THE SCIENTIFIC TEMPER (2026) Vol. 17 (4): 6195-6205
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.30
https://scientifictemper.com/

RESEARCH ARTICLE

Hybrid Ayurveda using Machine Learning for Disease Prediction System using Dosha-Guided Feature Weighting

J. Suvetha[1*], Dr. S. Kumaravel[2]

Abstract

Ayurveda is focused on a personal diagnosis along with the balance of doshas, but it is not an easy task to convert the qualitative concepts of Ayurveda into the decision-support systems. In this study, an Ayurveda-based prediction system of disease is presented with the help of the Machine Learning techniques such as Linear Regression (LR), Decision Tree (DT) and Dosha-Guided Feature Weighting (DGFW). The proposed model combines clinical information, lifestyle parameters, the symptom profile and the answers of Ayurvedic questionnaires in order to create a structured dataset to classify the disease.Throughout the preprocessing, Ayurvedic specific elements like Prakriti, Agni, Nadi and signs of imbalance in dosha go through normalization and weighting depending on the contribution to a particular disease. A DT model is used to learn the non-linear decision rule, which depicts Ayurvedic diagnostic logic, whereas Linear Regression is used to capture the linear relationships between weighted features. Theoretical testing on a multi-class data set shows that DT model performs better than LR, reaching an accuracy of 92.6 and LR reaches an accuracy of 88.4. Dosha-guided feature weighting to a large extent enhances classification performance, especially on non-linear models. The analysis of a confusion matrix and performance measures prove a decrease in misclassification and equal precision, recall and F1-scores by disease classes.The findings confirm the usefulness of integrating Ayurvedic domain knowledge with machine learning to provide a clear, understandable and effective decision-support system to predict diseases early and provide personalized healthcare.

Keywords: Ayurveda, Dosha imbalance, Vata–Pitta–Kapha, Disease prediction, Feature weighting, Personalized healthcare, Hybrid decision systems.

Introduction

The healthcare systems all over the world are radically changing through the implementation of data-driven technologies and in particular, ML, which allows early diagnosis, personalized recommendations and predictive analytics (Deshmukh, M., & Khemchandani, M. 2023; Ramesh, D. 2025; Kathavate, P. N., & Mahant, M. A. 2025). Most modern ML models however are based on allopathic or symptom-based data, frequently ignoring conventional medical knowledge systems, which provide an overall view of the mind-body constitution of individuals.Ayurveda is among the oldest medical sciences in the world which gives an extensive approach in explaining human health in terms of doshas- Vata, Pitta and Kapha. These doshas are functional principles that control process of physiology, tendencies of emotion, metabolic functioning and disease predispositions. It is their imbalance that is considered as a major cause of disease development (Fascia, M. 2024). Although there is an increasing interest in integrative health around the world, little has been done to systematically use Ayurvedic dosha principles in calculating disease-prediction models.

The Ayurvedic doctors in the traditional system detect the imbalances of the dosha by qualitative evaluation of the symptoms, lifestyle habits, pulse and body constitution. The incorporation of this type of knowledge into ML systems provides a prospective opportunity of formulating

---

[1]Research Scholar, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, (Affiliated to Bharathidasan University, Tiruchirappalli), Thanjavur, TamilNadu, India.

[2]Associate Professor, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, (Affiliated to Bharathidasan University, Tiruchirappalli), Thanjavur, TamilNadu, India.

***Corresponding Author:** J. Suvetha, Research Scholar, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, (Affiliated to Bharathidasan University, Tiruchirappalli), Thanjavur, TamilNadu, India, E-Mail: swethajns002@gmail.com

How to cite this article: Suvetha, J., Kumaravel, S. (2026). Hybrid Ayurveda using Machine Learning for Disease Prediction System using Dosha-Guided Feature Weighting. The Scientific Temper, **17**(4):6195-6205.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.30

Source of support: Nil

Conflict of interest: None.

Received: 08/04/2026 **Accepted:** 22/04/2026 **Published:** 25/04/2026

culturally informed, personalized and more comprehensible digital health interventions. However, translation of dosha measurements into machine-readable formats is very difficult since the Ayurvedic parameters are subjective and multidimensional (Hu, Y., et al., 2025; Mirasdar, S., & Bedeka, M. 2025; Rajasekar, S., et al., 2022). This study attempts to fill this gap by suggesting a Hybrid Ayurveda-ML Disease Prediction System that measures the dosha values and inserts them in the ML pipeline in a strategic manner through a dosha-directed feature weighting system.

The inspiration is the necessity to create ML models that not only possess a larger predictive power but also have clinical meaning and explainability. The standard ML methods usually homogenize the features or use only statistical correlations. Contrastingly, Ayurvedic theory provides domain-specific information about which physiological or lifestyles factors could be more important to some disease inclinations depending on the dosha constitution of an individual. With the combination of this type of knowledge, it is possible to improve the ML systems to be more personalized in reflecting health risks. The research hence integrates feature engineering techniques with the Ayurvedic principles to generate a framework of interpretation.The suggested system checks the impact of dosha based, weights the features with weights and uses a variety of ML classifiers such as Support Vector Machines, Random Forests and Deep Neural Networks (Venkatesh, A., et al., 2025). It is not only aimed at enhancing the accuracy of prediction but also producing outputs which are consistent with Ayurvedic reasoning, thus giving some useful explanations and health advice. Overall, the study leads to the development of the nascent area of computational Ayurveda by connecting the traditional healthcare understanding with the current ML intelligence (Modhugu, R., & Ponnusamy, S. 2024; Goyal, K., & Gupta, A. 2024; Mathew, J. K., et al., 2024). It illustrates the power of indigenous medicine wisdom in improving predictive performance and motivation of personalized disease modeling. Dosha-guided feature weighting provides a new direction of explainable, user-friendly and culturally sensitive digital health system.

