THE SCIENTIFIC TEMPER (2026) Vol. 17 (3): 5756-5766
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2026.17.3.03
https://scientifictemper.com/
RESEARCH ARTICLE

Development of an Ayurveda-Integrated Feature Engineering Framework for Disease Prediction

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

Abstract

A combination of the conventional Ayurvedic diagnostic knowledge and the recent computational intelligence should provide a direction to an improved way of improving the accuracy of diagnosing diseases and broadening the horizon of the entire healthcare provision. This paper introduces an Ayurveda-Based Feature Engineering (AFE) Framework in disease prediction with the assistance of machine-learning techniques. The systematically Ayurvedic diagnostic parameters of the Ayurvedic Prakriti constitution, Dosha imbalance, Agni condition, Nadi, and Astavidha Pariksha are systematically translated into structured machine-readable numerical features. To create a high-quality set of features that was consistent with the traditional medical reasoning and data science demands, a dataset gathered in the Ayurvedic hospitals and clinics was annotated with these parameters encoded. Several machine learning classifiers such as the random forest (RF), the support vector machine (SVM) as well as the navie bayes (NB) were trained and optimized using this improved dataset. Experiments indicate that using Ayurveda elements of diagnoses leads to a significant increase in predictive performance over traditional symptom-only models with significant improvements in accuracy, F1-score, and AUC measures. The presented AFE framework contributes to a powerful bridge between Ayurveda classical and contemporary predictive analytics and makes it possible to implement culturally-rooted and predictable disease-based forecasting systems. The contribution provides the foundation of the future study on integrative healthcare analytics and provides a scalable framework of building more sophisticated Ayurveda-informed clinical decision support systems.

Keywords: Ayurveda-Based Feature Engineering (AFE), Disease Prediction, Machine Learning Classifiers, Prakriti and Dosha Encoding, Integrative Healthcare Analytics.

Introduction

Ayurveda is the traditional Indian system of medicine, that has a history of thousands of years and a holistic view of health, disease prevention and treatment of a specific individual (Ahuja, N.,et al., 2024). In contrast to the present-day biomedical systems that are mainly concerned with physiological anomalies, Ayurveda pays more attention to the harmonies of the three-basic bio-energies Vata, Pitta, and Kapha, as well as other elements of diagnosis, including Prakriti (constitution of the body), Agni (digestive fire), Mala (heap), Nadi (pulse) and Astavidha Pariksha (eight-fold examination). These principles offer an extensive, patient-centric method of disease manifestation. Nevertheless, the qualitative and descriptive Ayurvedic diagnosis is very difficult to compute, which reduces the possibility of mass-digitizing healthcare and automated clinical decision support.

Recent developments in machine learning and artificial intelligence provide more opportunities than ever to convert conventional medical knowledge into predictive and analytical systems (Balkrishna, A., et al., 2024). The success of the models of disease prediction that rely on machine learning has been astonishing in the field of contemporary healthcare. However, majority of the available models are based more on biomedical measures, laboratory results, or symptomatic measures, frequently ignoring rich traditional

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

[2]Research Advisor and Associate Professor, PG and Research Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur(Dt), (Affiliated to Bharathidasan University, Tiruchirappalli), Tamil Nadu, India

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

How to cite this article: Suvetha, J., Kumaravel, S. (2026). Development of an Ayurveda-Integrated Feature Engineering Framework for Disease Prediction. The Scientific Temper, **17**(3):5756-5766.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.3.03

Source of support: Nil

Conflict of interest: None.

Received: 05/02/2026 **Accepted:** 10/02/2026 **Published :** 25/03/2026

diagnostic models like the Ayurveda. The way to close this gap is to develop new methodologies that will be able to transform qualitative Ayurvedic knowledge into structured and numerical forms which can be used in computational modeling (Baragi, U. C., & Ganer, J. M. 2025).

In order to solve these problems, this study suggests a new Framework that will encode classical Ayurvedic diagnostic characteristics in a systematic way as machine-readable characteristics. The framework translates subjective diagnostic information into numerical vectors with numbers, which, in turn, allows machine learning algorithms to apply deep-rooted Ayurvedic clinical knowledge (Chaturvedi, S., et al., 2024). This combination does not only increase the predictive accuracy but also makes the process of disease prediction culturally and contextually relevant to various groups.

The suggested framework is tested on various machine learning classifiers, such as Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Networks. Comparative experiments illustrate that the Ayurvedic feature engineering can be useful in enhancing the model performance when compared with traditional symptom-only model. The results highlight the possibility of integrative healthcare analytics where conventional medical expertise and contemporary computational methods are used to complement each other to further the predictive diagnostics.

The work of this study is relevant to the emerging area of Ayurveda informatics and preconditions the creation of intelligent, explainable, and culturally specific disease prediction system (Chouhan, P. N., & Joshi, A. 2025). Ayurvedic knowledge being included in the machine learning not only modernizes the traditional practice but also enhances the movement worldwide towards personalized and holistic healthcare.

Related Works

Umadevi, V., & Kamath, V. (2024) investigated how machine learning methods could be used to classify Ayurvedic herbs according to dosha-balancing. One of such works used a combination of multiple preprocessing methods, such as random oversampling and synthetic data creation with CTGAN, to deal with the issue of class imbalance. The data was separated into a training and testing sample in the proportions of 90:10 and six experimental analyses were carried out. Several classifiers were used to evaluate the predictive performance such as SVM, KNN, Random Forest, Decision Tree, and XGBoost. Among them, XGBoost proved to be the best, with no oversampling, which has 96% accuracy and high precision (97%), recall (96%), F1-score (96%), and low RMSE of 0.16. The results indicate that ensemble learning methods, in particular, XGBoost, can be used to achieve significant improvements compared to traditional models in the Ayurvedic herb classification.

