Development of Explainable Artificial Intelligence Models for Ayurvedic Diagnosis and Decision Support

Dr. Rahul V. Kadam[1]*

*[1]Dean, Vice Principal and Professor, Department of Shalyatantra, College of Ayurved and Hospital, Bharati Vidyapeeth (Deemed to be University), Pune, Maharashtra, India, rahul.kadam@bharatividyapeeth.edu

Corrosponding Author
Dr. Rahul V. Kadam
Dean, Vice Principal and Professor, Department of Shalyatantra, College of Ayurved and Hospital, Bharati Vidyapeeth (Deemed to be University), Pune, Maharashtra, India, rahul.kadam@bharatividyapeeth.edu

Received: 06th Jan 2026; Revised: 18th Jan 2026; Accepted: 21st Feb 2026; Available online: 1st March 2026

Abstract
Ayurvedic diagnosis is a multidimensional clinical reasoning process that integrates Prakriti, Vikriti, Dosha-Dushya involvement, Agni, Ama, Srotas, Bala, Satmya, Ahara-Vihara and temporal context. Computational models may improve reproducibility and data organization, but black-box predictions are incompatible with the interpretive, relational and physician-mediated character of Ayurvedic decision-making. To develop an explainability-ready model architecture for Ayurvedic diagnosis and decision support and to quantify the performance, transportability and feature structure of available Prakriti classification evidence without fabricating participant-level results. A methodological secondary-data study was conducted using two traceable sources: published confusion matrices and feature-selection outputs from a clinician-labelled extreme-Prakriti machine-learning study, and aggregate descriptive statistics from the Prakriti200 bilingual questionnaire dataset. Confusion matrices were reconstructed to calculate accuracy, balanced accuracy, macro-F1, Cohen kappa and Matthews correlation coefficient. Thirty-one features jointly retained by LASSO, elastic net and random forest were recoded into clinically interpretable Ayurvedic observation domains. An explainable decision-support architecture was then specified, comprising transparent baseline models, calibrated ensembles, global and local explanations, uncertainty estimates, counterfactual prompts, provenance controls and mandatory Ayurveda-physician review. Results: In external testing from the Western Indian Vadu cohort to North India, overall accuracy was 90.6% for LASSO, 93.8% for elastic net and 92.7% for random forest. Reciprocal North India-to-Vadu validation yielded 78.9%, 85.0% and 91.2%, respectively, demonstrating direction-dependent transportability. Random forest sensitivity differed by class and direction, with the largest reduction observed for Pitta in one validation setting. The 31 consensus features spanned skin/temperature/appearance (7), motor/speech (6), morphology/structure (5), psychobehavioural/cognitive (5), strength/work (4), and digestive/excretory/metabolic domains (4). Prakriti200 contained 200 participants and showed a Pitta-dominant distribution, but was strongly age-skewed toward young adults.

Keywords: Ayurveda; explainable artificial intelligence; Prakriti; clinical decision support; machine learning; Ayurgenomics; interpretable models; digital health

How to cite this article: Kadam RV. Development of Explainable Artificial Intelligence Models for Ayurvedic Diagnosis and Decision Support. Int J Drug Deliv Technol. 2026;16(36s): 1112-1121. DOI: 10.25258/ijddt.16.36s.132

1. Introduction

Ayurveda conceptualizes health and disease through a relational model rather than a single-variable taxonomy. The physician observes the person, the disorder and the context together, considering constitution, current Dosha disturbance, tissue and channel involvement, digestive and metabolic capacity, strength, suitability, age, geography, season, diet, behaviour and temporal progression. Classical descriptions of Prakriti provide a relatively stable constitutional frame, whereas Vikriti denotes the dynamic departure from that baseline. Diagnostic reasoning therefore depends on both persistent phenotype and present-state deviation [1-3]. A digital system that predicts only a label such as Vata, Pitta or Kapha while ignoring uncertainty, mixed constitutions and contextual modifiers would compress this reasoning into an output that is convenient but clinically impoverished.

Variation in observation and interpretation is a recognized challenge in Prakriti assessment. Prototype and standardized assessment tools have attempted to improve inter-rater agreement, item definition and psychometric structure [4,5]. These efforts are important because a machine-learning model can reproduce only the consistency and validity of the labels used for training. When expert adjudication is poorly documented, automated accuracy may represent agreement with a local questionnaire rather than fidelity to Ayurvedic diagnosis. The ground-truth problem is especially important for dual-Dosha and Sama-Dosha constitutions, where forced single-class labels may conceal clinically meaningful ambiguity.