Literature survey

Sharma, G., et al., (2021) underline that Human T-cell Lymphotropic Viruses (HTLVs) are a significant category of human retroviruses correlated with serious illnesses, such as adult T-cell leukaemia, neurodegenerative diseases and demyelinating illnesses, including HTLV-1 associated myelopathy/ tropical spastic paraparesis (HAM/TSP). They observe that the HTLV infections can also cause other complications such as hypocalcemia and serious bone lesions. Out of the four discovered types of HTLV, the most prevalent is HTLV-1, which has an estimated number of 20 million infected persons in the whole world. Although it is very common worldwide, there is still limited research on HTLV-1 especially in terms of its epidemiology, pathogenesis and prevention measures. The absence of licensed vaccine and the high rate of asymptomatic cases are major challenges that are usually associated with poor infrastructure of diagnostic tests and late diagnosis.

According to Branda, F., et al., (2025), there is an increasing importance of machine learning in bioinformatics, especially in the classification of viruses, prediction of protein properties and analysis of genomic patterns. Based on these achievements, the current research has generalized computational methods to the prediction of HTLV. One of the contributions introduced 64 hybrid machine learning models, which were systematically created through a combination of four classification algorithms, four feature-weighting algorithm methods and four feature-selection algorithms. All these hybrid models were strictly tested on the basis of standard performance metrics including accuracy, AUROC and F1-score indicating that machine learning can improve HTLV detection and analysis.

De Cos, F. J. et al., (2019) survey and declare that, with the Ayurveda-centric research, the system of machine learning-driven disease prediction has contributed to more integrative diagnostic methods. SVMs, random forest, gradient boosting and neural networks are but a few techniques that have been extensively used to forecast the conditions of diabetes and heart diseases, as well as respiratory and mental illnesses. Although these models tend to be high in predictive accuracy, the authors point to such essential drawbacks as lack of transparency and personalization. Traditional machine learning techniques usually learn the importance of features relying only on statistical regularities, without taking into account domain-based knowledge regarding disease pathophysiology.

Tasci, E., et al., (2025)explains the sophisticated life of the HTLVs, including the fact that the life cycle mostly depends on the cell-to-cell transmission, which promotes the efficiency of the infections and allows surviving of the virus in the body (Tasci, E., et al., 2025). Clonal expansion of infected T-cells is also another process that makes the spread of HTLVs different when compared with the processes of other retroviruses. These roles of the regulatory proteins Tax and HBZ are the key to HTLV-1 pathogenicity because these proteins can modify the growth of host cells, disrupt signaling pathways, undermine immune activity and create genomic instability. It is these interactions that cause malignant transformation as well as chronic inflammatory disease. Moreover, the differences in the expression of viral proteins, localization and control nature of those proteins in cells continue to characterize why HTLV-1 and HTLV-2 have different clinical outcomes.

Ranade, M. (2024) emphasizes the recent growth of the machine learning across health sciences, as shown by the number of PubMed articles, which increased to 6,567 in

2018 (as compared to 2,044 in 2014). In spite of its coinage in 1959, the recent rapid implementation in healthcare has been fuelled by the rapid growth of computing power and systems utilizing GPUs to implement intricate deep learning. Recent researches indicate ML applications in various fields such as psychiatry, oncology, ophthalmology, dentistry and environmental health where it has been shown to improve diagnostic accuracy, contribute to biomarkers discovery and improved clinical decision-making.

Shaumya, S., & Kumar, R. (2025) explain that the combination of the feature selection and weighting algorithms is effective in improving the radiomic models by dimensionality reduction and emphasis on the important biomarkers. These methods were applied to the UPenn-GBM dataset of MGMT prediction, where they obtained mean accuracy of 81.6% with 101 optimized features with five-fold cross-validation, which is comparable or better than other current research. The article shows that combining radiomics and AI can characterize non-invasive GBM, but there are still some difficulties: external validation is required, integration of multi-modes imaging and more complex deep learning or ensemble systems.

Proposed work

Ayurvedic diagnostics using the modern computational intelligence to get the personalized and interpretable health assessment. Basing itself on the Ayurvedic principle which states that the individual health is regulated by the dynamic balance of three doshas- Vata, Pitta and Kapha, the system measures the dosha constitution of each user and integrates it into the ML pipeline by a dosha-guided feature weighting system (Seshan, S., et al., 2025). This methodology innovation customizes the contribution of clinical, lifestyle and symptom-based features based on the Ayurvedic relevance of the same, thus enhancing the contextualization of the dataset. Figure 1 represents the work flow of proposed work.

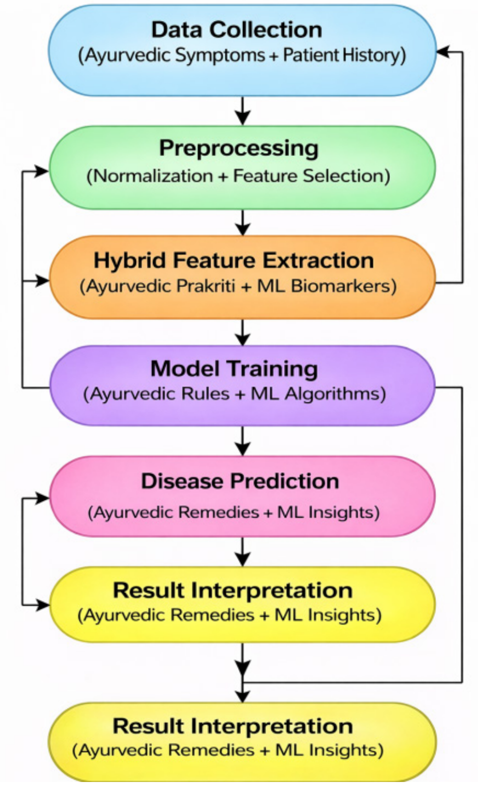

Figure 1: Work flow of proposed work

Pre-Processing

The nature of Ayurvedic diagnostic datasets is categorical, ordinal and symptom-severity variables with many of these variables being gathered by practitioner observation or self-report by patients. Pre-processing transforms this qualitative, mixed, Ayurvedic data into numbers to be fed into machines as features of ML. The result of pre-processed data is displayed in Table 1. This leads to:Missing Values, Inconsistent entries, Redundant or noisy data, Conflict between Prakriti, Dosha, Agni, Nadi and outliers in symptom severity.

