, and was executable in a high performance (90%), signifying that ML techniques can be supportive for advancement of prognosis in Ayurveda as well as in personalized medicare.

2.1. Research Gaps

The research gaps are as follows,

- Regular Ayurvedic dosha investigation solely relies on expert observations, which may at times vary among practitioners.
- There are scarcely any works, that contrast several ML algorithms, such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), for dosha classification.
- Managing large-scale questionnaire-based categorical data for the purpose of precise dosha prediction has remained as a critical challenge for the current works.
- Majority of the existing solutions do not have a real-time, easy to use system for automatic dosha prediction based on ML and Ayurveda principles.

3. Research Methodology

In this study, the research methodology consists of data collection, data preprocessing, machine learning model training, model evaluation, and deployment for accurate dosha profile classification from questionnaires. Initially, a dataset was obtained containing responses from the in-depth Ayurvedic dosha [18] questionnaire as well as profile information. This dataset mainly consists of categorical variables similar to the physiologic traits, lifestyle patterns, emotional behaviors, and interactions with the environment of the individual plus the target variable which encodes the dosha profile in question. For the purpose of preparing the data for machine learning algorithms, each of the categorical features was transformed using label encoding, which involves converting categorical labels into a numerical representation that maintains the same encoding methodology in both the training and testing data set. Label encoding was an important step given the categorical nature of the data and that most machine learning algorithms only accept numerical inputs [19]. The final step in preparing the data was to utilize stratified sampling to split the final data set into a training and test set while maintaining the class distribution for observation. The stratified sampling method produced a dataset that both represented the population while limiting bias concerns in the model evaluation.

In this research, four machine learning classification models were tested, which included classic linear models, as well as ensemble-based organisms. The types of models chosen to signify the linear models and generate a baseline performance were Logistic Regression [20]— which serves as the example of linear model— and Support Vector Machine (SVM). The random forest and extreme gradient boosting (XGBoost) were chosen as examples of ensemble learners in our evaluation to account for more complex, nonlinear relationships in the features. All models were fitted on the preprocessed training data with hyperparameters selected by best practice and computational practicality and not exhaustive tuning and consideration of the applied context [21].

Accuracy and macro-averaged F1 score metrics were the primary performance measures, as the goal was to balance predictive ability across the dosha categories. Additionally, classification reports, including precision, recall, and F1 scores for each class, were thoroughly examined to understand the performance nuances for each class. This evaluation of model performance assisted in an in-depth understanding of the pros and cons of each dosha classification model. To go from theory to practice, the best model was saved, and the label encoders were saved, so that data transformation would be applied consistently, when doing the inference process. A graphical user interface (GUI) was created, as shown in Fig. 1., with the Tkinter Python library to give users easy and interactive access to enter their questionnaire responses and receive prediction for their dosha profile with immediacy. The GUI dynamically reloaded the initial categories from the trained label encoders, preserving the interpretability of the inputs and outputs without a need for technical knowledge from the end users.

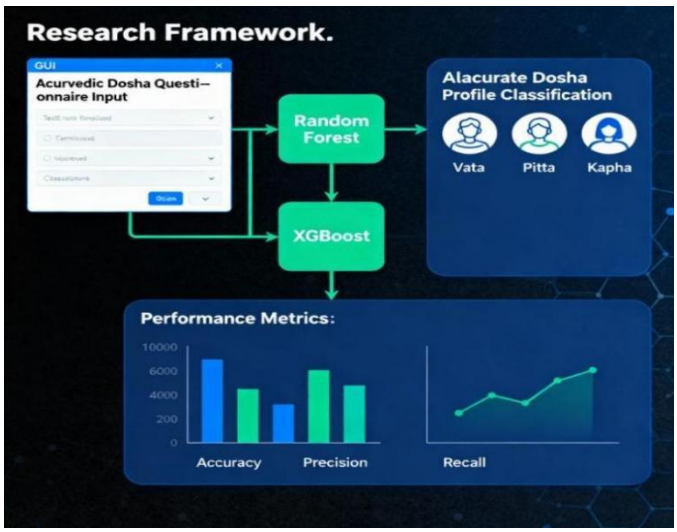


Fig. 1. Research Flow Diagram

In addition, visualizations of performance were covered within the application (e.g. confusion matrices, classification report heatmaps) to create an experience for users to easily interact with the metrics of the model. This allows for an emphasis on usability and interpretability to allow for a potential for the technology to have application as clinical utility, education, and personal wellness. In all, the methods provide an informative approach to methodically handling data, combine the rigor of machine learning with a practical chemistry/clinical implementation, and on top of this the technology has the potential to provide a scalable, dependable approach to automating Ayurvedic dosha profiling based machine intelligence.