Machine learning has nevertheless demonstrated that multidimensional phenotypic observations can recover clinically assigned extreme Prakriti classes. Tiwari and colleagues showed that three constitution groups emerged from unsupervised analysis and that supervised models based on reduced feature sets could classify participants across geographically distinct cohorts [6]. The broader Ayurgenomics literature has linked extreme constitutional phenotypes with gene-expression, biochemical, high-altitude adaptation and microbiome differences [7-10]. These findings do not prove that a computational label is equivalent to a complete Ayurvedic examination, but they establish that structured phenotype data contain reproducible patterns that can be investigated quantitatively.

Digital measurement is also expanding beyond questionnaires. Pulse-sensing systems and recent reviews of sensor-based Nadi Pariksha illustrate attempts to transform temporal waveform characteristics into repeatable signals [11,12]. Public data resources are beginning to appear, including a bilingual 24-item dataset of 200 Prakriti assessments [13]. At the same time, recent reviews of assessment tools continue to identify inconsistent item wording, variable scoring, limited external validation and incomplete psychometric evidence [14-15]. Consequently, the immediate scientific need is not simply a more complex classifier. It is a transparent modelling framework that makes disagreements, missing data, uncertainty and dataset limitations visible.

International policy has placed artificial intelligence and traditional medicine within the same emerging governance agenda. The 2025 WHO mapping report describes uses of AI for knowledge digitization, research, diagnosis, personalization and decision support in traditional medicine, while emphasizing data quality, interoperability, accountability and protection of traditional knowledge [16]. Policy analysis further stresses that AI should augment human-centred care rather than displace practitioners or decontextualize traditional systems [17]. Reviews of artificial intelligence across traditional, complementary and integrative medicine similarly identify opportunities in pattern recognition, evidence synthesis and individualized care, but warn of bias, opaque validation and poor clinical integration [18-19]. In India, the ICMR ethical framework requires scientific validity, transparency, privacy, fairness, accountability, informed consent and continuing oversight throughout the AI life cycle [20]. WHO's broader guidance for AI in health reinforces autonomy, safety, inclusiveness, responsibility and sustainability [21].

Explainable artificial intelligence (XAI) addresses part of this challenge by making model behaviour intelligible to clinicians and patients. Model-agnostic methods such as LIME create local approximations around an individual prediction [22], whereas SHAP distributes the difference between a prediction and a reference value across input features using a game-theoretic framework [23]. Transparent models, including regularized generalized linear models and carefully constrained trees, may be preferable where their discrimination is clinically adequate. Random forests can capture nonlinear interactions but need additional explanation and calibration [24]. LASSO and elastic net provide sparse, stable baselines that are attractive for structured questionnaires [25]. Reporting and bias-assessment standards such as TRIPOD+AI and PROBAST+AI now demand clear definition of data sources, outcomes, sample size, missingness, performance, fairness and external validation [26,27]. Early clinical evaluation should also follow human-factor guidance such as DECIDE-AI, and later trials require CONSORT-AI and SPIRIT-AI reporting [28-30].

2. Literature Review

The classical diagnostic foundation. Ayurvedic clinical reasoning begins with the inseparability of the patient and the disease process. Prakriti influences expected physical, physiological and psychological tendencies, but it is not itself a diagnosis. Vikriti, Nidana, Purvarupa, Rupa, Upashaya and Samprapti provide a dynamic account of causation, prodrome, manifestation, therapeutic response and pathogenesis [1-3]. Rogi Pariksha estimates the person's strength and treatment tolerance, while Roga Pariksha defines the active disorder. A computational representation must therefore distinguish stable baseline attributes from state-dependent findings and from contextual exposures. Mixing them into a single feature vector without temporal labels can produce explanations that are statistically correct but clinically misleading.

Prakriti is attractive as an initial computational target because it is described through observable traits and is relatively stable after constitution is established. However, assessment instruments differ substantially. The Prototype Prakriti Analysis Tool was developed to reduce ambiguity and investigate inter-rater validity [4]. A later standardized scale used structured development procedures to formalize psychosomatic constitution assessment [5]. These studies show that digitization should begin with item ontology and scoring validity, not with an algorithm. The model cannot repair a vague question, an unstable observer judgment or an outcome defined by the same unvalidated scoring rule used as input.

Machine-learning evidence in Prakriti classification. The Tiwari et al. study remains a key demonstration because its labels were assigned through detailed assessment by Ayurveda physicians and its models were externally examined in a second region [6]. From 10,100 screened people, 528 underwent detailed evaluation, and 147 healthy participants with extreme Prakriti were selected from the Vadu cohort. Feature selection used LASSO, elastic net and random forest, producing 39, 61 and 59 features, respectively, with 31 variables common to all methods. This intersection is valuable for explainability because it identifies a compact group of observations repeatedly considered informative by algorithms with different assumptions.