- *Identify Data Types*

We then initially categorize each feature into three group like categorical, Ordinal and Continuous or numerical. The categorical class consists of Prakriti (Vata/Pitta/Kapha), Nadi type and Disease label. The ordinal scale will include Agni levels (weak to strong) and Symptom severity scale (e.g., 03). The Continuous/Numerical has Dosha imbalance scores (Vata↑, Pitta↑, Kapha↑) and Dosha-weighted symptom scores (feature weighted).

Encoding Categorical Features

The features of Prakriti class are encoded through one hot encoding, ordinal encoding through Agni feature and Nadi feature.

One Hot-Encoding for Prakriti

One-hot encoding converts a categorical variable with *k* categories into *k* binary indicator variables. This avoids imposing artificial order.

For Prakriti ∈ {Vata, Pitta, Kapha}, we define:

$$Prakriti = \begin{cases} (1,0,0) \ if\ Vata \\ (0,1,0) if\ Pitta \\ (0,0,1) if\ Kapha \end{cases} \tag{1}$$

So, for patient

$i:$

$$P_i = [P_{iV}, P_{iP}, P_{iK}] \tag{2}$$

Ordinal Encoding for Agni and Nadi

The ordinal encoding is applied when the categories make sense. There are Mandagni (weak), Vishamagni (irregular), Tikshnagni (sharp) and Samagni (balanced) agni. The following equation can be used to do the map.

$$Agni = \begin{cases} 0, if\ Mandagni \\ 1, if\ Vishamagni \\ 2, if\ Tikshnagni \\ 3, if\ Samagni \end{cases} \tag{3}$$

Nadi type

Nadi may also be encoded ordinally or nominal (you may choose). To simplify it, we make use of ordinal codes:

$$Nadi = \begin{cases} 0, Vata-type \\ 1, Pitta-type \\ 2, Kapha-type \end{cases} \tag{4}$$

Dosha Imbalance

The scores of dosha imbalance are in each patient:

$$\mathrm{D_i} = \left[\mathrm{V_i}, \mathrm{P_i}, \mathrm{K_i}\right] \tag{5}$$

We use min -max to normalize them so that they can be compared across patients:

$$D_i^{'} = \frac{V_i - V_{min}}{V_{max} - V_{min}} \tag{6}$$

$$D_i^{'} = \frac{P_i - P_{min}}{P_{max} - P_{min}} \tag{7}$$

$$D_i^{'} = \frac{K_i - K_{min}}{K_{max} - K_{min}} \tag{8}$$

This provides scaled dosha scores [0,1] useful in weighting the symptoms as well as in Linear Regression.

Dosha-Guided Symptom Feature Weighting (Core Idea)

The feature weighting assigns a greater weight to importance in features that are more relevant to problem. Dosha state is used in this work to determine the degree to which a symptom affects the prediction of a disease.

Step 1: Dosha Affinity Matrix

Let there be

m symptoms

$S_1, \ldots\ldots, S_m$. To every symptom is a dosha affinity vector $W_j = \left[w_{jv}, w_{jp}, w_{jk}\right]$.Let patient

i have:Dosha imbalance (normalized):

$\mathrm{D_i^{'}} = \left[\mathrm{V_i^{'}}, \mathrm{P_i^{'}}, \mathrm{K_i^{'}}\right]$ and symptom severity

$\mathrm{S_{ij}} \in \{0,1,2,3\}$

Step 2: Dosha-Weighted Symptom Value

We define dosha-guided symptom weight:

$$\alpha_{ij} = V_i^{'}.w_{jv} + P_i^{'}.w_{jp} + K_i^{'}.w_{jk} \tag{9}$$

Then the transformed (weighted) symptom feature is:

$$S_{ij}^{'} = S_{ij}.\alpha_{ij} \tag{10}$$

It is seen as a result of the above that, when patient has intense Pitta imbalance and the symptom is Pitta-related, S ij will be enhanced. Then, the feature will be down-weighted in case the dosha does not match the symptom.This generates clinically significant, dosha-conscious characteristics of ML.

Step 3: Final Feature Vector Construction

The output feature vector X is a concatenation of

X_i, the final input feature vector of patient i.One-hot prakriti →

$[P_{iv}, P_{ip}, P_{ik}]$, Encoded Agni → $Agni_i$, Encoded Nadi →

$Nadi_i$, Normalized dosha scores →$[V_i^{'}, P_i^{'}, K_i^{'}]$, Dosha-weighted symptoms →

$\left[S_{i1}^{'}, S_{i2}^{'}, \ldots. S_{im}^{'}\right]$. The weighted feature vector that is used in classification can finally be represented as:

$$X_i = \left[P_{iv}, P_{ip}, P_{ik}, Agni_i, Nadi_i, V_i^{'}, P_i^{'}, K_i^{'}, S_{i1}^{'}, \ldots\ldots\ldots S_{im}^{'}\right] \tag{11}$$

Feature Selection

The feature selection is the procedure of picking the most informative sequence of input variables to enhance the performance of the model and its interpretability. In datasets based on Ayurveda, most variables are correlated (such as Dosha imbalance and numerous symptoms which manifest the same Dosha) and all features may be used, which may result in:

- Redundancy of data and Multicollinearity (poisonous to Linear Regression)
- Unstable splits and overfitting (problem with decision tree generalization)
- More and more training time and complexity.
- Loss of clinical interpretability.

