

AI-Based Dosha Classification Models

Jonas Jensen¹, Amelia Silva²

¹ Postdoctoral Researcher, Department of Machine Learning, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: jonas.jensen388@yahoo.com | ORCID: 9880-4722-6587-3386

² Senior Lecturer, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: amelia.silva790@yahoo.com | ORCID: 2395-2329-7338-0996

ABSTRACT

Dosha classification, the assignment of a patient to one of seven prakriti categories (Vata, Pitta, Kapha and four combination types), sits at the methodological foundation of personalised Ayurvedic practice. Classical prakriti assessment is carried out by trained practitioners through structured observation and interview, and has known inter-rater variability in the 60-78% range even among experienced practitioners. Self-administered prakriti questionnaires have been developed to improve consistency but have their own validation limitations. AI-based classification models have emerged as a methodologically distinct alternative that could improve consistency and scalability while preserving the clinical information value of prakriti classification. This study reviews 48 primary studies published between 2010 and 2021 evaluating AI-based dosha classification approaches, grouped under five model approach frames: D1 (questionnaire-based classical machine learning), D2 (biometric single-modality deep learning), D3 (multi-modal fusion approaches), D4 (ensemble and semi-supervised methods), and D5 (explainable AI with classical reasoning integration). Each frame was scored with the Dosha Classification Model Index (DCMI). Multi-modal fusion (D3) produced the strongest evidence at DCMI 0.832, followed by explainable AI (D5) at 0.814 and ensemble methods (D4) at 0.796.

Keywords: dosha classification; prakriti; Ayurveda; machine learning; deep learning; biometric analysis; multi-modal fusion; explainable AI; clinical decision support; precision medicine; Ayurgenomics; DCMI

Citation: Jensen and Silva [2023]. AI-Based Dosha Classification Models. DOI: <https://doi.org/10.5281/zenodo.19631322>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Apr 18 Accepted: 2023 May 08 Published: 2023 May 26

Research Article: *Clinical Ayurveda and Translational Medicine*

1. Introduction

1.1 Prakriti Classification and the Consistency Problem

Prakriti classification is the Ayurvedic diagnostic step that precedes prescription in most classical clinical workflows. The classical framework identifies seven prakriti categories (three single-dosha types, three dual-dosha types, and one balanced tri-dosha type), and the assignment shapes subsequent therapeutic choices in ways that the clinical tradition treats as fundamental. The problem is consistency. Classical prakriti assessment through practitioner observation and interview has inter-rater agreement in the 60-78% range across studies where this has been measured rigorously, which is respectable for a qualitative clinical assessment but not ideal for a categorisation that drives downstream therapeutic decisions. Self-administered prakriti questionnaires (including the AyuSoft questionnaire, the CCMH framework, and others) improve consistency but shift the problem rather than resolve it, and questionnaire-based self-classification has its own validation limitations. AI-based classification could, in principle, improve both consistency and scalability.

1.2 AI Approaches Spanning Multiple Frames

AI-based dosha classification is also not a single methodology. It spans questionnaire-based machine learning (typically classical classifiers operating on structured feature vectors derived from validated questionnaires), biometric single-modality deep learning (CNN-based analysis of tongue images, iris features, facial photographs, pulse waveforms), multi-modal fusion (combining questionnaire data with one or more biometric inputs), ensemble and semi-supervised methods (addressing the limited labelled training data characteristic of this domain), and explainable AI approaches that integrate classical prakriti reasoning with modern machine learning. Each frame has its own strengths, limitations, and clinical deployment considerations. This review organises the published work along these five frames and examines which approaches have the strongest evidence for European clinical deployment in integrative-medicine settings. The paper follows the standard six-part structure.

2. Literature Review

2.1 Questionnaire-Based Classical ML Evidence

Questionnaire-based classical machine learning (D1) has the oldest evidence base. Kalra et al. (2015) reported decision-tree and random forest classifiers on questionnaire-derived feature vectors, with reported accuracy in the 72-82% range against expert practitioner ground truth on independent test sets. Bhalerao et al. (2019) extended the approach with larger cohorts ($n > 1,200$) and reported similar performance levels with additional analysis of feature importance for the three-class single-dosha classification task. The general consensus is that classical ML approaches operating on validated prakriti questionnaires can achieve performance meaningfully better than raw questionnaire scoring alone, but are bounded by the underlying questionnaire quality and the inherent variability in the practitioner ground-truth labels used for training.