Kathavate, P. N., & Mahant, M. A. (2025)have concentrated on incorporating the traditional Indian Knowledge Systems (IKS) Ayurveda, Yoga, Vedic mathematics, and Jyotish Shastra with modern machine learning algorithms to improve the performance of the technology and its cultural relatability. These articles demonstrate the application of ancient concepts to the modern algorithms, including the use of Vedic mathematics to optimally calculate computations and using the idea of Ayurveda to help personalized healthcare frameworks. The integration is also meant to enhance the accuracy and efficiency of the algorithms as well as to maintain and advance the native knowledge traditions. Altogether, this research stream highlights the opportunities of integrating IKS with contemporary data-driven techniques to come up with culturally informed, innovative solutions in healthcare, education, and the policy sphere.

Vayadande, K., et al., (2024)considered the application of Ayurvedic diagnostic concepts and modern sensing technologies in prediction of diseases. The paper is aimed at integrating classical Dosha and Prakriti with pulse sensors and machine learning to diagnose the patient state and give him or her individual health guidelines. These models were compared, and Decision Tree has the highest accuracy (99.70%), then, Random Forest (89.21%), KNN (86.43%), and Logistic Regression (85.94). The study shows that integrating an ancient Ayurvedic pulse diagnosis with modern computational techniques can allow early diagnosis and advanced and personalized medicine.

Gandhmal, D.P., et al., (2024) created a software that could suggest appropriate herbs and drugs, depending on the reported symptoms by the user and being backed by authoritative texts and databases of Ayurveda. The system standardized names of diseases as well as gave extensive information about medicinal properties. Also, a chatbot based on Ayurvedic was developed to provide natural healthcare advice on a case-to-case basis by extracting appropriate remedies out of a broad symptom-medicine database. The study also covered the state-of-the-art language processing techniques to enhance the comprehension and the accuracy of the responses of the chatbot. On the whole, the analysis shows that by combining classical Ayurvedic information with software and NLP tools (Mirasdar, S., & Bedeka, M. 2025), one may easily develop convenient, effective, and easily accessible healthcare interventions.

Nayak, S. K., et al. (2025) relevance and sustainability of the School Ayurveda and Yoga Health Program (SAHYP) that was adopted in seven Nepal districts were evaluated. The program gave emphasis on preventative healthcare by encouraging healthy lifestyles, teaching Medicinal plants to students and educating them about Yoga practices. Results indicated high levels of student acceptance, with the participants seeing the benefits of Ayurveda- and

Yoga-based practices in helping them in maintaining their well-being. The paper has found that SAHYP is contextually appropriate and useful, but needs to work on its long-term sustainability by facing a number of challenges such as insufficient funding, lack of proper infrastructure and lack of trained employees. Such health-promotive initiatives should be maintained by strengthening the community involvement, institutional support and coordination with the local government.

Bhusal, N., et al., (2024) have discussed globalization of Ayurveda in other regions like India, United States, Europe and Middle East. India being the pioneer of the system has established effective regulatory frameworks, clinical research infrastructures and policy support systems that will make Ayurveda an element within mainstream healthcare. Research points to Ayurvedic therapy- such as Panchakarma and herbs- as effective in treatment of chronic illnesses with reduced side effects. Ayurveda is gaining recognition in such nations as the U.S. and Germany as part of the complementary and alternative medicine, but the variability in regulations and the quality control problems are problematic. The increasing cultural acceptance in the Middle East and Southeast Asia has further promoted the adoption due to the joint research programs. Ayurveda has a potential worldwide, but its further adoption is inhibited by intermittent regulations, the absence of clinical testing on large scale, and unstandardized preparations.

Singh, A. (2025)highlights Ayurveda as an individualized medical system that is holistic in nature and aims at balancing dosha and mind-body health. Classical literature and contemporary research indicate that Ayurvedic methods that include herbal therapy, detoxification and lifestyle modification are applicable in the management of chronic illnesses and general well-being. Although there are certain limitations in the research, Ayurveda has a good preventive, immuno modulatory and anti-inflammatory potential. The increased attention to integrative medicine is occurring worldwide, so combining modern healthcare with Ayurveda makes the latter an even more applicable and sustainable system of complementary healthcare. Begum, S., et al., (2024) emphasise the importance of school health programs based on Ayurveda as a holistic method of healthcare delivery to a community. This school of thought focuses on the promotive, preventive, and curative services that would benefit the general population health. The strategy will use Ayurvedic concepts in school health programs in order to improve the health of students and the community at large by promoting healthy lifestyles among them in the long term.

Proposed System

The suggested scheme presents an integrative and computational paradigm, which analytically converts the traditional knowledge of Ayurveda diagnostics to determinable feature representations that can be used by machine learning to predict the disease. The system will combine conventional assessment values like Prakriti, Dosha imbalance, Agni, Nadi and Astavidha Pariksha with the current artificial intelligence methodology to develop a strong culturally aware predictive model (Gaikwad, S. 2025). The authors used five significant modules illustrated in figure 1 to make up the architecture of the proposed system they (1) Data Acquisition, (2) AFE (3) Machine Learning Model Development, (4) Evaluation and Validation, and (5) System Deployment and User Interface. Figure 1 below shows the work flow detail of proposed work.