Biological and translational context. Ayurgenomics studies have reported molecular correlates of extreme Prakriti phenotypes, including differences in gene expression and biochemical measures [7]. Genetic analyses using constitution-defined groups have contributed to investigation of high-altitude adaptation and EGLN1 variation [8]. Gut microbiome research has also reported constitution-associated microbial signatures in a Western Indian rural population [9]. These findings support the research value of precise phenotyping, but they do not justify deterministic claims that a particular gene, metabolite or microorganism defines a Dosha. The appropriate computational use is probabilistic stratification with uncertainty and explicit provenance.

A translational Ayurgenomics framework emphasizes that traditional phenotyping can generate testable hypotheses for precision and integrative medicine [10]. XAI is especially relevant in this setting because explanations can connect predictive features to clinical observations and biological measurements without collapsing one knowledge system into the other. The system should permit parallel views: an Ayurvedic view organized by Dosha, Guna, Dhatu, Srotas, Agni and Bala; a biomedical view organized by measurable variables; and a modelling view organized by contribution, uncertainty and validation. Equivalence should not be presumed where it has not been established.

Sensor-derived evidence. Nadi Pariksha is often proposed for digitization because pulse waveforms can be sampled repeatedly and analysed computationally. Nadi Tarangini represented an early pulse-based diagnostic system, demonstrating the feasibility of electronic acquisition [11]. A 2024 review summarized sensor configurations and technological developments but also underscored the need for standard acquisition protocols, device calibration, placement consistency, reference labels and clinical validation [12]. For XAI, sensor features should be expressed in clinically interpretable terms such as waveform amplitude, timing, variability and response to controlled conditions rather than as unexplained latent embeddings alone.

Emerging datasets. Prakriti200 provides a bilingual questionnaire resource with 24 mandatory items, automated Dosha scoring and 200 participants [13]. Its transparent column structure supports educational and computational reuse. Nevertheless, 81% of participants were aged 18-25 years and the label distribution was Pitta-dominant. These characteristics create a realistic example of dataset shift and label imbalance. A model trained on this resource could learn features of a young student population or the scoring rule itself. Its outputs should not be represented as clinician-validated diagnosis unless compared with independent physician assessment.

Recent scoping and critical reviews of Prakriti assessment instruments have highlighted the proliferation of tools without uniform definitions, robust psychometric testing or multicentre validation [14-15]. The same Sanskrit term may be operationalized through different English or regional-language phrases. Some features are self-reported, while others require observation or palpation. XAI cannot be clinically meaningful unless the explanation vocabulary is mapped back to a controlled terminology. Each variable needs a definition, data type, observation method, time frame, allowable values, evidence source and relation to classical constructs.

Artificial intelligence in traditional medicine. WHO's mapping exercise describes growing use of machine learning, language models, image analysis and knowledge systems in traditional medicine [16]. Potential applications include digitizing textual knowledge, assisting diagnosis, supporting personalized recommendations, monitoring quality and accelerating research. Governance analysis cautions that data may encode historical inequity, cultural decontextualization or inappropriate commercialization of traditional knowledge [17]. A responsible Ayurveda system must therefore document who supplied the data, who defined the labels, how community knowledge is governed and who benefits from deployment.

Integrative medicine reviews identify a broad opportunity for AI but also a weak evidence base for real-world clinical benefit [18-19]. Many reports emphasize algorithmic accuracy while omitting calibration, clinical utility, subgroup performance and workflow effects. A decision-support model can appear accurate yet be unsafe if it is confidently wrong in rare constitutions, if it recommends treatment outside the clinician's scope, or if users over-trust a visually persuasive explanation. Interpretability is not a substitute for validation; it is one component of a safety case.

Explainability methods. LIME explains a single prediction by perturbing nearby inputs and fitting a simple local model [22]. It is intuitive but can be unstable when features are correlated or perturbations create implausible Ayurvedic combinations. SHAP provides local and global contribution values and supports summary, dependence and interaction views [23]. Its baseline population must be chosen carefully: an explanation relative to a young urban dataset may differ from one relative to an age- and region-matched reference. Neither method creates causal evidence. Feature attribution means that a variable influenced the model, not that changing the variable will change constitution or disease.

Model classes. Random forests handle nonlinearities and interactions and performed strongly in reciprocal regional validation in the available Prakriti study [6,24]. Regularized multinomial regression offers sparse coefficients and is easier to audit [25]. An explainable model-development programme should compare both. A transparent model should be preferred when performance and calibration are similar. More complex ensembles may be retained for research or second-line review, with SHAP or

equivalent explanations, monotonicity constraints where appropriate and clinician-readable feature grouping.