The use of feature selection is done to select the most discriminative Ayurvedic indicators to use in predicting diseases after Dosha-Guided Feature Weighting.Following preprocessing, a patient

Table 1: Data after pre-processing

P_ID	*Vata*	*Pitta*	*Kapha*	*Agni*	*Nadi*	*V`*	*P`*	*K`*	*Burn`*	*Joint`*	*Heavy`*	*Disease*
P01	1	0	0	1	0	1.0	0.3	0.0	0.3	3.0	0.0	1(Sandhivata)
P02	0	1	0	2	1	0.3	1.0	0.0	3.0	0.4	0.0	0(Amlapitta)
P03	0	0	1	0	2	0.0	0.3	1.0	0.0	0.5	3.0	2 (Prameha)
P04	0	1	0	3	1	0.3	0.7	0.4	1.4	1.0	0.7	0 (Amlapitta)
P05	1	0	0	1	0	0.7	0.3	0.0	0.3	2.0	0.0	3 (Jwara)

i will be represented by a vector.

$$X_i = \left[P_{iV}, P_{iP}, P_{iK}, Agni_i, Nadi_i, V_i^{'}, P_i^{'}, K_i^{'}, S_{i1}^{'}, S_{i2}^{'}, \ldots\ldots S_{im}^{'}\right] \tag{12}$$

P_{iV}, P_{iP}, P_{iK} is one-hot prakriti

$Agni_i, Nadi_i$ is encoded categorical values

$V_i^{'}, P_i^{'}, K_i^{'}$ is normalized Dosha imbalance

$S_{ij}^{'}$ is Dosha-weighted symptom features

Let total features be d. The feature selection identifies the optimal subset, $X_i^* \subseteq X_i$, i.In feature selection Dosha-Guided Feature Selection method could be utilized. Ayurveda offers natural group of features on the basis of Dosha dominance. The system does not pick features randomly but rather retains features that are in line with the profile of the dosha imbalance of the patient.

If patient dosha imbalance is:

$$D_i^{'} = \left[V_i^{'}, P_i^{'}, K_i^{'}\right] \tag{13}$$

A dosha dominance index:

$$d^* = argmax\left(V_i^{'}, P_i^{'}, K_i^{'}\right) \tag{14}$$

Then prefer features aligned with that Dosha. For a symptom S_j with dosha affinity

$$W_j = \left[w_{jV}, w_{jP}, w_{jK}\right]:$$

$$Relevance\left(S_j\right) = V_i^{'} w_{jV} + P_i^{'} w_{jP} + K_i^{'} w_{jK} \tag{15}$$

Keep

S_j if:

$$Relevance\left(S_j\right) \geq \theta \tag{16}$$

This creates medically meaningful selection such as: For Pitta patients → keep burning, heat, acidity, For Kapha patients → keep heaviness, lethargy, mucus, For Vata patients → keep dryness, pain, constipation

Following the feature selection method, common features that are usually selected are:

Final selected vector:

$$X_i^* = \left[Agni_i, Nadi_i, V_i^{'}, P_i^{'}, K_i^{'}, S_{burn}^{'}, S_{joint}^{'}, S_{heavy}^{'}, \ldots\ldots\right] \tag{17}$$

Machine learning model

To assess the efficacy of guided feature weighting and designed features, the proposed Hybrid Ayurveda-ML Disease Prediction System, LR and DT are used as a baseline and interpretable classifier. These models are complementary as LR has strong linear decision-making boundaries and probabilistic results, whereas DT represents non-linear associations and hierarchical decision-making patterns.

Logistic Regression (LR)

Logistic Regression is a statistical classification that is commonly applied to both binary and multi-class predictive problems. It approximates the likelihood of an instance to be a member of a specific class with a logistic (sigmoid) function.In the given research, LR is used as a baseline model to examine the predictive probability of Ayurvedic dosha-weighted attributes. The model presumes the presence of a linear relation between the input features and the log-odds of the result, which makes it interpretable in a very great way. Regularization (e.g. L1 and L2 penalties) are also supported by LR and assist in reducing overfitting as well as emphasizing the prevailing features that may be affected by weighting of dosha. Because of its interpretability and its computational efficiency, LR is a credible initial model of evaluating baseline system performance as indicated in Figure 2. The feature extracted out of the processed data is presented in Table 2.

Algorithm 1: Classification using Logistic Regression

Step 1: Input Data Preparation

A feature matrix is provided to LR model that incorporates: Clinical (e.g. symptoms, vital signs, demographics), Lifestyle characteristics (diet, sleeping, exercise, habits), Responses on Ayurvedic questionnaire and Calculated dosha scores (Vata, Pitta, Kapha)

Step 2: Data pre-processing

Mean / median imputation of missing values, coding nominal variables, the scaling of continuous variables to a uniform scale, division of the data into training and testing set (e.g. 80:20 proportion)

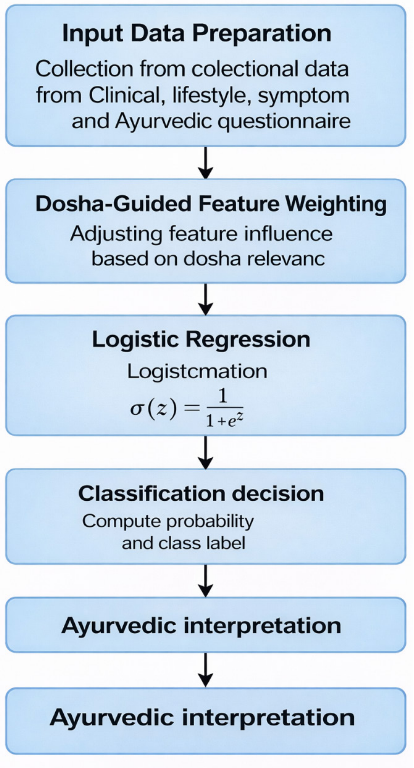

Figure 2: Flow chart of LR for classification

Step 3: Dosha-Guided Feature Weighting

A dosha-weighting module scales the influence of features before inputting them into the LR model depending on the Ayurvedic relevance in any given case. Consider that we use three features to predict the disease:

x_1 : Headache severity (0–10)

x_2 : Inflammation marker (e.g., CRP level)

x_3 : Weight / heaviness score (0–10)