Data Acquisition

The Ayurveda-based disease prediction system proposed has the Data Acquisition step as its foundation. During this stage, patient information will be gathered, formatted, cleaned, and formatted to act as input in the AFE module and ML models.

The data employed in this research has been gathered by visiting Ayurvedic clinics, wellness centers and consultation documents with the assistance of licensed Ayurvedic doctors (Vaidyas). The main aim of this step is to obtain descriptive Ayurvedic diagnostic data and also the symptoms and confirmed diagnosis labels (Gautam, A., et al., 2025). Five hundred four hundred and fifty-five raw records were gathered under different sources, 33 records were filtered out during the pre-processing phase because some of the major missing fields were found like key diagnostic fields (Nadi, Prakriti, Dosha), In complete symptom descriptions or final diagnosis labels. Finally a total of 512 patient records had been gathered in this study.

The dataset is made of three types of fields: Demographic and Background Fields containing 5 fields i.e., patent ID, Age, Gender, Region and lifestyle type. There are 15 parameters in Ayurvedic Diagnostic field: Vata, pitta, kapha, Nadi speed,depth, movement, Mala Quality, Mutra Quality, Strotas Quality, Dosha imbalance, Rasa status, Ritu-seasons effect. These parameters are in accordance with Astavidha Pariksha (eightfold exam), Prakriti and Nadi Pariksha.

Symptoms and clinical field (Headache, fat, body pain, fever or body temperature, Appetite Level, Sleep Quality, Digestion Issues, Skin Condition, Breathing Difficulty, Swelling, Mental Stress Level and Other Ayurveda-specific symptoms) 12 attributes. The symptoms assist in the mapping of Lakshana (clinical symptoms) to Dosha imbalance.

Ayurveda-Based Feature Engineering

The Ayurveda-Based Feature Engineering module is the main innovation of the desired disease prediction system. It can be used to convert the Ayurvedic diagnostic principles, which are usually qualitative, textual, and practitioner-based, into machine-readable numerical input in the form

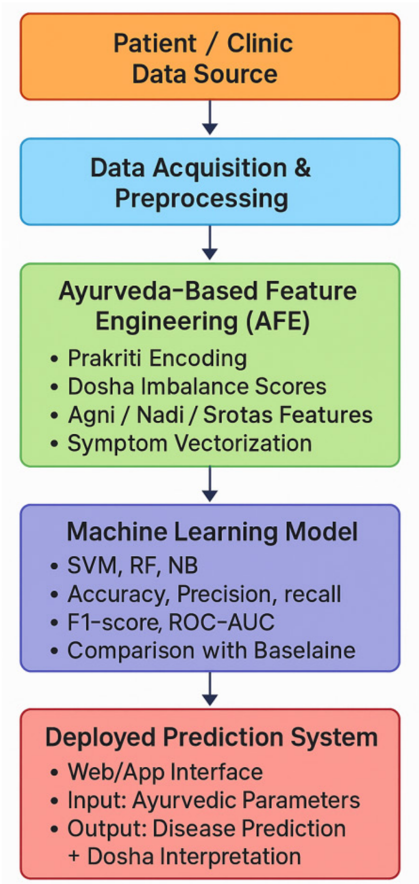

Figure 1: Flow diagram of proposed work

of a machine learning classification model, such as SVM, RF, and NB.

AFE makes certain that the wisdom of ancient diagnostics has a beneficial impact on predictive analytics but does not affect its interpretability or compliance with classical Ayurvedic logic. One of Ayurvedic diagnosis is Dosha balance, Prakriti, Agni, Nadi, Mala, Mutra, and clinical symptoms (Lakshanas). Such tests are subjective and descriptive in nature. Instead, ML algorithms need to have quantitative and standardized features. AFE bridges this gap by:

- Encodings of Ayurvedic assessments in text format into numerical formats.
- Populating weighted feature vectors of dosha dominance.
- Organizing symptomdosha relations.
- Lessening noise and inconsistency of practitioner-based assessments.
- Development of a comprehensive representation with an Ayurvedic view of human physiology.

Prakriti Encoding (Constitution Mapping)

Prakriti is the natural make-up of a patient: Vata, Pitta or Kapha, or a mix between them. As prakriti is a multi-valued and categorical factor, it is coded through one-hot vector define below in equation (1)

$$P=\left[P_v,P_p,P_k\right] \tag{1}$$

Here,

$$P_d=\begin{cases}1, & if\ dosha\ d\ is\ dominant\\ 0, & otherwise\end{cases} \tag{2}$$

For dual prakriti (e.g., Vata-Pitta):

$$P=\left[1,1,0\right] \tag{3}$$

Through this way A 3-dimensional numerical vector of the constitutional tendencies applied by all ML models can be able to obtain. Table 1 above shows the sample data values of prakriti.

Dosha imbalance

Doshas are paired up with symptoms (Lakshanas) and are related according to the traditional Ayurvedic texts including Charaka Samhita and Susruta Samhita.