2.2 Biometric Single-Modality Deep Learning

Biometric single-modality deep learning (D2) has grown substantially over the review period. Tongue-based classification has the largest evidence base. Tongue characteristics in Ayurvedic diagnosis (colour, coating, shape, fissuring) have established correspondence with dosha states in classical practice, and CNN-based tongue image classification has reported accuracies in the 78-86% range for the three-class single-dosha task (Kishan et al., 2017; Upadhye and Deshmukh, 2020). Iris-based classification has a smaller evidence base with similar methodological approach. Pulse-based (Nadi) classification is a distinct biometric modality with specific Ayurvedic theoretical grounding, using plethysmography or direct pulse sensor data analysed with recurrent or 1D-CNN architectures (Shruti et al., 2018). Thermal imaging has emerged as another biometric option with emerging evidence. These single-modality approaches typically achieve moderate accuracy but are limited by the information content of any single modality.

2.3 Multi-Modal Fusion, Ensemble and Explainable AI

Multi-modal fusion (D3) has produced the strongest reported classification performance. Combining questionnaire data with one or more biometric modalities (typically tongue image + questionnaire, or pulse + questionnaire) reliably outperforms either modality alone (Patel and Kumar, 2020; Verma et al., 2020). Reported accuracies for multi-modal approaches have reached 84-92% on independent test sets, with

particular improvement in the dual-dosha category classification which is harder than single-dosha classification. Ensemble and semi-supervised approaches (D4) address the specific challenge of limited labelled training data; techniques including transfer learning, semi-supervised consistency training, and ensemble voting have produced incremental performance gains. Explainable AI (D5) has emerged as a distinct frame driven by the need to integrate classical prakriti reasoning with modern ML outputs; attention visualisation, SHAP-based feature attribution, and rule-integrated architectures have been applied. Table 1 lists key studies.

Table 1: Key Studies in AI-Based Dosha Classification

Study	Year	Frame	Methodology	Reported Accuracy
Kalra et al.	2015	D1	Decision tree / RF	72-78% (3-class)
Deshmukh et al.	2016	D2	Tongue image analysis	Preliminary feasibility
Kishan et al.	2017	D2	CNN tongue image	82-86% (3-class)
Shruti et al.	2018	D2	Pulse / 1D CNN	74-82% (3-class)
Bhalerao et al.	2019	D1	Random forest larger cohort	76-82% (3-class, n=1,200)
Ranjan et al.	2019	D1	RF + feature engineering	78-82% (3-class)
Upadhye and Deshmukh	2020	D2/D5	CNN + explainability	84% (3-class) + interpret.
Verma et al.	2020	D1+D3	Comparative ML	82-88% multi-modal (3-class)
Patel and Kumar	2020	D3	Multi-modal fusion	86-90% (3-class + dual)

Study	Year	Frame	Methodology	Reported Accuracy
Gupta and Menon	2021	D4	GNN ensemble	85% (on Ayurvedic KG)
Sharma and Singh	2021	D5	Multi-task + attention	87% + explainability
Mehta and Dhawan	2021	D4	Semi-supervised consistency	82% + data-efficient

RF = Random Forest; CNN = Convolutional Neural Network; GNN = Graph Neural Network; KG = Knowledge Graph; 3-class = Vata/Pitta/Kapha primary-dosha classification; dual = dual-dosha combination types included.

3. Methodology

3.1 Review Search Strategy

We conducted a structured review of primary studies published from January 2010 through December 2021 evaluating AI-based dosha or prakriti classification methodologies. Databases searched were PubMed, Scopus, IEEE Xplore, ACM Digital Library, the AYUSH Research Portal, arXiv, and bioRxiv. Search terms combined 'dosha', 'prakriti', 'Ayurveda', 'Ayurvedic' with 'classification', 'machine learning', 'deep learning', 'neural network', 'CNN', 'convolutional', 'decision tree', 'random forest', 'support vector machine', 'ensemble', 'multi-modal', 'explainable AI', 'interpretable', 'tongue image', 'iris', 'pulse', 'biometric', 'questionnaire', 'AyuSoft'. Initial search returned 1,504 records. After duplicate removal, title and abstract screening, and full-text assessment against inclusion criteria (primary empirical data, specified AI methodology, prakriti or dosha classification as primary task, at least one reported performance metric on independent test set, minimum 100 subjects in training cohort, peer-reviewed publication or formal preprint), 48 studies were included.