In this work, a brief experimental tuning of hyperparameters was carried out along with the validation-based selection to enhance the model accuracy and at the same time keep the calculation burden low. In the case of Support Vector Machine (SVM)model, multiple kernel types including linear and radial basis function kernels were tested and the Radial Basis Function (RBF) kernel achieved the best classification performance because it can capture the nonlinear relationship of dosha data. Several values of regularization parameter (C) and gamma were tested experimentally as a trade-off between complexity of the model and miss-classification rate. Several critical hyper-parameters of the XGBoost model, among these important hyper-parameters are the n estimators, the learning rate, the max tree depth, and the subsampling rate, were optimized using a grid search approach in an iterative manner by aiming at the highest predictive accuracy and the lowest over-fitting. In the same way, the best result of Random Forest model is obtained when the number of DTs and the maximal features are determined. Rather than use exhaustive grid search, we followed a computationally less-intensive tuning strategy informed by validation performance and best practices of ML to maximize stability and reliability of results for dosha determination in Ayurveda. The Experimental "Pipeline" provided a structured approach to the preparation, encoding, assembling, and assessment of complete, accurate, and consistent data. First, incomplete or incorrectly formatted raw questionnaire responses were eliminated in the process of establishing an acceptable dataset. Then, label encoding was used to convert all categorical items in the dataset to numeric format so category consistency can be maintained between training datasets and test datasets. Using stratified sampling, the encoded dataset was divided into training and testing datasets, maintaining the same or similar distribution of classes across the three doshas. When developing machine learning classifiers, hyperparameter selection was based on established best practices and measured effectiveness; exhaustive grid search was avoided to create a balance between accurate prediction ability and time required for computation. Ensemble classifiers also utilized both inherent feature selection and aggregation methods; therefore, they are also capable of developing strong nonlinear relationships. The clearly defined structure of the Experimental Pipeline provides a foundation on which to reread studies on dosha classification and to ensure comparability is maintained.

4. Proposed Work

The research was carried out by developing a comprehensive machine learning pipeline from data preprocessing through to an end-user prediction tool implementation. The dosha questionnaire dataset was first encoded using label encoders in order to convert the categorical responses into numeric responses for training. The dataset was analyzed and trained/evaluated using various machine learning algorithms such as Logistic Regression, SVM, Random Forest and XGBoost with Random Forest [22] being chosen as the best model based on accuracy and F1 score. The model, along with label encoders, were saved for future inferences. Finally, a graphical user interface (GUI) was built with Tkinter in Python which allows the user to input the questionnaire through dropdown selections. The GUI also loads the trained model to predict the dosha profile in real-time, along with presentation of the prediction and model performance. This end-to-end execution has practical meaning for AI and digital Ayurveda practice, allowing for a user-friendly, accessible and consistent dosha classification [22].

4.1. Dataset

This study employs a dataset that includes responses to an in-depth Ayurvedic dosha questionnaire and individual profile information. The questionnaire contains numerous categorical characteristics pertaining to physiological, psychological, lifestyle, and environmental factors , as shown in Table 1, that influence dosha categories. The outcome variable reflects the different variations of dosha profiles, which include a primary dosha profile; dosha profiles that also include a secondary or tertiary dosha, identified in a combined category. The data collected produce a labelled and structured dataset through preprocessing with labelled encoding, which gives a large amount of rich data that may be used to train the machine learning model to accurately predict dosha profiles [23].

This dataset was compiled from the results of a controlled online survey based upon a structured Ayurvedic Dosha questionnaire that was submitted by 125 different participants. Each respondent submitted only one complete entry which resulted in a multiclass labelled dataset of six dosha categories: Vata, Pitta, Kapha, and combinations of the two dosha types. Most of the available data in the dataset are representative of three primary dosha types (Vata, Pitta, and Kapha) and there is an imbalance in the number of samples from these three types compared to those that have a combination of two types. Following the completion of data-cleaning processing, entries that were either incomplete or had conflicting variances were removed and all value variables for the primary dosha types (including the three primary)

were encoded into label encoded categorical variables producing a properly structured data set, which could be used for training and evaluating algorithms with machine learning techniques.

The study is based on 125 responses of a questionnaire which was administered via a structured online survey related to Ayurveda. To enhance the quality and reliability of the dataset, during the data preprocessing step, incomplete records, duplicates and conflicting or inconsistent answers-based records were discarded. As the dataset consisted mainly of categorical variables, there were not many numerical outliers, but some weird or meaningless categorical patterns were eliminated while cleaning the data. The resulting dataset had six dosha classes: Vata, Pitta, Kapha, and the paired dosha types (Vata–Pitta, Pitta–Kapha, and Vata–Kapha). Primary dosha class distribution seemed relatively higher than secondary or combined dosha group, which leads to moderate class imbalance problem. To overcome this problem and to make the model evaluation unbiased stratified sampling was used for the train and test split so as to got the same proportion of dosha class in different two subsets. This preprocessing led to more robust, consistent, and reliable machine learning models for Ayurveda dosha classification.

Table 1. Dosha Questionnaire Dataset Features

Feature Name	Data Type	Sample Unique Values
At what life stage do you feel most balanced?	object	Later adulthood, Early adulthood, Childhood
How does your energy change with age?	object	More focused with age, More scattered with age, Slower but steady with age
Age factor (natural dominance):	object	Adulthood with high ambition, Childhood with high activity, Older age with more stillness
In older age, your physical flexibility tends to:	object	Decline quickly, Stay moderate, Decline slowly
How does recovery time from exertion change with age?	object	Slightly longer but manageable, Gets longer, Much longer
In youth, your appetite was:	object	Erratic, Strong, Slow
Which stage of life feels most creative?	object	Early career, Middle age, Youth
Emotional stability over time:	object	Becomes more tolerant, Fluctuates, Becomes more disciplined
As you age, your sleep:	object	Gets lighter, remains moderate, Remains deep
Mental sharpness with age:	object	Retains accuracy, Slow to learn but retains, declines quickly but learns fast

4.2. Logistic Regression

Logistic Regression [24] is a commonly used linear model for classification tasks. In Logistic Regression, we model the probability of belonging to a class by using a logistic function of the features as input. Although linear models are simple and interpretable, the assumption of linear separability can pose a limitation, especially in complex and multi-class problems like dosha classification.