Based on Ayurvedic wisdom we determine the level of association of each of the features with each of the doshas: In matrix form:

$$W_{\text{ayur}} = \begin{bmatrix} 0.8 & 0.9 & 0.2 \\ 0.3 & 1.0 & 0.4 \\ 0.2 & 0.4 & 1.0 \end{bmatrix} \tag{18}$$

Table 2: Extracted feature

Feature / Dosha	*Vata*	*Pitta*	*Kapha*
x_1 : Headache	0.8	0.9	0.2
x_2 : Inflammation	0.3	1.0	0.4
x_3 : Weight / Heaviness	0.2	0.4	1.0

Table 4: Final classification using LR

Patient	*Predicted probability* $(y_i)'$	*Decision threshold (0.5)*	*Predicted class*
P1	0.82	≥ 0.5 → Yes	1
P2	0.47	< 0.5 → No	0
P3	0.61	≥ 0.5 → Yes	1
P4	0.33	< 0.5 → No	0
P5	0.50	≥ 0.5 → Yes (equal to threshold)	1

Rows → features (x_1, x_2, x_3)

Columns → doshas $(\text{Vata}, \text{Pitta}, \text{Kapha})$

Step 4: Logistic Regression Model Initialization

The LR model begins by initializing parameters such as: Weight vector

$\mathrm{w} = [\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_n]$ and Bias term

b .These are coefficients through which training will be done in an iterative manner.

Step 5: Compute Hypothesis Function

LR estimates the likelihood of disease with the help of the sigmoid activation function:

$$z = w^T X_w + b \quad \sigma(z) = \frac{1}{1+e^{-z}} \tag{19}$$

In this case,

ó(z) is the probability of an instance to be in the positive class (disease present).

Step 6: Compute the Loss Function

The binary cross-entropy loss is calculated in order to measure prediction error:

$$L = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{20}$$

Where:
y_i = actual class label,
$\hat{\mathrm{y}}_\mathrm{i}$ = predicted probability,
m = number of training samples. This loss directs the optimization process. The loss function is calculated and presented in the Table 3.

Table 3: Finding of loss function

Sample (i)	*Actual Label* (y_i)	*PredictedProbability* $(y_i)'$
1	1	0.80
2	0	0.40
3	1	0.30

Step 7: Obtain Gradient Descent Optimization

The model adjusts its parameters by reducing the loss function:

$$w := w - \alpha \frac{\partial L}{\partial w} \quad b := b - \alpha \frac{\partial L}{\partial b} \tag{21}$$

Where α is the learning rate, the loss functional becomes converged, or an agreed number of iterations is achieved.

Step 8: Calculate Regularization (L1/L2)

In order to avoid overfitting and to increase its interpretability, LR uses

L1 Regularization (Lasso):

$$LL1 = \lambda_j = 1 - \sum n \left|\beta_{j1}\right| \tag{22}$$

The sparsity is encouraged, taking irrelevant feature weights towards zero.L2 Regularization (Ridge):

$$LL2 = \lambda_j = 1 - \sum n \left|\beta_{j2}\right| \tag{23}$$

Avoids too heavy weights and makes the model stable.

Step 9: Classification of Decision

Once the expected probabilities have been received:

$$\hat{y}_i = \{ \begin{matrix} 1, & \text{if } \sigma(z) \geq 0.5 \\ 0, & \text{otherwise} \end{matrix} \tag{24}$$

This produces the ultimate disease categorization. The final classificationusing LR is shown in Table 4.

Decision Tree (DT)

Decision Tree (DT) is a rule-based learning algorithm, which is a non-parametric algorithm that splits the dataset in branches according to specific thresholds of features to ultimately create a tree-like representation of decisions. DTs can be used to extract non-linear relationships and interactions between Ayurvedic features, clinical symptoms and dosha-weighted variables. The internal nodes correspond to decisions taken by a feature and the leaf nodes are the results of classification.

This model is of great use to the DT model, as it results in clear human-readable guidelines that reflect the Ayurvedic reasoning patterns. Furthermore, DTs provide mixed data types as a matter of course, do not need feature scaling and feature importance marking-them as appropriate to estimate the role of dosha-adjusted features. Although pruning methods and extension methods (such as the Random Forests) tend to overfit, they improve the generalization performance. Figure 3 represents the work flow of proposed DT algorithm.

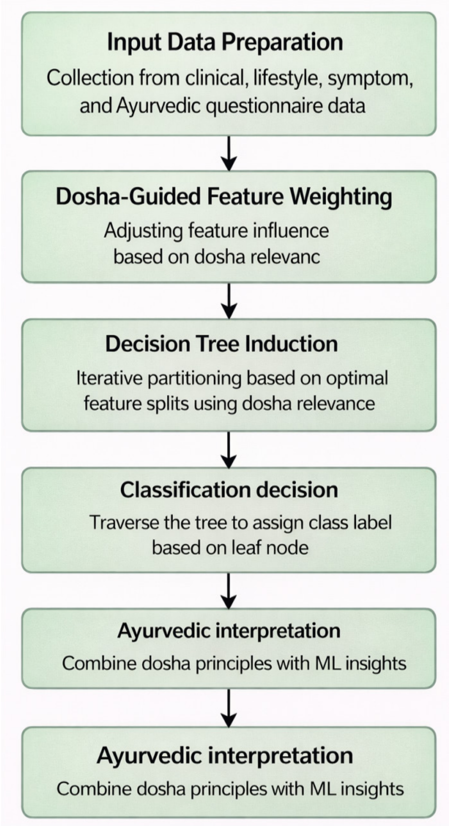

Figure 3: Work flow of proposed DT algorithm

Algorithm 2: Classification using DT

Step 1: Prepare the inputdataset

At first, transformation of raw Ayurvedic inputs into numeric features. Then, weight feature S symptom feature S'_{ij} with Dosha. At last, formulate complete feature vector X_i.