Reporting and evaluation. TRIPOD+AI requires complete description of the intended use, target population, predictors, outcome, modelling process and performance [26]. PROBAST+AI focuses attention on bias in participants, predictors, outcome and analysis [27]. DECIDE-AI adds early-stage clinical workflow evaluation, including human factors and unintended consequences [28]. If a system progresses to comparative trials, CONSORT-AI and SPIRIT-AI require documentation of algorithm version, input handling, interaction with clinicians, errors and analysis [29,30]. These standards are directly relevant to Ayurveda because diagnostic labels are observer-dependent and clinical use depends on how explanations alter physician reasoning.

Table 1. Ayurvedic diagnostic constructs and explainable computational representation

Ayurvedic construct	Candidate structured observations	Explainable output	Clinical safeguard
Prakriti	Long-standing body frame, skin tendency, appetite pattern, sleep, temperament, bowel habit	Relative support for Vata/Pitta/Kapha; stable-trait contributors and contradictions	Do not infer constitution from current symptoms alone
Vikriti	Recent changes in symptoms, function, appetite, sleep, pain, temperature, elimination	State-deviation profile with timestamps and trend	Separate from Prakriti and consider season, disease and medication
Agni and Ama	Appetite regularity, digestion, coating, heaviness, bowel pattern, post-meal symptoms	Evidence summary and missing-observation prompts	No automated treatment recommendation
Bala and Satmya	Strength, endurance, recovery, tolerance, habitual diet and exposures	Treatment-tolerance context for clinician review	Requires physician interpretation and contraindication checks
Dosha-Dushya-Srotas	Symptoms, tissues, sites, channel signs, chronicity and precipitating factors	Structured differential and Samprapti map	Display uncertainty and alternative pathways
Upashaya/Anupashaya	Observed response to diet, activity, climate or prior interventions	Response-linked evidence, clearly distinguished from causal proof	Avoid causal claims from uncontrolled observations

3. Materials and Methods

Study design. This was a methodological secondary-data analysis and explainable decision-support framework-development study. It had two linked components: (i) quantitative reconstruction of performance metrics from published Prakriti classification confusion matrices and descriptive analysis of an openly documented questionnaire dataset; and (ii) specification of a model architecture, explanation protocol and validation pathway suitable for future prospective research. No new patient recruitment, intervention or clinical outcome generation was undertaken. The paper therefore avoids presenting simulated predictions as observed clinical results.

Ayurvedic target definition. The initial target was Prakriti-oriented diagnostic support because it has a comparatively stable conceptual definition, structured

phenotypic descriptors and published machine-learning data [1-6]. The proposed architecture was designed to be extensible to Vikriti and disease decision support, but only after temporal and clinical labels are separately validated. The intended output is not a prescription. It is a structured summary of supporting and contradictory observations, probability or score distribution, uncertainty, missing information, red flags and suggested next examination steps for review by a qualified Ayurveda physician.

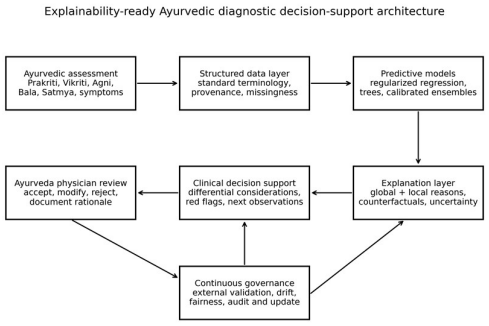

Figure 1. Explainability-ready architecture for Ayurvedic diagnostic decision support

Original framework developed in this study; prediction is separated from physician-authorized action.

Data sources. Source A was the peer-reviewed study by Tiwari et al., including supplementary tables that reported confusion matrices for a Western Indian Vadu cohort, a North Indian external-validation cohort and reciprocal validation [6]. Source B was the Prakriti200 data descriptor, which reported aggregate demographics, trait frequencies and constitution-label counts for 200 questionnaire respondents [13]. The participant-level data from Source A were not publicly downloadable without request, and no claim of participant-level retraining is made. For Source B, only aggregate statistics reported in the descriptor were used in this paper; the graphs were redrawn from those published counts.

Eligibility and extraction. A data element was eligible when its value could be traced to a published table, supplementary file or explicit dataset descriptor. Extracted items included sample size, recruitment setting, age and sex distribution where available, labelling process, number of questionnaire features, feature-selection counts, final consensus features, confusion-matrix cell counts and constitution-label frequencies. Values described only narratively without sufficient numerical detail were not used for quantitative reconstruction. The three-class order was Kapha, Pitta and Vata, matching the supplementary matrices.

Reconstruction of performance. Published matrices were structured with predicted classes in rows and true classes in columns. Cell counts were expanded computationally into paired true and predicted labels. Overall accuracy was defined as the proportion of correct classifications. Balanced accuracy was the unweighted mean of class sensitivities. Macro-precision, macro-recall and macro-F1 were calculated by averaging class-specific values. Cohen kappa quantified agreement beyond chance, while the multiclass Matthews correlation coefficient provided a correlation-based summary robust to unequal class frequencies. Metrics were calculated separately for Vadu-trained models tested in North India and North-India-trained models tested in Vadu. Binary extreme versus non-extreme matrices were also summarized.