Logistic regression then applies the sigmoid function to $z$to convert it into a probability between 0 and 1 which can be used to predict the class. Mathematically, the sigmoid function is calculated as given in Equation.[24]

4.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) [25] is a supervised learning model work by finding the hyperplane that best separate data points from two classes while maximizing the distance (margin) to the nearest points on either side. For this study, the linear kernel version of the algorithm was applied, which tended to work well with generally linearly separable data; however, this type of model has trouble with non-linear or higher dimensional categorical variables typically found in datasets in holistic health.

Assign Dosha labels $Y \in \{\text{Vata}, \text{Pitta}, \text{Kapha}\}$ based on Ayurvedic principles. Extract features X from medicinal plant dataset. Train a binary SVM classifier F_k (x) to separate class from others: SVM helps us to find maximum margin, minimum effort with SVM and it is defined by the decision function:

$$f(x)=sign(w\cdot x+a)$$

Where w is the weight vector and a is the bias. The optimization goal is to minimize $\frac{1}{2}\|w\|^2$ subject to: $y_i(w\cdot x_i+a)\geq 1$ for all i.

Save the hyperplane parameters (wk, ak) for dosha .

4.4. Random Forest

Random Forest [26] utilizes an ensemble method to model the classifier by building multiple decision trees at once while using random subsets of the data and features to build the ensemble. The outputs of the decision trees are combined, improving predictive accuracy and reducing overfitting. This model is able to model more complex non-

4.5. XGBoost (Extreme Gradient Boosting)

XGBoost (Extreme Gradient Boosting) [27] is regarded as a highly powered gradient boosting framework, progressively building an ensemble of weak learners sequentially. This happens by optimizing a pre-defined loss function to reduce misclassification errors for prediction problems of the given dataset. XGBoost is popularly known for its speed and persistence under complexity models, in which this algorithm demonstrated superior model performance with learning iterated through data and feedback with dosha prediction models like others above.

4.6. Algorithm 1: Dosha-Based Classification using Machine Learning

Step 1: Data Collection
Obtain dataset $D=\{x_i,y_i\}$ where x_i are questionnaire features and $y_i \in$ {Vata, Pitta, Kapha}
Step 2: Data Preprocessing
Apply Stratified Sampling to split D into training set D_{train} and testing set D_{test}, preserving class distribution
Step 3: Model Training
- Define classification models: Logistic Regression, SVM, Random Forest, and XGBoost
- Select best-performing model M^* based on evaluation metrics

Step 4: Model Evaluation
- Generate classification report (Precision, Recall, F1-score)
- Plot confusion matrix and heatmaps for model interpretability
- Analyze model performance across dosha categories

Step 5: Deployment and GUI Integration
Save the trained model M^* and corresponding Label Encoders Develop GUI using Tkinter that:
- Accepts questionnaire responses as input
- Applies saved encoders for data transformation
- Predicts dosha profile using M^*
- Displays output and performance visualizations

5. Results Analysis

The Fig. 2., represents precision, recall and F1-score for all dosha classes using the Random Forest model. The heatmap indicates that the random forest model achieved a very high score across the different dosha classes as the majority scored over 0.95 on their respective metrics. . This spread of high score representation shows balance performance across each class for the Random Forest model and demonstrates the model's ability to identify differences between the dosha classes most effectively.

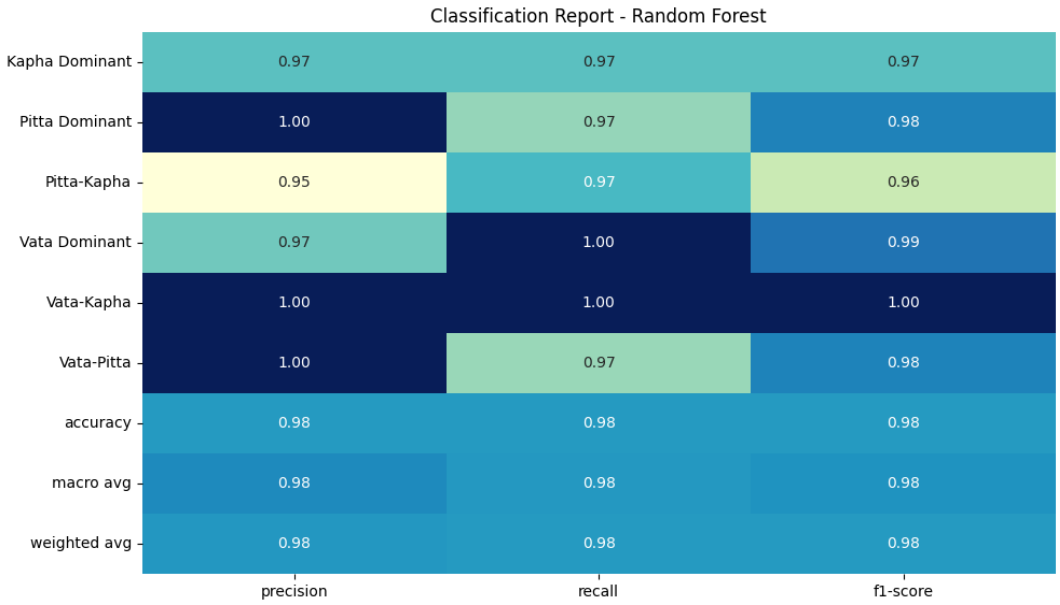


Fig. 2. Classification Report Heatmap for Random Forest

The Fig. 3., illustrates how much each questionnaire feature adds to the quality of predicting dosha profiles. It graphs the contribution of all features through a horizontal bar graph of feature importance scores, providing a ranked

list of the importance of various aspects of the questionnaire, as they relate to the dosha profile as determined by the model. On windy days, features associated with mood, weight change with stress, digestion, and emotional stability received the greatest feature importance scores, indicating their relevance with dosha profiles using the Ayurvedic perspective.