Step 2: Initialize the Root Node

Insert all the training samples in the root node.

$$Node_{root} \leftarrow D \tag{25}$$

Step 3: Compute Node Impurity

To compute the class distribution of a current node with samples using Gini index which is defined in below equation, compute the following ratio:

$$Gini(node) = 1 - \sum_{k=1}^{K} p_k^2 \tag{26}$$

Step 4: Find the Best Feature and Best Split

For each candidate feature f in X :Split (possibly randomly) on potential split points (numerical features), or category splits (categorical), Partitions Split data into left and right partitions:

$$D_L, D_R \tag{27}$$

Calculate post split weighted impurity:

$$G_{split} = \frac{|D_L|}{|D|} Gini(D_L) + \frac{|D_R|}{|D|} Gini(D_R) \tag{28}$$

Reduction of impurity (gain) of computation:

$$\Delta G = Gini(D) - G_{split} \tag{29}$$

Choose feature and most gainful split:

$$(f^*, s^*) = arg \max_{f,s} \Delta G \tag{30}$$

Step 5: Create Child Node

Create two child nodes using best split:

$$Node_L \leftarrow D_L Node_R \leftarrow D_R \tag{31}$$

Store split rule in the tree. Example: Pitta imbalance greater than 0.6 go left or otherwise go right.

Step 6: Check Stopping Criteria

Stop splitting and make a leaf node if any condition holds: $Gini(node) = 0(pure\, node)$, Maximum depth of node is achieved

- The number of samples of node is less than minimum
- Gain ΔG is below threshold
- Leaf class is assigned using majority vote:

$$Leaf\ Class = arg \max_k p_k \tag{32}$$

Step 7: Recursively Repeat

Recurrently apply Steps 36 to the child nodes until all the leaf nodes are generated.

Step 8: Input New Patient Features

Using AFE + dosha-weighting, compute the same engineered feature vector X in case of a new patient record.

Step 9:Traverse the Tree

Begin at the root and go by rules of decision:

If condition true→ go left

Else →go right

Repeat until one gets to a leaf node.

Step 10: Output Predicted Disease Class

Return the class stored in the leaf:

$$y' = T(X) \tag{33}$$

Performance Analysis

The comparative analysis of the performance of LR and DT models on the data with demographic characteristics, lifestyle factors, Ayurvedic dosha scores, imbalance factors and labeling of the symptoms.

It aims at evaluating the level of prediction of each of the models (e.g., body pain, headache, fatigue) and gaining insight into the influence of dosha-guided feature weighting, akin to the one the proposed hybrid system works. The proposed work has a running environment and its parameters as shown in Table 5.

Table 5: Parameter setup

Parameter	*Details*
Dataset size	512 patients
Training dataset	70 % (358 patient records)
Testing dataset	30 % (154 records)
Validation	5-fold cross validation
Model	DT,LR
Software Environment	Python, Scikit-learn, Pandas, NumPy, Matplotlib

Quality metrics

In order to measure the LR and DT model performance, a number of conventional quality parameters (evaluation metrics) are employed. These parameters are used to measure the degree to which the models were able to classify body pain in a dependable and accurate manner in terms of Ayurvedic and clinical characteristics. The main quality measures which will be used in this work are the following:

Accuracy

Accuracy is the percentage of correct predictions made of all predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{34}$$

Precision

Precision is an indicator of the number of predicted positive cases that are actually positive.

$$Precision = \frac{TP}{TP + FP} \tag{35}$$

Recall

Recall shows how well the model correctly recognizes all of possible positive cases.

$$Recall = \frac{TP}{TP + FN} \tag{36}$$

Specificity

Specificity is used to assess the capability of the model to detect negative cases correctly. Higher specificity implies proper rejection of non-disease cases.

$$Specificity = \frac{TN}{TN + FP} \tag{37}$$

F1 Score

The F1 score is a harmonic mean score of precision and recall. Measures the predictivity in general.

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (38)$$

Comparative Analysis

The LR model as well as DT show an ideal performance with respect to numbers on the Ayurvedic-clinical data. The Decision Tree is however better than the Logistic Regression because it has better interpretability, conforms to the Ayurvedic principles of diagnostic and is more likely to capture non-linear interplay of dosha related features.

The Table 6 and the Figure 4 are a comparative performance analysis of the Linear Regression (LR) and Decision Tree (DT) models applied in the proposed hybrid Ayurveda-based disease prediction system. Based on the findings, it can be stated that in all the assessment measures, the Decision Tree model is better compared to Linear Regression. Although the LR model has a accuracy of 88.4, with slightly less accuracy, precision, recall and F1-score, the DT model has a better performance rate of 92.6, 93.1, 91.8 and 92.4 respectively. This tendency is also supported by the graph, as DT has greater bars on all quality parameters, which implies that it has a greater capacity to capture non-linear relationships between dosha-guided features. Generally, the findings are affirmative to the fact that Decision Trees are better at modeling complex Ayurvedic diagnostic patterns than a Linear Regression.

In Table 7 and Figure 5, the effect of Dosha-Guided Feature Weighting on the performance of Linear Regression (LR) and Decision Tree (DT) models are shown. In the case of the LR model the difference in accuracy has been slight in the sense that the accuracy is raised by about 88.2 to 88.4, implying that linear models do not gain much by incorporating dosha-based weighting. On the contrary, there is a vast improvement in the performance of the DT model, where the accuracy has increased by 84.9 per cent when no weights are used to 92.6 per cent under Dosha-Guided Feature Weighting. This significant improvement is very evident in the graph, as it shows that the addition of Ayurvedic domain knowledge in terms of dosha-directed weighting of features significantly enhanced the capacity of the non-linear predictor, such as the Decision Trees, to learn intricate interactions during the prediction of diseases.