Now There shall be

m symptoms expressed as:

$$S=\left\{s_1,s_2,\ldots\ldots s_m\right\} \tag{4}$$

Every symptom is associated with the dosha influence weights:

$$W=\begin{matrix} w_{v1} & w_{p1} & w_{k1}\\ w_{v2} & w_{p2} & w_{k2}\\ w_{vm} & w_{pm} & w_{km}\end{matrix} \tag{5}$$

Where, $w_{vd}=1$ when symptom causes Vata, 0 otherwise. Symptom presence is encoded as:

$$s_i=\begin{cases}1, & if\ symptom\ present\\ x, & otherwise\end{cases} \tag{6}$$

Dosha imbalance scores are calculated as:

$$D_v=\sum_{i=1}^{m}s_i.w_{vi} \tag{7}$$

$$D_p=\sum_{i=1}^{m}s_i.w_{pi} \tag{8}$$

$$D_k=\sum_{i=1}^{m}s_i.w_{ki} \tag{9}$$

The dosha imbalance vector is obtained as:

$$D=\left[D_v,D_p,D_k\right] \tag{10}$$

Table 1: Sample encoding of Prakriti encoding

Prakriti	*Vata*	*Pitta*	*Kapha*
Vata-Pitta	1	1	0
Pitta	0	1	0
Kapha	0	0	1

Table 2: Sample dosha encoding

Symptoms	*Vata*	*Pitta*	*Kapha*
Dry Skin	1	0	0
Burning Sensation	0	1	0
Heaviness	0	0	1

The end result of this process A 3-feature numerical vector [Vata_imbalance, Pitta_imbalance, Kapha_imbalance] can be achieved. Table 2 below presents the data of dosha imbalance.

Agni encoding

The strength of metabolic and digestive processes is determined by Agni. Agni scales are traced to numerical scale.

$$Agni = \begin{cases} 0, \ Madagni \\ 1, \ Vishamagni \\ 2, \ Tikshnagni \\ 3, \ Samagni \end{cases} \tag{11}$$

In this encoding A single ordinal feature numerically indicating metabolic strength in the end.

Nadi Pariksha Feature Extraction

The assessment of Nadi (pulse) is in terms of various characteristics that are utilized in this study like pulse speed, depth and movement. These features are transformed into three simple numerical characteristics. Feature Extraction using Nadi Pariksha is shown in Table 3.

Let Nadi attributes be:

$$Speed\ N_s, Depth\ N_d, Movement\ N_m \tag{12}$$

Encode them as:

$$\begin{aligned} N_s &\in \{0,1,2\} \in \{Slow, Medium, Fast\} \\ N_d &\in \{0,1,2\} \in \{Deep, Middle, Surface\} \end{aligned} \tag{13}$$

Both Movement Corresponds to dosha

$$N_m = \begin{matrix} [1, & 0, & 0,](Vata-type\,movement) \\ [0, & 1, & 0](Pitta-type\,movement) \\ [0, & 0, & 1](Kapha-type\,movement) \end{matrix} \tag{14}$$

The feature vector of Nadi is, therefore, the following:

$$N = \left[N_s, N_d, N_{mv}, N_{mp}, N_{mk}\right] \tag{15}$$

Table 3: Feature Extraction using Nadi Pariksha

Attribute	*Value*	*Encoding*
Speed	Medium	1
Depth	Middle	1
Movement	Snake-like	Pitta=1

Srotas, Mala, Mutra, and Other Diagnostic Parameters

Ayurvedic examinations (Astavidha Parika) comprises of Mala Quality, Mutra Quality, Srotas Quality. In this step Some categorical conviction will be coded into numerical mappings that create 3-5 supplementary features.

$$Srotas = \begin{cases} 0, \ Normal \\ 1, \ Mild\ obstruction \\ 2, \ Severe\ obstruction \end{cases} \tag{16}$$

$$Mala = \begin{cases} 0, \ Normal \\ 1, \ Hard \\ 2, \ Loose \end{cases} \tag{17}$$

$$Mutra = \begin{cases} 0, \ Clear \\ 1, \ Cloudy \\ 2, \ Frequent \end{cases} \tag{18}$$

Symptom Vectorization

The clinical symptoms are represented as weighted or binaries. They are converted to ordinal and binary numerical features as shown in Table 4.

All the symptoms will be coded as either:

Binary

$$s_i = 1 if\ present, 0\, is\, absent \tag{19}$$

Ordinal

$$s_i \in \{0,1,2,3,4\} \tag{20}$$

Weighted

Based on intensity:

$$s_i \in [0,1] \tag{21}$$

These form a symptom vector:

$$S = [s_1, s_2, \ldots . s_m] \tag{22}$$

Final Feature Vector Construction

Once all the above elements are processed, a single feature vector is constructed. The concatenation of all engineered features takes place:

$$F = [P, D, Agni, N, Srotas, Mala, Mutra, S] \tag{23}$$

In expended form:

$$F = \left[P_v, P_p, P_k, D_v, D_p, D_k, Agni, N_s, N_d, N_{mv}, N_{mp}, N_{mk}, Srotas, Mala\right] \tag{24}$$

Table 4: Symptom Vectorization

Symptoms	*Type*	*Example value*
Headache	Binary	1
Fatigue	Binary	1
Burning Sensation	Weighted	0.8
Appetite Level	Ordinal	2
Sleep Quality	Ordinal	1

Then when total engineered features d

$$F \in R^d \tag{25}$$

In this study work 25-35 features are normalized and inputted into ML algorithms as SVM, RF and NB algorithms.