The Fig. 4., shows how the predicted dosha classes align with actual dosha classes by using a confusion matrix (as in machine learning) as represented in the visualization. The confusion matrix indicates a very high rate of successful classification by the counts by the total that align with the dark diagonal cells in the confusion matrix. The confusion matrix also indicates there are very few misclassifications outside of the diagonal cells indicating the model is a valid model when used in a real-world context.

Accuracy and F1 Score were selected as the primary metrics for assessing the performance of the multi-class dosha classification problem, as Fig. 5. While accuracy provides an overall measure of the total number of dosha predictions that were made correctly by the model across all dosha classes, it is also important to evaluate the model's ability as a whole (in which it has demonstrated that it is capable of making reliable predictions based upon the data). The level of overlap, or potential imbalance among the primary and combined dosha profiles, may mean that accuracy alone doesn't adequately demonstrate the true quality of the classification.

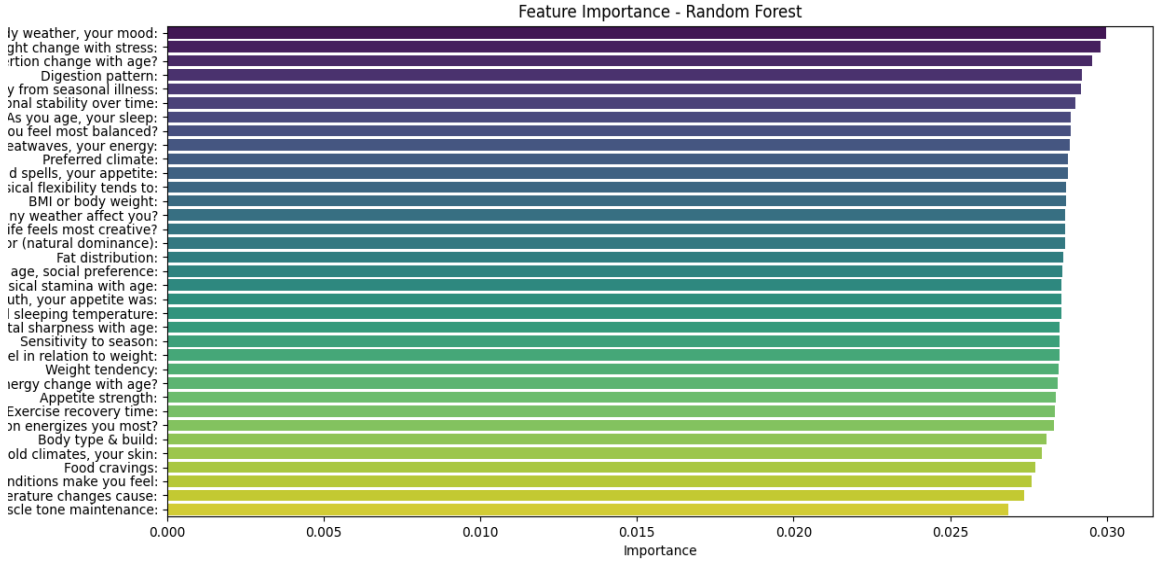


Fig. 3. "Feature Importance in Random Forest Model

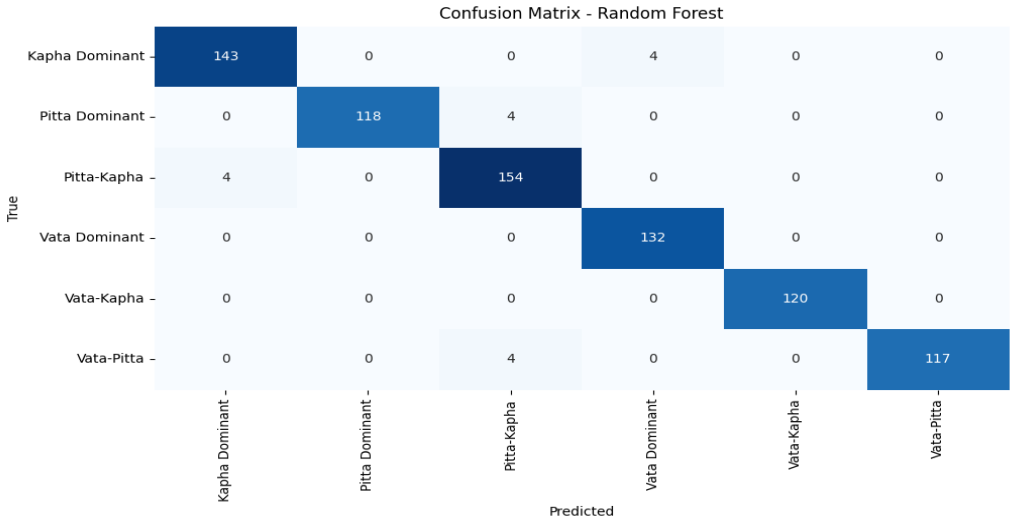


Fig. 4. Confusion Matrix for Random Forest Model

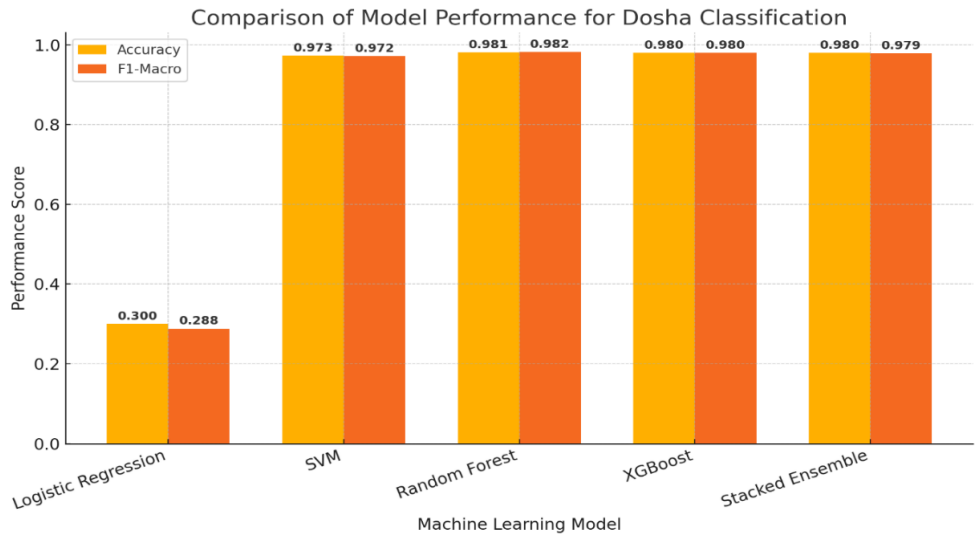


Fig. 5. Comparison of Model Accuracy and F1 Score

This is why the F1-score—specifically, the macro-averaged F1-score—has been included as a way to create a balance between precision and recall for all classes, while ensuring that the minority dosha combination also has equal weight during the evaluation. Therefore, when assessing predictive performance in a complex multi-class Ayurvedic classification environment, the combination of these two metrics together can create an accurate and dependable image for evaluating that model's performance.

Accuracy and Macro F1-score as the dosha classification task is a multi-class classification problem with imbalanced samples among different classes. Accuracy is the proportion of correctly predicted dosha profiles over all predicted dosha profiles of the model. While accuracy can help gauge how well a model is performing in aggregate, it can be misleading when there are fewer samples for certain dosha classes than others. To address this shortcoming, the Macro F1-score was also employed. F1-score is the harmonic mean of precision and recall, precision is how many of the predicted classes are really correct, recall is how many of the real classes are successfully predicted. Macro F1-score The macro-averaged F1-score is obtained by calculating the F1-score of each dosha category individually and then averaging the scores across all classes. An important aspect of this is that Macro F1-score seemed to be effective in assessing balanced model performance in multi-class Ayurveda dosha classification. Hence, reference to both considerations in the form of accuracy coupled with the Macro F1-score gives a fuller and more trustworthy picture of the overall efficiency, robustness, and fair treatment in respect of all dosha types. The figure 5, provides a clear view that Random Forest produced the best results overall, achieving an accuracy of 0.981, with an F1 - Macro value of 0.982, with XGBoost (accuracy = 0.980, F1 = 0.980) and the Stacked Ensemble model (accuracy = 0.980, F1 = 0.979) following closely, meaning all ensemble models showed excellent precision and stability across all dosha types, validating their ability to model nonlinear and hierarchical relationships in the data. SVM, although not an ensemble model, also achieved good performance with an accuracy of 0.973 and 0.972 for F1 - Macro, plus the kernel-based optimization enabled identification of complex boundaries among the overlapping dosha classes. Logistic Regression also dropped significantly in terms of performance (1.0-300, F1 = 0.288), indicating that the model may not have been able to model the relational interdependencies within the holistic health dataset well either. In conclusion, the figure shows that the ensemble tree-based models (Random Forest and XGBoost) are best fitted to this dosha classification task.

Random Forest and XGBoost outperform linear model methods of prediction in terms of modelling complex, non-linear relationships and many high order interactions among factors that make up the classification of Ayurvedic Doshas, which do not rely solely on the contribution of individual variables (like we see in Linear Regression and Logistic Regression), but instead rely on multiple interdependent combinations or configurations of the various categories of physiological, psychological and lifestyle-related factors that work together, as a group or ensemble of variables in a hierarchical and contextual manner. Linear Models (Logistic Regression), by design, make predictions based on making linear separations between the doshas (and their corresponding profiles) based upon the additive contribution of the various features/factors making up the classification of each Dosha profile. The natural boundaries between doshas (and the profiles corresponding to them) are much more fluid than what is accounted for by the methods of prediction we referred to previously, since they are affected by a multitude of different external factors that contribute to one's overall health/well-being, in addition to the classification of Doshas as their own separate entity.

While Linear Models can model the relationships between the different components that make up Dosha classifications as being additive in nature, Tree-based Ensemble prediction model methods have the capability to partition the feature space of the dosha classifications into different separate "bundles". Then the prediction paths for all

the partitioned bundles are then aggregated together, which allows for the modelling of non-linear characteristics of Dosha classification and for modelling the combinations and subtle interdependencies between the various attributes of the three categories of features/factors that form the basis of the Ayurvedic Dosha classifications. Random Forest builds its robustness through the use of bagging and random sampling of features, and thus reduces both variance and overfitting. Conversely, XGBoost utilizes a gradient boosting scheme to correct the residual errors from the previous iteration in order to build highly separated decision boundaries from the data. Therefore, due to their ability to model the holistic, multidimensional nature of doshas, both Random Forest and XGBoost are highly effective approaches for modelling dosha-based health data.