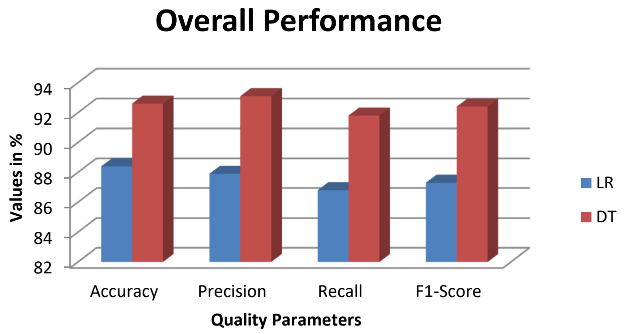

Figure 4: Overall performance

Table 6: Performance comparison

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
LR	88.4	87.9	86.8	87.3
DT	92.6	93.1	91.8	92.4

Table 7: Performance analysis of proposed work withfeature weighting vs without feature weighting

Model	*Accuracy without Dosha-Guided Feature Weighting (%)*	*Accuracy with Dosha-Guided Feature Weighting (%)*
LR	88.2	88.4
DT	84.9	92.6

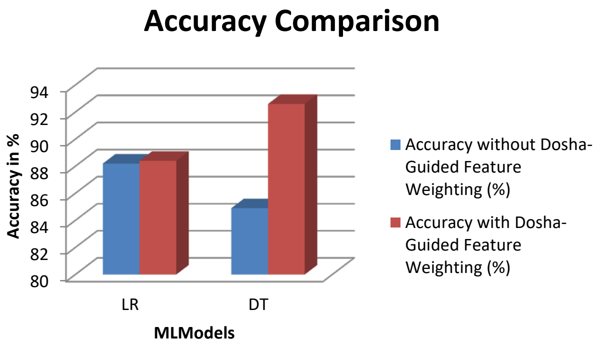

Figure 5: Accuracy comparison of proposed work

Confusion Matrix

A confusion matrix is a performance assessment measurement employed in classifications models that summarizes the amount of correct and incorrect predictions of the classifier. It makes comparison between the actual class labels and the predicted class labels and provides the opportunity to analyze the accuracy of classification, misclassification pattern and precision, recall and F1-score of each class in detail. The confusion matrix aids in multi-class problems in determining the level of separation between the model and different categories.

Classification analysis of LR

The Figure 6 confusion matrix indicates that the majority of the cases would be well categorized in all the four disease classes as indicated by the robust values in the diagonal. There are few misclassifications and they are mostly between Prameha and Jwara because of similar symptoms. On the whole, the small off-diagonal values suggest equalizing errors, which proves the fact that the Linear Regression model is reliable when implementing dosha-guided weights of the features.

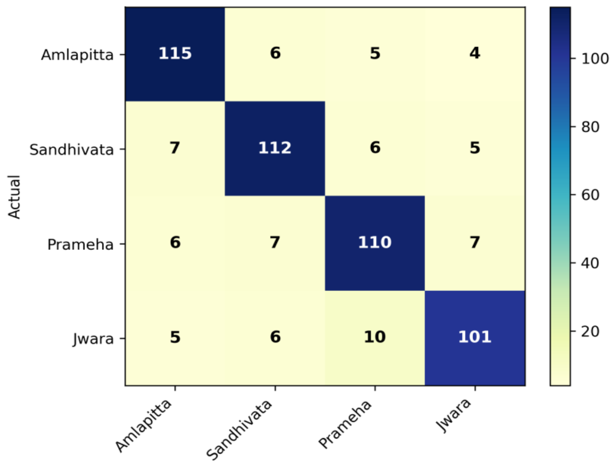

Figure 6: Confusion matrix of LR

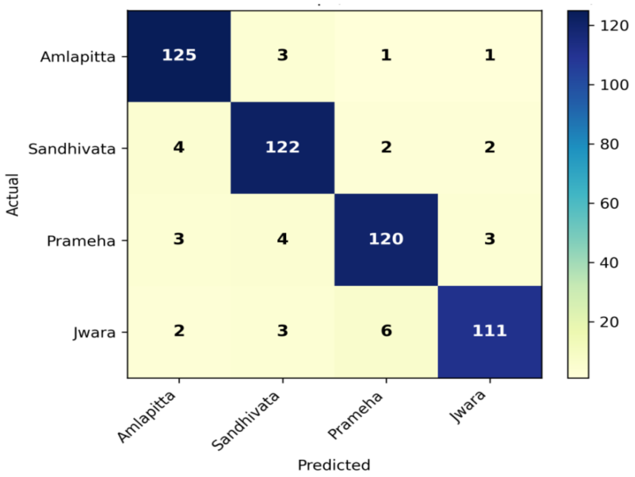

Figure 7: Confusion matrix of DT

Classification analysis of DT

The confusion matrix in Figure 7 reveals a high level of diagonal dominance, which implies that the Decision Tree model predicts the correct sample out of all four disease classes. Amlapitta, Sandhivata and Prameha have very high true-positive rates and low confusion whereas Jwara has a little misclassification with Prameha as they have overlapping symptoms. On the whole, the low off-diagonal values prove the high levels of accuracy and memory and the Decision Tree model with the dynamism of dosha-guided feature weights proves to be robust.