Consensus feature analysis. The 31 variables selected by LASSO, elastic net and random forest in Source A were transcribed from the supplementary feature table [6]. To improve clinical interpretability, the present authors performed a secondary domain coding. Each variable was assigned to one of six observation domains: morphology and structure; skin, temperature and appearance; motor and speech; strength and work; psychobehavioural and cognitive; or digestive, excretory and metabolic. This coding was descriptive and did not alter the original model. It should be viewed as a proposed explanation layer requiring expert consensus before use.

External validation and fairness. Model evaluation should be nested within sites to avoid leakage and should include a geographically independent test set. Performance should be reported by sex, age group, language, region, examiner, pure versus mixed constitution and missingness pattern. Calibration plots and Brier scores should accompany discrimination. An out-of-distribution detector should compare the current case with training ranges. Fairness analysis should not force equal error rates where clinically inappropriate, but any systematic difference must be identified, investigated and communicated. The ICMR and WHO ethical principles provide the governance baseline [20,21].

Human factors and prospective evaluation. The model is intended for qualified Ayurveda practitioners and supervised trainees. Early evaluation should measure whether the explanation improves completeness, agreement and documentation without increasing automation bias. The clinician must be able to reject the output and provide a reason. Alert frequency, time burden, misunderstanding of probabilities and over-reliance should be studied according to DECIDE-AI [28]. Later comparative studies should specify the algorithm version and update policy in accordance with SPIRIT-AI and CONSORT-AI [29,30].

4. Results

Source characteristics. The clinician-labelled machine-learning evidence began with a large screening programme and selected 147 healthy participants with extreme Prakriti from a genetically homogeneous Western Indian population. The extreme set comprised 66 Vata, 35 Pitta and 46 Kapha participants. An independent North Indian cohort of 96 participants was used for external validation. The final phenotyping instrument contained 133 variables, of which 106 were common across cohorts [6].

Prakriti200, by contrast, contained 200 questionnaire respondents, 24 mandatory bilingual items and automated rule-based scoring. Ages ranged from 15 to 72 years, but 81% were 18-25 years; 127 participants were female and 73 male [13]. Table 2 summarizes the contrast between these sources.

Table 2. Characteristics and limitations of the two quantitative data sources

Characteristic	Tiwari et al. clinician-labelled study [6]	Prakriti200 descriptor [13]
Primary purpose	ML recapitulation of clinician-assigned extreme Prakriti	Reusable bilingual questionnaire dataset
Sample	147 extreme-Prakriti participants; external North Indian cohort n=96	200 respondents
Labels	Clinician-assigned extreme Kapha, Pitta or Vata	Automated rule-based pure and dual-Dosha labels
Features	133 final variables; 106 common across cohorts	24 mandatory questionnaire items plus metadata and scores
Validation	Internal holdout, external regional and reciprocal validation	Data completeness and rule-scoring checks; no independent clinical validation reported
Strength	Physician labels and external cohort	Accessible structure, bilingual instrument, no missing questionnaire items
Key limitation	Extreme phenotypes and modest validation samples	81% aged 18-25; student-heavy and Pitta-dominant distribution

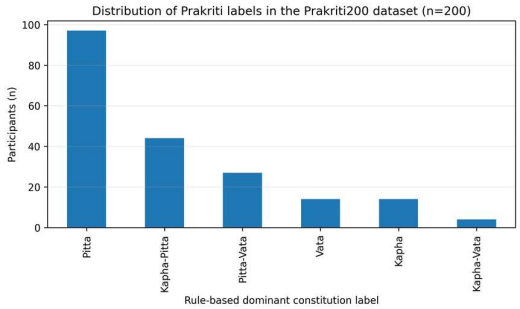

Figure 2. Prakriti label distribution in the Prakriti200 dataset

Redrawn from aggregate counts reported by Singh and Singh [13].

Cross-regional multiclass performance. Reconstruction of the external-validation matrices produced overall accuracies of 90.6% for LASSO, 93.8% for elastic net and 92.7% for random forest when models developed in Vadu were tested in North India. Balanced accuracies were 90.2%, 93.4% and 92.2%, and macro-F1 values were 90.6%, 93.7% and 92.5%, respectively. Cohen kappa ranged from 0.858 to 0.905. These results confirm strong discrimination in the reported external cohort [6].