Despite the ensemble models, especially Random Forest and XGBoost, perform really well in terms of accuracy and Macro F1-score in this work, the results should be interpreted with caution with regard of the relatively small size of the dataset and the lack of external independent validation. The sample size of this dataset was only 125 responses to the questionnaire, which may pose a challenge to the generalization of the trained models to a wider population. Stratified sampling and adequate preprocessing were used to minimize bias and preserve class balance on training and testing, however, overfitting cannot be ruled out, especially times when complex ensemble-based classifiers are considered. As such, the performance should be interpreted as a promising preparatory results and not a definitive evidence for large-scale clinical reliability. Results are mainly indicative of the scope of implementing ensemble learning strategies for Ayurveda dosha classification. Further study should include larger multi-center datasets, cross-cultural subject groups, external validation datasets, and longitudinal assessment to additionally confirm the robustness, stability, and real-world applicability of the proposed system.

The Fig 6, depicts the graphical user interface (GUI) for the research-created Dosha Profile Classifier. The interface provides with the users to choose personal/lifestyle responses via dropdown menus tailored to each questionnaire feature. The model then predicts the dosha profile most likely to fit use of the selections made - "Vata Dominant" in this example. Another component of the GUI is the ability to show the best model performance metrics, in this case indicating that the Random Forest model demonstrates an accuracy and F1 macro score of 0.98. Overall, the implementation of this case study provides a useful and interactive tool for real-time Ayurvedic dosha classification via machine learning.

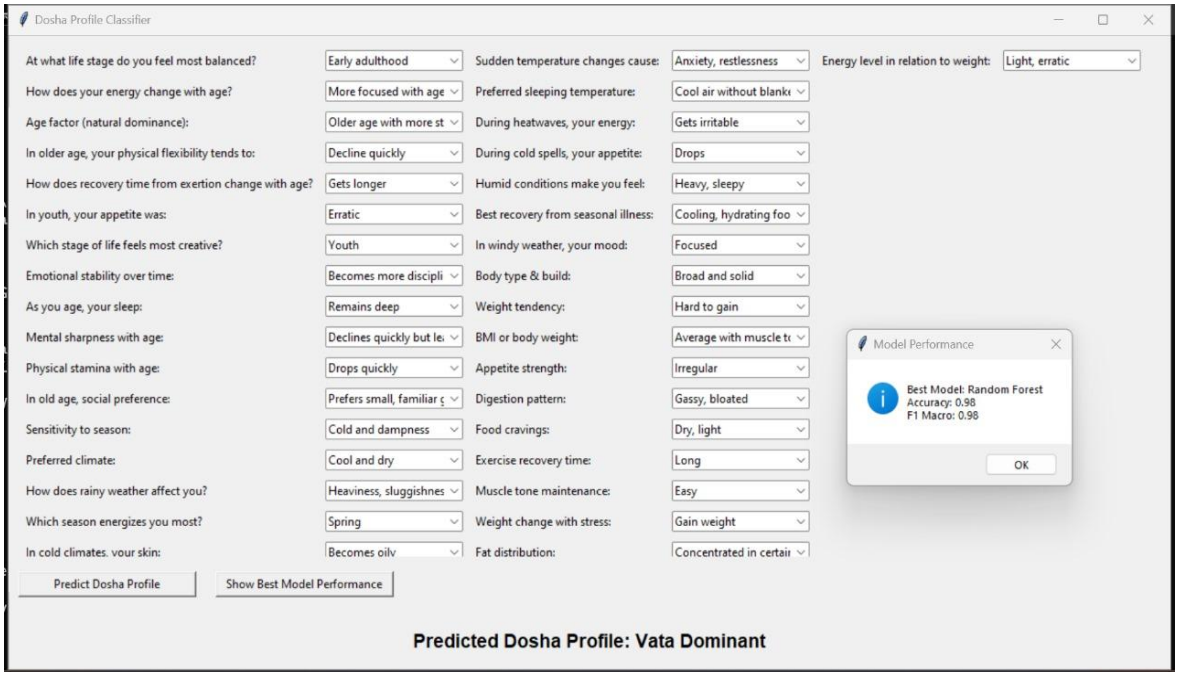


Fig. 6. GUI for Dosha Classification

The GUI created in this work was mainly oriented to illustrate the realization of the proposed machine learning based dosha classification system. The assessment was concentrated on funtionality, prediction accuracy, navigation, and real-time response status, user-friendliness rather thanandardized user evaluation metrics for a formal usability testing evaluation. Some informal testing was done to make sure the dropdowns, the model loading, the prediction generation and the visualization are working correctly and consistently with various input from the users. Even though the interface was relatively simple, interactive, and easy to use for non-technical users, extensive usability testing with larger user groups, user satisfaction analysis, task completion studies, and standardized usability scales were not conducted in the present study. Future works may entail end-to-end human-centric usability evaluation, which can

further enhance the user experience, interface design, accessibility, and practical application of the intelligent Ayurveda dosha prediction system.

6. Conclusion

This study demonstrates that individuals can accurately identify their dosha type using self-reported questionnaire data that utilizes a combination of machine learning and a comprehensive framework of Ayurveda. In addition, Random Forest and XGBoost outperformed linear regression in predicting dosha based on questionnaire responses because ensemble learning methods are able to accommodate the complex interactions among physiological, psychological and lifestyle factors that give rise to dosha characteristics. Furthermore, the performance of these classifiers is enhanced through systematic preprocessing and feature encoding, while the user-friendly graphical interface created for this project enables easy application of personalized wellness information by users and clinicians alike. The results from this study support the potential of using digital dosha classification as a decision support system for Ayurvedic practitioners; however, there are limitations to this approach. The classification of doshas was based solely on self-reported data, which creates a possibility of bias being present and affecting the reliability of the classification's accuracy. The limited number of participants and the various cultural backgrounds of participants participating in the study restrict the ability to generalise the findings of this study to other cultural groups and populations.