Machine Learning Model

To improve the accuracy and interpretability of the disease prediction system based on Ayurveda, the proposed system uses three popular machine learning classifiers namely, RF, SVM and NB. These models work based on the numerically theorized Ayurvedic characteristics based on Prakriti assessment, dosha imbalances scores, Nadi parameters, Agni evaluation, and symptom vectors. All of the algorithms have their own advantages: the idea of Random Forest is good at learning the decision through ensemble, SVM is good at the separation of complicated, nonlinear health patterns, and Naive Bayes has the ability to learn the classification in a fast and probabilistic way that can be used in symptom-based diagnosis. Collectively, these models can help the system to interpret holistic Ayurvedic diagnostic data and produce data-intensive and precise disease forecasts in support of integrative and personalized medical care.

Integration of ML Models With Proposed System

Based on the AFE feature engineering module, the patient records are transformed to a numerical feature set:

$$F = [f_1, f_2, \ldots.. f_d] \in R^d \tag{26}$$

where Prakriti encoding is one of the features, the dosha imbalance scores are also part of the features, Agni, Nadi, Mala, Mutra, lifestyle indicators, and vectors of symptoms.

The multiclass classification is the machine learning task:

$$y = \{c_1, c_2, \ldots.. c_k\} \tag{27}$$

where c_k represents diseases such as Amlapitta, Sandhivata, Prameha, etc.

Thus the objective is,

$$y' = argmax\, P(c \mid F) \tag{28}$$

To approximate y' RF, SVM, and NB as predictive models.

Random Forest

Random Forest is an ensemble learning algorithm, which relies on decision trees + bagging. The process of Rf in proposed work is given in Figure 2 below. It is very effective in Ayurvedic datasets since:

- processes categorical and numerical mixed features.
- interpretable feature significance (e.g. dosha imbalance significance)
- RF constructs various decision trees and compounds their predictions.

Let,

$T = \{h_1, h_2, \ldots.. h_n\}$ and $F = ayurved\ feature\, vector$

Each tree predicts a class:

$$h_t(F) - \rightarrow c_j \tag{29}$$

Final prediction via majority voting:

$$y' = argmax \sum_{t=1}^{n} I(h_t(F) = c_j) \tag{30}$$

Where,

$I(.)$ is an indicator function.

Algorithm 1: RF in Proposed System

Input: Ayurveda feature matrix X, disease labels y

Output: class y' as Predicted disease

- for t=1 to N trees Draw bootstrap sample from x Select random subset of features Grow decision tree using Gini Index defined below:

$$Gini = 1 - \sum (p_i)^2 \tag{31}$$

- Store trained tree h_t For a new patient F Pass F into all trees Collect predictions $\{h_1(F), h_2(F), \ldots\ldots h_{N(F)}\}$
- Return majority voting result y'

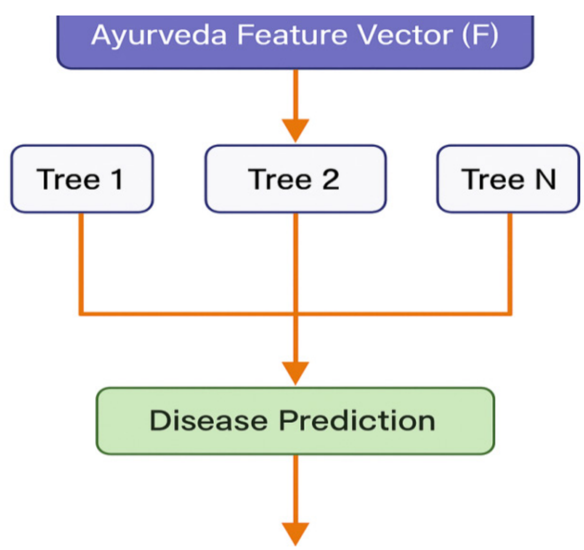

Figure 2: Flow diagram of RF for proposed work

Support Vector Machine

SVM is a margin based classifier that is appropriate in high dimensional biomedical data. It draws hyperplane dividing classes of diseases according to Ayurvedic features designed. The Proposed work flow with SVM classification is shown in Figure 3.

For two class case:

$$f(F) = w^T F + b \tag{32}$$

SVM optimizes:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{33}$$

Subject to:

$$y_i\left(w^T F_i + b\right) \geq 1 \tag{34}$$

For Non-linear Ayurvedic data, kernel trick is applied:

$$K\left(F_i, F_j\right) = \phi\left(F_i\right)^T \phi\left(F_j\right) \tag{35}$$

Common kernel used is,

$$K = exp - \gamma \left\|F_i - F_j\right\|^2 \tag{36}$$

Algorithm 2: SVM for Proposed System

Input: Ayurveda feature matrix X, labels y

Output: Predicted class y'

- Normalize continuous features (Agni, Dosha scores, symptoms)
- Choose kernel (RBF recommended)
- Train SVM by maximizing margin
- For new patient feature vector F:

Compute decision function using the below

$$y' = argmax_j\left(w_j^T F + b_j\right) \tag{37}$$

Output predicted disease

Navie Bayes

Naive Bayes is a Bayesian classifier, which is founded on the Bayes and conditional independence. The figure 4 shows the work flow of proposed work with NB.

- Useful for Ayurveda because:
- Training, which can be interpreted quickly.
- Does not work with continuous data, supports few symbols and categorical dosha.