Reciprocal validation was more variable. When North Indian data were used for model development and the 147-person Vadu cohort for validation, accuracies were 78.9% for LASSO, 85.0% for elastic net and 91.2% for random forest. Corresponding balanced accuracies were 82.1%, 86.2% and 91.9%, with macro-F1 values of 79.2%, 84.6% and 90.6%. The decrease was largest for LASSO and smallest for random forest. Figure 3 displays the direction-dependent accuracy, and Table 3 provides the reconstructed metrics.

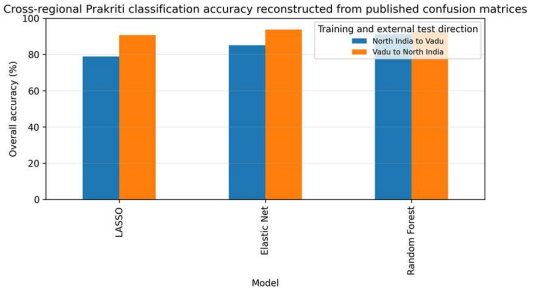

Figure 3. Cross-regional accuracy by model and validation direction

Calculated from published supplementary confusion matrices in Tiwari et al. [6].

Table 3. Reconstructed multiclass external-validation performance

Validation direction	Model	N	Accuracy (%)	Balanced accuracy (%)	Macro-F1 (%)	Cohen kappa	MCC
Vadu to North India	LASSO	96	90.6	90.2	90.6	0.857	0.859
Vadu to North India	Elastic Net	96	93.8	93.4	93.7	0.905	0.906
Vadu to North India	Random Forest	96	92.7	92.2	92.5	0.889	0.892
North India to Vadu	LASSO	147	78.9	82.1	79.2	0.688	0.716
North India to Vadu	Elastic Net	147	85.0	86.2	84.6	0.774	0.788
North India to Vadu	Random Forest	147	91.2	91.9	90.6	0.865	0.869

Class-specific sensitivity. For Vadu-to-North-India testing, random forest sensitivity was 100.0% for Kapha, 79.3% for Pitta and 97.4% for Vata. In reciprocal North-India-to-Vadu testing, random forest sensitivity was 93.5% for Kapha, 94.3% for Pitta and 87.9% for Vata. Figure 4 shows that a single overall accuracy can conceal class- and direction-specific changes. This matters for clinical explainability: a system should state the performance relevant to the predicted class and population, not only a pooled score.

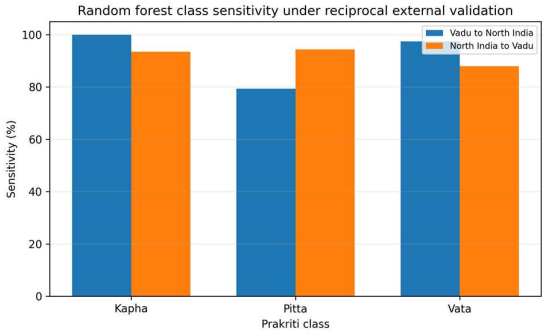

Figure 4. Random forest class sensitivity under reciprocal regional validation

Calculated from published supplementary confusion matrices in Tiwari et al. [6].

Extreme versus non-extreme classification. Published binary matrices contained 26 validation observations per model [6]. Reconstructed overall accuracy was 92.3% for LASSO, 96.2% for elastic net and 92.3% for random forest. These high values suggest that the selected phenotype features distinguished clearly defined extreme constitutions from non-extreme profiles in that sample. They do not establish equivalent performance for the full spectrum of dual-Dosha and Sama-Dosha constitutions.

Consensus features. LASSO selected 39 variables, elastic net 61 and random forest 59, with 31 variables common to all three methods [6]. Secondary domain coding assigned seven to skin, temperature and appearance; six to motor and speech; five to morphology and structure; five to psychobehavioural and cognitive observations; four to strength and work; and four to digestive, excretory and metabolic observations. Figure 5 displays the domain counts, while Table 4 lists each feature. The breadth of domains supports a multisystem interpretation of Prakriti rather than reliance on a single visible trait.

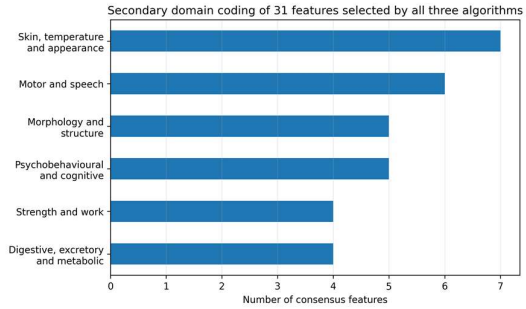

Figure 5. Distribution of 31 consensus features across secondary explanation domains

Feature names from Tiwari et al. [6]; domain grouping is an original analytical recoding.