There was also a large predictive accuracy level achieved through machine learning techniques - but digital dosha classification should only be seen as a decision support system, rather than a replacement for skilled practitioners of Ayurveda. AI systems must also address an ethical framework that covers a number of considerations regarding the application of AI technology within health care settings, as this includes issues surrounding privacy, informed consent, and how the predictions produced by the AI are used to make decisions about an individual's health. The trust between users, their clinicians, and the AI systems they are using depends on the ability to explain how the AI system is operating. Challenges in order to deploy AI systems successfully in the case of real-world health settings include the data standardization, integration of AI systems into existing health care routines and practices, and compliance with regulations.

To improve the robustness, fairness, and clinical relevance of the digital Ayurvedic systems developed in this work, future research will continue examining how multimodal data can be combined, the potential for tracking a person's dosha over time, and the use of explainable AI methodologies. This work provides the foundation for the future of digital Ayurvedic Systems to ethically combine the medical wisdom of Ayurveda with modern computerized technologies (such as Artificial Intelligence) to allow for the development of personalized, proactive health care solutions.

All the Declarations and Statements

Author Contributions Statement

Rama Bhardwaj – Conceptualization, Methodology, Proposed research ideas, Constructed the overall framework, Model Training, Validation, and Performance Evaluation: Drafted the manuscript, contributed to the literature survey, and documented the technical background of the study.
Rakesh Kumar Saxena – Supervised project execution and coordinate project milestones and deadlines

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Declaration

None.

Data Availability Statement

The dataset supporting the findings of this study contains sensitive questionnaire-based profiles and is not publicly available as it is research work and simultaneously to protect the privacy and anonymity of the participants.

Ethical Declarations

This study involves survey-based data collection from willing human volunteers. All methods were carried out in accordance with relevant guidelines and regulations.

Acknowledgments

Not applicable.

Declaration of Generative AI in Scholarly Writing
Not applicable.

Abbreviations
None.

Appendix A\B\C…, with appendix tile
None.

References

[1] Vayadande, K., Sukte, C. D., Bodhe, Y., Jagtap, T., Joshi, A., Joshi, P., Kadam, A., & Kadam, S. (2024). *Heart disease diagnosis and diet recommendation system using Ayurvedic dosha analysis*. EAI Endorsed Transactions on Internet of Things. https://doi.org/10.4108/eetiot.6016

[2] Shariff, F., Huda, L. A., Maaz, M. H. K., Zeeshan, M., & Rehman, A. (2025). *Ayurvedic medicine recommendation*. IRE Journals, 9(6).

[3] Kumar, S. (2024). *Ayurvedic-based disease prediction and drug recommendation system*. Srusti Management Review, Special Issue, 27–37.

[4] Dwivedi, J. (2025). Ayurvedic Perspective on Vata Dosha and Neurological Disorders using Machine Learning Techniques. SGS - Engineering & Sciences, 1(4). https://spast.org/techrep/article/view/5660

[5] Sil, P., Tiwari, R., Garisetti, V., Baskaran, S. P., Dhanaseelan, F. H., Srivastava, S., & Samal, A. (2026). *Computational investigation of single herbal drugs in Ayurveda for diabetes and obesity using knowledge graph and network pharmacology*. arXiv. https://arxiv.org/abs/2601.21643

[6] Wang, C., Wang, F., Guo, R., Liang, Y., Liu, K., & Yu, P. S. (2024). Confidence-aware fine-tuning of sequential recommendation systems via conformal prediction. *arXiv Preprint arXiv:2402.08976*. https://doi.org/10.48550/arXiv.2402.08976

[7] Vayadande, K., Ghadekar, P. P., Sawant, A., Shelke, M. P., Shirsath, B., Bhande, B., & Samgir, S. (2024). Identification of Ayurvedic medicinal plant using deep learning. *International Journal of Intelligent Systems and Applications in Engineering, 12*(15s), 678–693.

[8] Garg, S. (2021). *Drug recommendation system based on sentiment analysis of drug reviews using machine learning*. arXiv. https://arxiv.org/abs/2104.01113

[9] Fathi, P., & Charkari, N. M. (2024). Herb target prediction using protein complexes detection and machine learning methods in heterogeneous network. *Kurdish Studies, 12*(4), 65–77.

[10] Ji, X., Li, Q., Liu, Z., Wu, W., Zhang, C., Sui, H., & Chen, M. (2024). Identification of active components for sports supplements: Machine learning-driven classification and cell-based validation. *ACS Omega, 9*(10), 11347–11355. https://doi.org/10.1021/acsomega.3c09234

[11] GunaChandra, S., Mounika, P., Yaswanth, K., & Prasad, M. K. S. R. (2024). Deep learning for medicinal plant identification and utilization: Leveraging ResNet for enhanced recognition and applications. *International Refereed Journal of Engineering and Science (IRJES), 13*(2), 159–165.