Specification of the bayes theorem can be realised by specification:

$$P\left(c_j | F\right) = \frac{P\left(F | c_j\right) P\left(c_j\right)}{P(F)} \tag{38}$$

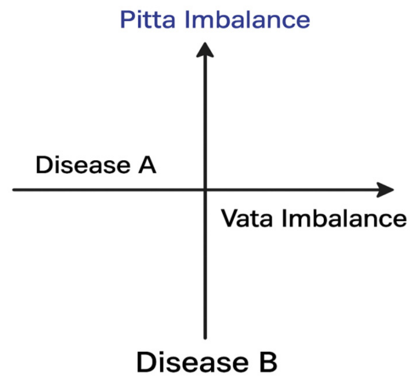

Figure 3: SVM for Proposed work

Since features assumed conditionally independent:

$$P\left(F | c_j\right) = \sum_{i=1}^{d} P(f_i \mid c_j) \tag{39}$$

Thus,

$$y' = arg \max_{c_j} \sum_{i=1}^{d} P(f_i \mid c_j) \tag{40}$$

Algorithm 3: NB for Ayurveda classification

Input: Feature matrix X, disease labels y

Output: Predicted disease y'

- Calculate prior probability

$$P\left(c_j\right) = \frac{count\left(c_j\right)}{total\ records} \tag{41}$$

- For each feature f_i

Compute conditional probability $P(f_i \mid c_j)$

- For new patient feature vector F

For each class c_j:

$$score_j = P\left(c_j\right) * \Pi P(F_i \mid c_j) \tag{42}$$

- Return class with highest $score_j$

Result and Discussion

The analysis of the results determines the efficiency of the presented AFE framework along with three ML models RF, SVM and NB in predicting the disease. The main aim is to find out the accuracy of these models in placing the patients in Ayurvedic disease types using the engineered features like Prakriti type, score in dosha imbalance, Agni levels, Nadi characteristics, lifestyle and symptom vectors. Table 5 below shows the parameter setup of proposed work.

Quality Parameters

The quality parameters (or evaluation metrics) are used to determine the quality of the ML models in the classification of diseases.

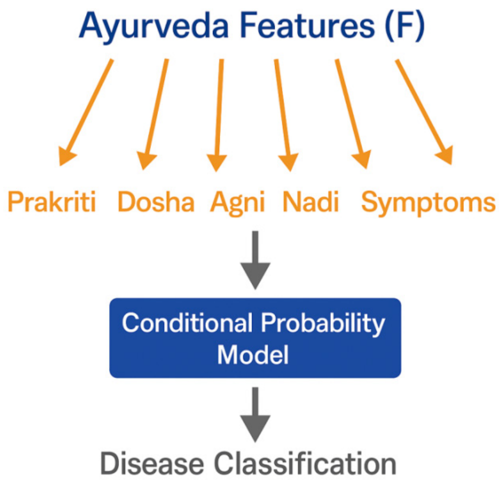

Figure 4: NB for proposed work

Accuracy

Measures overall correctness:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (43)$$

Precision

Precision is a performance measurement in classification to estimate the number of the correct cases of the positive cases predicted.

$$Precision = \frac{TP}{TP + FP} \quad (44)$$

Recall

Recall is a classification performance measure that is used to determine the ability of a model to recognize real positive cases.

$$Recall = \frac{TP}{TP + FN} \quad (45)$$

F1-Score

F1-Score is the harmonic mean of Precision and Recall providing one measure of model accuracy when false positives and false negative are both important.

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	RF, SVM, NB
Software environment	Python, Scikit-learn, Pandas, NumPy, Matplotlib

$$F1 = 2X \frac{Precision\ X\ Recall}{Precision + Recall} \quad (46)$$

ROC-AUC Curve

Measures probability model models differences between disease classes:

$$AUC = \sum_{0}^{1} TPR(FPR)\, d(FPR) \quad (47)$$

Performance Comparison

Three classifiers, including RF, SVM and NB, were trained and tested on the proposed engineered feature set after the application of the proposed AFE. It aimed to determine the predictability of these models to indicate the correct Ayurvedic disease classification according to the scores of Prakriti, dosha imbalance, Agni, Nadi, and symptom vectors.

The 512 records data was partitioned into 70 percent training and 30 percent testing and the experiment was repeated again with 5-fold cross-validation. Accuracy, Precision, Recall and F1-Score of the models were macro-averaged and presented as below.

Random Forest outperformed both of the other models compared based on Table 6 with a mean accuracy of about 95 and highest precision, recall and F1-score, which showed that the model has great predictive power with only a few false positives and false negatives. SVM was also doing well and achieved approximately 92 percent accuracy, which is indicative of the fact that the Ayurvedic feature space that has been engineered can be classified effectively with the help of margin. Naive Bayes retrieved the poorest accuracy of about 86, but it was still tolerable since it is simple and it is based on the assumptions of independence of features. The overall performance analysis is shown in Figure 5.

Effect of Ayurveda-Based Feature Engineering

In order to demonstrate the advantage of proposed AFE module, the outcome of AFE can be contrast the same classifiers trained without AFE (using raw demographic and simple symptom features only) and with AFE.