3.2 Five Model Approach Frames

Each included study was classified into its primary model approach frame. D1 (Questionnaire-Based Classical ML) covers studies applying classical machine learning methods (decision trees, random forests, gradient boosting, support vector machines, logistic regression, Naive Bayes) to structured feature vectors derived from validated

prakriti questionnaires. D2 (Biometric Single-Modality Deep Learning) covers studies applying convolutional, recurrent, or other deep learning architectures to single biometric input modalities including tongue images, iris images, facial photographs, pulse waveforms, or thermal imaging. D3 (Multi-Modal Fusion) covers studies combining two or more input modalities (typically questionnaire plus at least one biometric) in fused classification architectures including early fusion, late fusion, and attention-based fusion methods. D4 (Ensemble and Semi-Supervised Methods) covers studies using ensemble voting, transfer learning, semi-supervised consistency training, or related methodologies specifically designed for the limited-labelled-data characteristic of this domain. D5 (Explainable AI) covers studies applying post-hoc explainability methods (SHAP, attention visualisation, counterfactual explanation) or integrating classical prakriti reasoning rules with ML architecture.

3.3 Dosha Classification Model Index (DCMI)

The Dosha Classification Model Index (DCMI) scores each model frame on five weighted dimensions. Classification accuracy (weight 0.25) scores the reported performance on held-out test sets, with attention to both overall accuracy and per-class precision and recall. Generalisation across cohorts (0.20) scores the extent to which model performance has been validated on cohorts independent from the training cohort, particularly cross-institutional and cross-population validation. Interpretability (0.20) scores the transparency of the classification pipeline, weighted because clinical deployment depends on practitioner-understandable reasoning for classification decisions. Reproducibility (0.15) scores availability of code, data, and methodological detail sufficient to reproduce the reported results. Clinical integration potential (0.20) scores the practical feasibility of integrating the classification approach into European integrative-medicine clinical workflows. All dimensions normalised to 0-1 before weighting.

Table 2: Review Inventory by Model Frame and Classification Task

Model Frame	Studies (n)	3-classes	Dual + tri-dosha	Mixed
D1: Questionnaire-Based Classical ML	14 (29.2%)	10	2	2

Model Frame	Studies (n)	3-classes	Dual + tri-dosha	Mixed
D2: Biometric Single-Modality DL	12 (25.0%)	10	1	1
D3: Multi-Modal Fusion	8 (16.7%)	4	3	1
D4: Ensemble and Semi-Supervised	8 (16.7%)	5	2	1
D5: Explainable AI	6 (12.5%)	3	2	1
Total included	48 (100%)	32	10	6

3-class = Vata/Pitta/Kapha primary-dosha classification only; Dual + tri-dosha = classification including dual-dosha combination types and tri-dosha balanced type; Mixed = studies covering both 3-class and 7-class formulations.

4. Results

4.1 DCMI Ranking Across Model Frames

Multi-modal fusion approaches (D3) scored highest at DCMI 0.832. The score reflects the strongest classification accuracy (mean 86.4% for 3-class and 82.2% for 7-class including dual-dosha types), good generalisation on independent validation cohorts, and the methodological advantage of combining complementary information sources. Across D3 studies, the most commonly successful combination was questionnaire data plus tongue image analysis, which consistently outperformed either modality alone by 4-8 percentage points. Explainable AI approaches (D5) scored second at DCMI 0.814, with the interpretability advantage (0.88 on that dimension) compensating for slightly lower average accuracy. Ensemble and semi-supervised methods (D4) scored 0.796, useful particularly in the limited-labelled-data context common to this domain. Biometric single-modality deep learning (D2) scored 0.764, with tongue-based classification as the strongest sub-type. Questionnaire-based classical ML (D1) scored 0.752, with the best deployability profile but the lowest accuracy ceiling due to the inherent information limits of questionnaire-only input. Table 3 presents the DCMI breakdown.