[12] Qian, Y., Wang, X., Cai, L., Han, J., Huang, Z., Lou, Y., et al. (2024). Model informed precision medicine of Chinese herbal medicines formulas–A multi-scale mechanistic intelligent model. *Journal of Pharmaceutical Analysis, 14*(4), 100914. https://doi.org/10.1016/j.jpha.2024.100914

[13] Pushpa, B. R., Rani, N. S., Chandrajith, M., Manohar, N., & Nair, S. S. K. (2024). On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach. *Ecological Informatics, 84*, 102611. https://doi.org/10.1016/j.ecoinf.2024.102611

[14] Jin, Y., Ji, W., Shi, Y., Wang, X., & Yang, X. (2023). Meta-path guided graph attention network for explainable herb recommendation. *Health Information Science and Systems, 11*(1), 5. https://doi.org/10.1007/s13755-023-00208-1

[15] Jin, Z., Zhang, Y., Miao, J., Yang, Y., Zhuang, Y., & Pan, Y. (2023). A knowledge-guided and traditional Chinese medicine informed approach for herb recommendation. *Frontiers of Information Technology & Electronic Engineering, 24*(10), 1416–1429. https://doi.org/10.1631/FITEE.2200457

[16] Yang, X., & Ding, C. (2023). SMRGAT: A traditional Chinese herb recommendation model based on a multi-graph residual attention network and semantic knowledge fusion. *Journal of Ethnopharmacology, 315*, 116693. https://doi.org/10.1016/j.jep.2023.116693

[17] Zhao, Z., Ren, X., Song, K., Qiang, Y., Zhao, J., Zhang, J., & Han, P. (2023). PreGenerator: TCM prescription recommendation model based on retrieval and generation method. *IEEE Access, 11*, 103679–103692. https://doi.org/10.1109/ACCESS.2023.3317221

[18] Bidve, P., Mishra, S., & Annapurna, J. (2023). *Enhancing Ayurvedic diagnosis using multinomial Naive Bayes and K-modes clustering: An investigation into prakriti types and dosha overlapping*. arXiv. https://arxiv.org/abs/2310.02920. https://doi.org/10.3390/math11163504

[19] Khatua, D., Sekh, A. A., Kutum, R., Mukherji, M., Prasher, B., & Kar, S. (2023). Classification of Ayurveda constitution types: A deep learning approach. *Soft Computing, 27*(9), 5309–5317. https://doi.org/10.1007/s00500-022-07653-8

[20] Patel, M., Shah, V., & Joshi, R. (2025). *A decision tree-based approach for disease prediction and Ayurvedic drug recommendation*. International Journal of Intelligent Systems and Applications in Engineering, 13(1), 77–89.
[1] Rao, N., Kulkarni, P., & Deshmukh, S. (2024). *Status of artificial intelligence and machine learning in Indian traditional medicine*. Journal of Medicinal Plants and Comparative Research, 8(3), 145–158.
[2] Sharma, K., Gupta, R., & Bansal, P. (2024). *Identification of medicinal leaves and recommendation system using deep learning*. International Journal of Intelligent Systems and Applications in Engineering, 12(4), 225–234.
[3] Ahmed, S., Khan, M., & Hussain, T. (2025). *An automated system for medicinal plant identification and usage recommendation*. International Journal of Advanced Computer Science and Applications, 16(1), 66–75.
[4] Lee, W. Y., Lee, Y., Lee, S., Kim, Y. W., & Kim, J. H. (2022). A machine learning approach for recommending herbal formulae with enhanced interpretability and applicability. *Biomolecules, 12*(11), 1604. https://doi.org/10.3390/biom12111604
[5] Malik, O. A., Ismail, N., Hussein,B. R., & Yahya, U. (2022). Automated real-time identification of medicinal plants species in natural environment using deep learning models—A case study from Borneo Region. *Plants, 11*(15), 1952. https://doi.org/10.3390/plants11151952
[6] Azadnia, R., Al-Amidi, M. M., Mohammadi, H., Cifci, M. A., Daryab, A., & Cavallo, E. (2022). An AI based approach for medicinal plant identification using deep CNN based on global average pooling. *Agronomy, 12*(11), 2723. https://doi.org/10.3390/agronomy12112723
[7] Joseph, A., Thomas, M., & Roy, P. (2024). *HerbId: A medicinal plant identification and recommendation framework*. AIP Conference Proceedings, 3122(1), 080026. https://pubs.aip.org/aip/acp/article/3122/1/080026/3298914/HerbId-A-medicinal-plant-identification-and

Authors' Profiles



Rama Bhardwaj is currently an Assistant Professor in the Department of Computer Science and Engineering at JECRC College, Jaipur, India. She is pursuing her Ph.D. in Computer Science from Poornima University, Jaipur, focusing on advanced Machine Learning concepts. She holds an integrated B.Tech. and M.Tech. degree in Computer Science & Engineering from Jayoti Vidyapeeth Women's University. Her core research interests lie in Machine Learning, computational intelligence, intelligent healthcare recommendation systems, and decentralized architectures. She has published multiple research papers in international journals and IEEE/IET conference proceedings, along with a published patent in hybrid classification algorithms for digital health solutions.



Dr. Rakesh Kumar Saxena, M. Tech (Dayalbagh Educational Institute, Agra, 1999) and Ph.D. (RGPV Bhopal, 2016), is currently working as Professor in Computer Science Department at Poornima University, Jaipur, Raj. India. He is Fellow member of The Institution of Engineers India, and member of IACSIT, ISTE Delhi. He was awarded as Chartered Engineer by IE (I) in 2017. His research focuses on soft computing, Artificial Intelligence Techniques-Macjine Learning, Deep Learning, Natural Language Processing and Computer Visions and has led projects funded by AICTE, MSME, and the World Bank. He has published 43 research papers and Book Chapters in IEEE Xplore, AIP, and other Scopus/SCI-indexed journals and Books published by Wiley, IGI Global.

How to cite this paper

Rama Bhardwaj, Rakesh Kumar Saxena, "Discrete Optimization Analysis of Ensemble Learning in Ayurvedic Dosha Classification", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.16, No.3, pp. 323-335, 2026. DOI:10.5815/ijwmt.2026.03.21