A comparative study of the classification accuracy of three ML models, including RF, SVM, and NB, with and without the proposed AFE is provided in the Figure 6 and Table 7. The performance of the models, in any case, is substantial with the addition of AFE, which illustrates the significance of Ayurvedic diagnostic characteristics in the prediction of the disease. Random Forest is best improved,

Table 6: Overall performance of classifiers with AFE

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
RF	94.8	95.2	94.5	94.7
SVM	91.6	92.1	91.4	91.6
NB	86.3	87.5	85.9	86.2

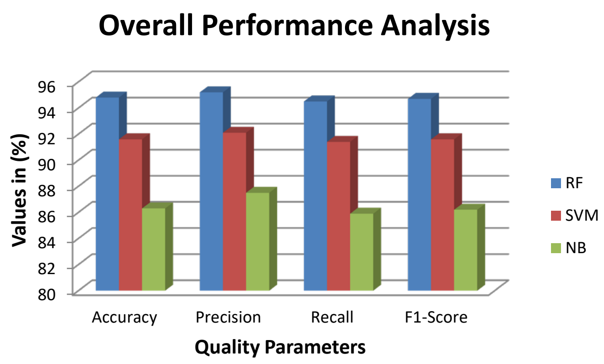

Figure 5: Overall performance analysis

as it achieves approximately 89% and then almost 96% which means it is capable of taking advantage of engineered features. The improved margin-based classification with transformed feature space also favours SVM, improving its performance by about 86 percent to 92 percent. Despite being the least predictive, Naïve Bayes, however, increases its prediction capability by about 80 per cent to almost 87 per cent, which validates the fact that even the simplistic predictive models are more predictive when they are augmented with structured Ayurvedic characteristics. On the whole, the graph shows clearly that AFE significantly improves the accuracy of the model, which justifies the usefulness of the offered method.

Confusion Matrix

The following are the sample confusion matrices of the three classifiers that are to be employed in the Ayurveda based disease predictive system. These are the degrees of the suitability of each of the models in terms of categorizing four classes of diseases namely: Amlapitta, Sandhivata, Prameha and Jwara following the application of AFE.

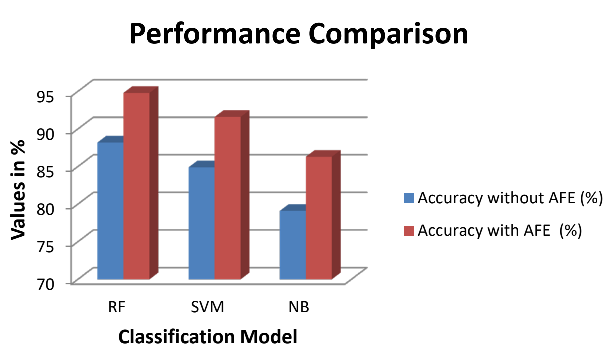

Figure 6: Performance comparison

Table 7: Performance analysis of proposed work AFE vs without AFE

Model	*Accuracy without AFE (%)*	*Accuracy with AFE(%)*
RF	88.2	94.8
SVM	84.9	91.6
NB	79.1	86.3

Confusion Matrix for RF

According to the confusion table in Figure 7, the Random Forest model is outstanding as its accuracy is approximately 95 percent. The majority of predictions are on the diagonal which means a very high rate of accuracy in the classification of all the four classes of diseases. The highest performance is seen in amlapitta and Sandhivata (135 and 133 correct predictions respectively and very few misclassifications). Prameha also shows good separation with correct predictions of 125 but few samples are mixed with Jwara. Jwara has good classification results having 92 correct predictions and low level of overlap with other classes. In general, the low off-diagonal values reveal that the model has low false positives as well as false negatives which represent high robustness and reliability of the model on all classes.

Confusion Matrix for SVM

The SVM model in Figure 8 has an accuracy of approximately 89 which is good with respect to overall performance although slightly lower than that of the Random Forest. The confusion matrix indicates that most of the predictions are along the diagonal line implying that the classifier is able to predict the majority of the samples in all four types of diseases. The high percentages in correct prediction among amlapitta and sandhivata (130 and 125 respectively) with some misclassifications dispersed in other classes. Prameha has also good performance of 118 correct predictions but there is some overlap with Sandhivata and Jwara. Jwara documents 86 accurate classifications yet somewhat more confusions with Prameha. All in all, SVM works well on issues of class boundaries, yet it exhibits moderate misclassification in scenarios where the patterns of symptoms are similar hence the relative low accuracy compared to RF.