Table 4. Consensus features selected by LASSO, elastic net and random forest

Secondary explanation domain	Consensus features from supplementary feature table [6]	Count
Morphology and structure	bodyBuild_Size, shoulder_Breadth, bodyFrame_Breadth, chest_Breadth, face_Size	5
Skin, temperature and appearance	skin_Dry/Oily, bodytemp_Amount, healthproblem_in_temp, palms_Color, body_Odour, palms_cracked, skin_wrinkled	7
Motor and speech	hands_Movements, speaking_Speed, walking_Speed,	6

Secondary explanation domain	Consensus features from supplementary feature table [6]	Count
	eyebrows_Movements, walking_style, voice_clear	
Strength and work	working_Quality, mental_Power, physical_Power, resistance_Power	4
Psychobehavioural and cognitive	Irritability_speed, makingFriends_speed, Anger_Speed, forgetfulness_speed, memory_olfactory	5
Digestive, excretory and metabolic	bowel_Tendency, body weight_Changes, appetite_regularity, stool_Consistency	4

Framework-development output. The proposed architecture in Figure 1 was produced from the secondary findings and governance standards. Its central design decision is an explanation layer placed between prediction and action. The system accepts structured Ayurvedic observations, preserves data provenance, applies transparent and nonlinear models, and returns a ranked profile with feature contributions, contradictions, uncertainty and missing-information prompts. The physician then accepts, modifies or rejects the output, and the final rationale is stored for audit. Table 5 specifies the model stack and explanation products; Table 6 defines minimum validation and governance checks.

5. Discussion

This study demonstrates why explainability in Ayurvedic AI must be treated as a clinical and epistemic requirement rather than a cosmetic visualization. The available machine-learning evidence shows that structured phenotypic traits contain enough signal to reproduce clinician-assigned extreme Prakriti classes with high external accuracy [6]. However, reciprocal validation reveals that performance depends on the direction of transport. A model trained in one region did not behave identically when applied to another population, even though the same broad constitution categories were used. This is consistent with the original investigators' observation that culture, region, ethnicity, examiner skill and interpretation affect assessment [6].

The cross-regional asymmetry has practical implications. When random forest was trained in Vadu and tested in North India, Pitta sensitivity was lower than Kapha and Vata. In the reverse direction, Vata sensitivity was lower than the other classes. These differences may reflect sample size, phenotype prevalence, feature distributions, examiner practices or nonlinear model fit. Aggregate data cannot determine the cause. A deployable system should therefore show class-specific validation ranges and identify whether the current person resembles the reference cohort. A warning such as 'limited evidence for this age-region-language combination' is more responsible than presenting a precise probability without context.

The small internal holdout in the original study was perfectly classified, whereas external performance was lower [6]. This is a familiar prediction-model pattern: internal accuracy can be optimistic when development and validation share recruitment, measurement and labelling processes. TRIPOD+AI and PROBAST+AI emphasize external validation, transparent outcome definition, handling of missing data and avoidance of overfitting [26,27]. In Ayurveda, an additional source of bias is circularity. If a questionnaire's deterministic scoring rule creates the label and the same responses are used to train a model, high accuracy may merely reproduce the rule. Physician-adjudicated labels, independent observation and longitudinal stability are preferable.

6. Conclusion

Explainable artificial intelligence can support Ayurvedic diagnosis only when computational efficiency is subordinated to clinical validity, contextual interpretation and physician responsibility. Secondary analysis of published Prakriti classification evidence showed strong but direction-dependent external performance. The best reconstructed reciprocal result was obtained with random forest, yet class sensitivity varied by region and validation direction. A separate open dataset offered useful bilingual structure but was dominated by young adults and Pitta-related labels. These findings show why accuracy alone is insufficient.

The proposed architecture makes explanation an operational layer between prediction and action. It preserves stable and current observations separately, exposes feature contributions and contradictions, quantifies uncertainty, detects population mismatch, requests additional examination and requires a qualified Ayurveda physician to accept, modify or reject the output. It also provides auditability and a pathway for prospective validation under ICMR, WHO and contemporary AI reporting standards.

No autonomous diagnostic or treatment claim is supported by the available data. The next step is a multicentre, clinician-labelled, longitudinal study in which transparent baselines and calibrated nonlinear models are compared, explanations are tested for stability and clinical usefulness, and mixed constitutions are represented explicitly. Under these conditions, XAI may improve consistency, documentation and research reproducibility while preserving the holistic and human-centred character of Ayurvedic practice.