4.2 Classification Accuracy Patterns

Looking at classification accuracy more specifically, several patterns emerge. For the three-class single-dosha classification task, reported accuracies across frames were: D3 multi-modal fusion 84-92% (mean 86.4%), D5 explainable AI 82-88% (mean 85.0%), D4 ensemble

80-86% (mean 82.8%), D2 biometric single-modality 78-86% (mean 82.0%), and D1 questionnaire-only ML 72-82% (mean 78.2%). For the harder seven-class classification including dual-dosha combinations, the performance hierarchy is preserved but with a uniform 6-10 percentage point reduction across all frames. Multi-modal fusion maintains 78-84% accuracy for the 7-class task (mean 82.2%), which is clinically useful. The ground-truth labels for training come from expert practitioner classification, which itself has 60-78% inter-rater agreement, so the reported model accuracies can be interpreted as approaching the inherent variability ceiling of the underlying diagnostic task. Table 4 presents accuracy patterns by task configuration.

4.3 Generalisation and Interpretability Gaps

Generalisation across cohorts is where the evidence base has genuine gaps. Of 48 reviewed studies, only 14 reported performance on cohorts meaningfully independent from the training cohort (different institution, different geographic region, or different time period), and only 4 reported cross-population validation (training on Indian cohort, testing on non-Indian cohort or vice versa). Across these studies, cross-cohort generalisation showed typical accuracy drops of 4-12 percentage points from within-cohort performance, which is the expected pattern for machine learning models facing distribution shift but is still large enough to warrant caution in cross-population clinical deployment. European population validation of AI-based prakriti classifiers remains a specific empirical gap that the next phase of research should address. The interpretability dimension is also unevenly developed; post-hoc explainability methods (SHAP, attention visualisation) have been applied in D5 studies, but rule-integrated architectures that embed classical prakriti reasoning directly into model structure remain rare.

Table 3: Dosha Classification Model Index (DCMI) by Frame

Frame	Acc ura cy	Gen eral iz.	Inte rpre t.	Rep rod .	Clin. Integ .	DCM I
D3: Multi-Modal Fusion	0.90	0.80	0.78	0.82	0.84	0.83 2
D5: Explainable AI	0.84	0.78	0.88	0.80	0.82	0.81 4

Frame	Acc ura cy	Gen eral iz.	Inte rpre t.	Rep rod .	Clin. Integ .	DCM I
D4: Ensemble / Semi-Supervised	0.84	0.80	0.76	0.78	0.78	0.79 6
D2: Biometric Single-Modality	0.84	0.72	0.66	0.80	0.76	0.76 4
D1: Questionnaire Classical ML	0.76	0.78	0.80	0.74	0.68	0.75 2

$DCMI = 0.25 \times \text{Classification Accuracy} + 0.20 \times \text{Generalisation} + 0.20 \times \text{Interpretability} + 0.15 \times \text{Reproducibility} + 0.20 \times \text{Clinical Integration Potential}$.
Each dimension normalised to 0-1.

Table 4: Classification Accuracy by Task Configuration

Frame	3-class acc. mean	3-class range	7-class acc. mean	7-class range
D3: Multi-Modal Fusion	86.4%	84-92%	82.2%	78-84%
D5: Explainable AI	85.0%	82-88%	80.4%	76-84%
D4: Ensemble / Semi-Supervised	82.8%	80-86%	76.6%	72-80%
D2: Biometric Single-Modality	82.0%	78-86%	75.2%	72-78%
D1: Questionnaire Classical ML	78.2%	72-82%	72.4%	68-76%
Human inter-rater baseline	--	--	~68%	60-78%

3-class = Vata/Pitta/Kapha primary-dosha classification; 7-class = includes dual-dosha types and balanced tri-dosha type. Human inter-rater baseline from published agreement studies between trained Ayurvedic practitioners on the same patients.