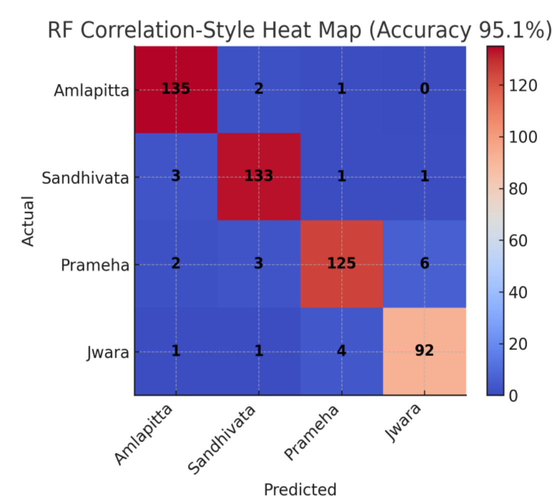

Figure 7: Confusion matrix for RF

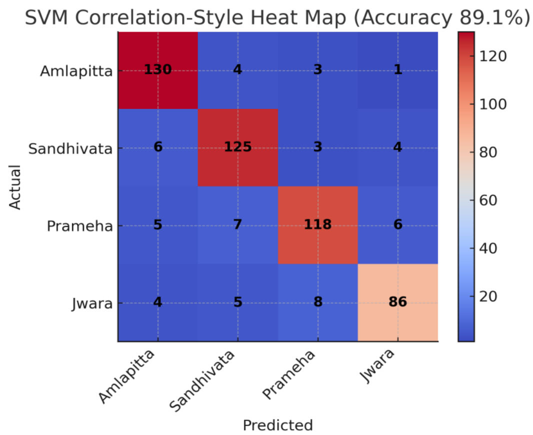

Figure 8: Confusion matrix of SVM

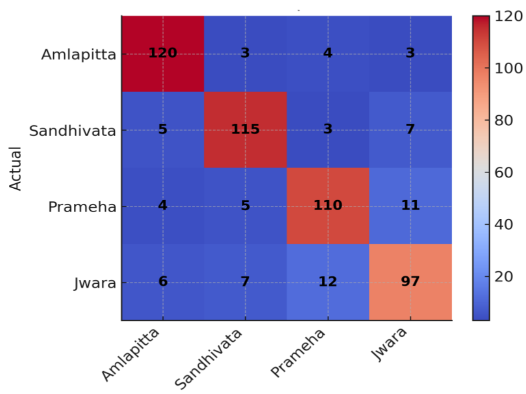

Figure 9: Confusion matrix of NB

Confusion Matrix for NB

The confusion matrix of NB in Figure 9 indicates that the model has a strong performance regarding all the four disease classes with low misclassification.

Amlapitta, Sandhivata, and Prameha have very high numbers of correct predictions (120, 115 and 110 respectively), which shows that the classifier can differentiate between their patterns of features. Jwara also performs well with accuracy of 97 correct predictions although it gives a little bit more overlap with Prameha and Sandhivata. The off-diagonal values are not high across the matrix indicating that false positive and false negative are also minimal. The general impression of the model is that, the model has strong and dependable classification abilities, correctly classifying most of its samples in all the categories.

Conclusion

This research has come up with a useful disease forecasting framework using Ayurveda as its basis by combining the traditional characteristics of diagnosis with the contemporary machine learning models. The models performed well using engineered Ayurvedic qualities like Prakriti, Dosha imbalance, Nadi, Agni and symptom patterns with the highest accuracy of Random Forest (94.8%), and the next two SVM (91.6) and Naive Bayes (86.3). The use of confusion matrix and heat-map analysis confirmed that there was minor misclassification and high feature discriminability between different disease categories. In general, the study shows that an integration of Ayurvedic knowledge and machine learning can substantially enhance the early detection of diseases and aids in the applications of personalized and holistic healthcare.

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 consistent support on our research work.

References

Ahuja, N., Bhardwaj, P., Pathania, M., Sethi, D., Kumar, A., Parchani, A., ... & Phadke, A. (2024). Yoga Nidra for hypertension: A systematic review and meta-analysis. *Journal of Ayurveda and Integrative Medicine, 15*(2), 100882. doi.org

Balkrishna, A., Sharma, N., Srivastava, D., Kukreti, A., Srivastava, S., & Arya, V. (2024). Exploring the safety, efficacy, and bioactivity of herbal medicines: Bridging traditional wisdom and modern science in healthcare. *Future Integrative Medicine, 3*(1), 35–49.

Baragi, U. C., & Ganer, J. M. (2025). Innovative Ayurvedic formulations: Bridging tradition with modern science & technology. *Journal of Ayurveda and Integrated Medical Sciences, 10*(6), 1–3.

Begum, S., Begum, S., Venkatakrishna, K. V., & Asha, S. A. (2023). Role of Ayurveda in school health. *International Journal of Medicine and Public Health*.

Bhusal, N., Sapkota, S., Upadhyay, V., Thakur, S., & Panthi, S. (2024). Relevancy and sustainability of Ayurveda education and alternative medicine in Nepal. *Curriculum Development Journal*.

Chaturvedi, S., Kumar, N., Tillu, G., & Patwardhan, B. (2021). Research, biomedicine and Ayurveda: From evidence-based medicine to evidence-informed healthcare. *Indian Journal of Medical Ethics, 6*(4), 301–305.

Chouhan, P. N., & Joshi, A. (2025). Integrative Ayurveda strategies for global health restoration: Bridging tradition with scientific innovation. *Ayurveda Rising*, 124.

Gaikwad, S. (2025). Exploring the relevance of indigenous medicinal practices in India: A comprehensive analysis. *International Journal of Sciences and Innovation Engineering, 2*(4), 37–50.

Gandhmal, D. P., Moin, S., Perampalli, A., Nadaf, S. S., Chitkoti, S. S., Raut, R. M., & Kurle, N. A. (2024). *AYURAID: Revolutionizing Ayurvedic learning and practice using OpenAI and TF-IDF.* [Unpublished manuscript or Preprint].

Gautam, A., Asutkar, S., Nagpure, D., & Dagar, M. (2025). Ayurveda meets surgery: Exploring ksharasutra's role in modern anorectal disorders. *Multidisciplinary Reviews, 8*(7), Article 2025216.

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

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

Nayak, S. K., et al. (2025). A decision tree-based approach for disease prediction and Ayurvedic drug recommendation. *Ingénierie des Systèmes d'Information*. Advance online publication.

Singh, A. (2025). The role of Ayurvedic principles in holistic health management: A comprehensive review of ancient practices and modern applications. *International Journal of Environmental Sciences*. Advance online publication.

Umadevi, V., & Kamath, V. (2024). Assessing classification approaches for categorizing Ayurvedic herbs. *Multimedia Tools and Applications, 84*, 6711–6735.

Vayadande, K., Sukte, C. D., Bodhe, Y. U., 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*.