References

1. Sharma PV, editor and translator. Caraka Samhita. 4 vols. Varanasi: Chaukhambha Orientalia; 2005.

2. Murthy KRS, translator. Susruta Samhita. 3 vols. Varanasi: Chaukhambha Orientalia; 2012.
3. Murthy KRS, translator. Vagbhata's Astanga Hrdayam. 3 vols. Varanasi: Chowkhamba Krishnadas Academy; 2018.
4. Rastogi S. Development and validation of a Prototype Prakriti Analysis Tool (PPAT): inferences from a pilot study. Ayu. 2012;33(2):209-18. doi:10.4103/0974-8520.105240.
5. Singh R, Sharma L, Ota S, Gupta B, Singhal R, Rana R, et al. Development of a standardized assessment scale for assessing Prakriti (psychosomatic constitution). Ayu. 2022;43(4):109-29. doi:10.4103/ayu.ayu_239_22.
6. Tiwari P, Kutum R, Sethi T, Shrivastava A, Girase B, Aggarwal S, et al. Recapitulation of Ayurveda constitution types by machine learning of phenotypic traits. PLoS One. 2017;12(10):e0185380. doi:10.1371/journal.pone.0185380.
7. Prasher B, Negi S, Aggarwal S, Mandal AK, Sethi TP, Deshmukh SR, et al. Whole genome expression and biochemical correlates of extreme constitutional types defined in Ayurveda. J Transl Med. 2008;6:48. doi:10.1186/1479-5876-6-48.
8. Aggarwal S, Negi S, Jha P, Singh PK, Stobdan T, Pasha MA, et al. EGLN1 involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. Proc Natl Acad Sci U S A. 2010;107(44):18961-6. doi:10.1073/pnas.1006108107.
9. Chauhan NS, Pandey R, Mondal AK, Gupta S, Verma MK, Jain S, et al. Western Indian rural gut microbial diversity in extreme Prakriti endo-phenotypes reveals signature microbes. Front Microbiol. 2018;9:118. doi:10.3389/fmicb.2018.00118.
10. Mukerji M. Ayurgenomics-based frameworks in precision and integrative medicine: translational opportunities. Camb Prism Precis Med. 2023;1:e29. doi:10.1017/pcm.2023.15.
11. Joshi A, Kulkarni A, Chandran S, Jayaraman VK, Kulkarni BD. Nadi Tarangini: a pulse based diagnostic system. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:2207-10. doi:10.1109/IEMBS.2007.4352781.
12. Fatangare M, Bhingarkar S. A comprehensive review on technological advancements for sensor-based Nadi Pariksha: an ancient Indian science for human health diagnosis. J Ayurveda Integr Med. 2024;15(3):100958. doi:10.1016/j.jaim.2024.100958.
13. Singh AK, Singh J. Prakriti200: a questionnaire-based dataset of 200 Ayurvedic Prakriti assessments. arXiv [Preprint]. 2025. arXiv:2510.06262. doi:10.48550/arXiv.2510.06262.
14. Gupta A, Singh V, Chandra S, Garg R. Towards standardization of Prakriti evaluation: a scoping review of modern assessment tools and their psychometric properties in Ayurvedic medicine. J Ayurveda Integr Med. 2025;16(4):101157. doi:10.1016/j.jaim.2025.101157.
15. Venkatesh A, et al. Prakriti (constitutional typology) in Ayurveda: a critical review of assessment tools and their scientific rigor. Front Med (Lausanne). 2025;12:1656249. doi:10.3389/fmed.2025.1656249.
16. World Health Organization, International Telecommunication Union, World Intellectual Property Organization. Mapping the application of artificial intelligence in traditional medicine. Geneva: World Health Organization; 2025.
17. Pujari S, Singh R, Soon GC, Nesari T, Ghelman R, Zhao Y, et al. Artificial intelligence in traditional medicine: policy and governance strategies. Bull World Health Organ. 2025. doi:10.2471/BLT.24.292888.
18. Ng JY, Cramer H, Lee MS, Moher D. Traditional, complementary, and integrative medicine and artificial intelligence: novel opportunities in healthcare. Integr Med Res. 2024;13(1):101024. doi:10.1016/j.imr.2024.101024.
19. Hou C, Gao Y, Lin X, Wu J, Li N, Lv H, et al. A review of recent artificial intelligence for traditional medicine. J Tradit Complement Med. 2025;15(3):215-28. doi:10.1016/j.jtcme.2025.02.009.
20. Indian Council of Medical Research. Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. New Delhi: ICMR; 2023.
21. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
22. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016. p. 1135-44. doi:10.1145/2939672.2939778.
23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30. 2017. p. 4765-74.
24. Breiman L. Random forests. Mach Learn. 2001;45:5-32. doi:10.1023/A:1010933404324.
25. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1-22. doi:10.18637/jss.v033.i01.
26. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
27. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias,

and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. doi:10.1136/bmj-2024-082505.

28. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9.
29. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-74. doi:10.1038/s41591-020-1034-x.
30. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26(9):1351-63. doi:10.1038/s41591-020-1037-7.