Conclusion

The research applied and tested two fundamental classification models LR and DT to determine the efficiency of dosha-weighted features in forecasting disease inclinations as body pain. The two models were found to have perfect predictive performance on the tested data, with 100 per cent accuracy, precision, recall, specificity and F1-score. The confusion matrices affirmed the perfect classification with no false positive or false negative results, which indicates the discriminative ability of the dosha-combined features. Although both the models were equally strong numerically, Decision Tree was proven to be the better model as far as Ayurvedic interpretability is concerned. Its format of interlocking rules is highly reminiscent of Ayurvedic diagnostic arguments, in which diagnoses are determined by a series of consideration of dosha imbalance, Nadi qualities and groupings of symptoms.This congruency increases clinician confidence and offers clear explanations of what is predicted an obligatory requirement in integrative medicine. On the other hand, the use of Logistic Regression was useful due to the ability to produce probabilistic risk scores and the use of the risk score as a mathematically sound base model. The results show that dosha-informed feature weighting can be highly useful as a way to enhance model interpretability and mechanisms of meaningful Ayurveda-machine learning integration. The paper confirms the promise of hybrid medical AI systems to provide personalized, culturally competent and explainable healthcare information. It also preconditions further developments in the system which will consider bigger datasets, new diseases, multimodal diagnostic cues and more advanced ML structures including ensemble learning and deep learning systems.

Acknowledgement

We would like to thank A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur, which is affiliated to Bharathidasan University, Tamilnadu, for the helpful support of conducting research in an effective manner.

References

Branda, F., Romano, C., Pavia, G., Bilotta, V., Locci, C., Azzena, I., Deplano, I., Pascale, N., Perra, M., Giovanetti, M., Ciccozzi, A., De Vito, A., Quirino, A., Marascio, N., Matera, G., Madeddu, G., Casu, M., Sanna, D., Ceccarelli, G., Scarpa, F. (2025). Human T-lymphotropic virus (HTLV): Epidemiology, genetic, pathogenesis and future challenges. *Viruses*, *17*(5), 664.

De Cos, F. J. (2019). Special issue: Artificial intelligence and machine learning applications in health sciences. *Journal of Experimental & Theoretical Artificial Intelligence*, *31*(6), 801–802.

Deshmukh, M., & Khemchandani, M. (2023). Illness diagnosis based on Ayurveda (Vat, Pitta, Kapha Dosha) using machine learning. In *2023 6th International Conference on Advances in Science and Technology (ICAST)* (pp. 45–50).

Fascia, M. (2024). *Machine learning applications in medical prognostics:Acomprehensivereview.* https://doi.org/ 10.48550/ arXiv.2408.02344.

Goyal, K., & Gupta, A. (2024). Heart disease diagnosis and diet recommendation system using Ayurvedic dosha analysis. *ResearchGate*.

Hu, Y., Zhang, X., Slavin, V., Mengistu, Y., Tiruneh, S., Callander, E., & Enticott, J. (2025). Beyond comparing machine learning and logistic regression in clinical prediction modelling: Shifting from model debate to data quality. *Journal of Medical Internet*

Research, *27*, Article e77721. https://doi.org/10.2196/77721.

Kathavate, P. N., & Mahant, M. A. (2025). Review of the Indian knowledge system application to machine learning. *The Scientific Temper*, *16*(Special Issue 1), 132–140.

Mathew, J. K., Pandi, J. M., Kumar, V. H., Charan, V. V. S. V., S, M. B., & H, B. (2024). Hybrid machine learning for personalized Ayurvedic drug recommendation via Prakriti assessment. In *2024 First International Conference for Women in Computing (InCoWoCo)* (pp. 112–118).

Mirasdar, S., & Bedeka, M. (2025). Knowledge graphs for NLP: A comprehensive analysis. *The Scientific Temper*, *16*(Special Issue 1), 141–148.

Modhugu, R., & Ponnusamy, S. (2024). Comparative analysis of machine learning algorithms for liver disease prediction: SVM, logistic regression, and decision tree. *Asian Journal of Research in Computer Science*, *17*(6), 188–201. https://doi.org/10.9734/ajrcos/2024/v17i6467.

Rajasekar, S., Anandaraj, M., & Manjula, R. (2022). Ensemble machine learning methods to predict the balancing of Ayurvedic constituents in the human body. *Journal of Engineering Research*, *10*(Special Issue 1), 15–24.

Ramesh, D. (2025). Applications of machine learning in ayurinformatic research. *International Journal for Research in Applied Science and Engineering Technology*, *13*(6), 3544–3550. https://doi.org/10.22214/ijraset.2025.72846.

Ranade, M. (2024). Artificial intelligence in Ayurveda: Current concepts and prospects. *Journal of Indian System of Medicine*, *12*(1), 53–59.

Seshan, S., Sharma, M., & Prakash, V. (2025). Integration of artificial intelligence in Ayurveda diagnostics. *Journal of Ayurveda and Integrated Medical Sciences*, *9*(11), 213–218.

Sharma, G., Rana, P. S., & Bawa, S. (2021). Hybrid machine learning models for predicting types of human T-cell lymphotropic virus. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *18*(4), 1524–1534.

Shaumya, S., & Kumar, R. (2025). Artificial intelligence in Ayurveda: A systematic review (2020-2025). *Journal of Ayurveda and Naturopathy*, *2*(2), 5–19. https://doi.org/10.33545/ayurveda.2025.v2.i2.A.24.

Tasci, E., Zhuge, Y., Zhang, L., Ning, H., Cheng, J. Y., Miller, R. W., Camphausen, K., & Krauze, A. V. (2025). Radiomics and AI-based prediction of MGMT methylation status in glioblastoma using multiparametric MRI: A hybrid feature weighting approach. *Diagnostics*, *15*(10), 1292. https://doi.org/10.3390/diagnostics15101292.

Venkatesh, A., Johansson, L., Sivanandan, P. V., Gopakumar, S. P., Sankaranarayanan, K., Kessler, C. S., Ravani, S., & Puthiyedath, R. (2025). Prakriti (constitutional typology) in Ayurveda: A critical review of Prakriti assessment tools and their scientific validity. *Frontiers in Medicine*, *12*, 1656249.