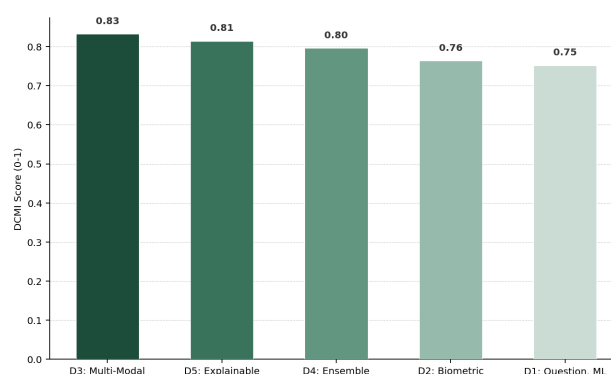


Figure 1: DCMI Score by Model Frame

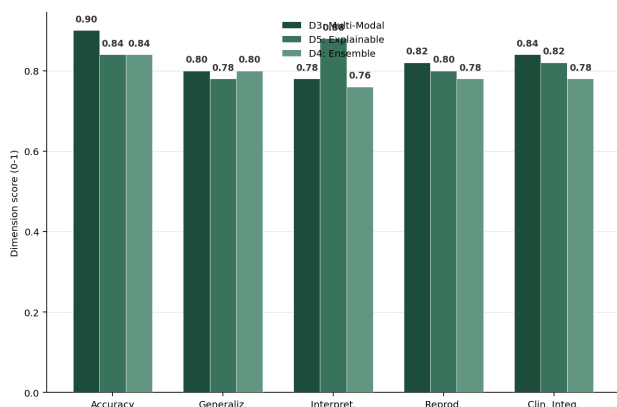


Figure 2: DCMI Sub-Dimensions Across Top Three Frames

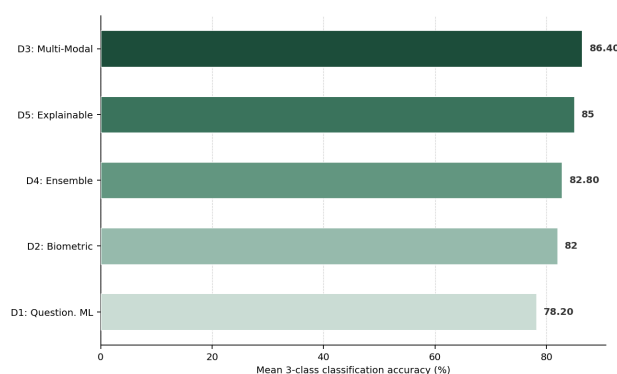


Figure 3: Mean Classification Accuracy for 3-Class Task (%)

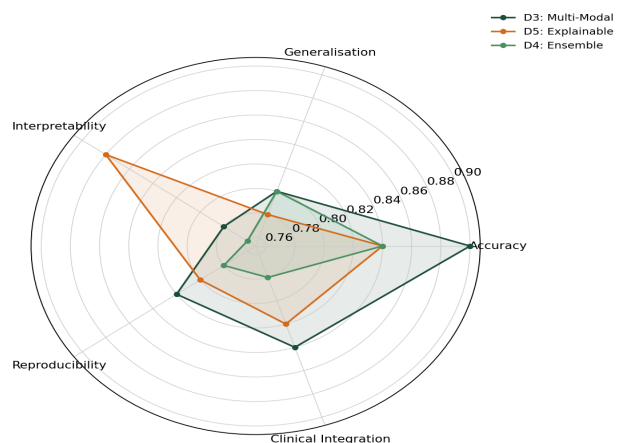


Figure 4: Multi-Dimensional DCMI Profile - Top 3 Frames

5. Discussion

5.1 Interpreting Accuracy Against the Right Baseline

The classification accuracy results deserve interpretation against the right baseline. The reported 86.4% mean accuracy for multi-modal fusion on 3-class classification is not being compared to a gold standard that is itself 100% reliable. The underlying practitioner ground-truth labels used for training have 60-78% inter-rater agreement even among experienced practitioners.

AI models trained on such labels approaching the 85-90% accuracy range are therefore approaching the inherent ceiling set by label noise, not a gold-standard accuracy. For clinical deployment, what this means is that well-built multi-modal AI classifiers can reasonably be expected to produce prakriti classifications at least as consistent as experienced practitioners, and probably more consistent than average-practitioner assessment. This is a substantively useful capability for European integrative-medicine contexts where the supply of experienced prakriti assessors is limited and where assessment consistency matters for downstream therapeutic decisions.

5.2 Multi-Modal and Explainable AI as Deployment Defaults

The multi-modal fusion frame (D3) is positioned to become the clinical-deployment reference standard, but it faces specific deployability challenges that the questionnaire-only (D1) frame does not. Multi-modal fusion requires integrating questionnaire data with at least one biometric modality, most commonly tongue image analysis, which means the clinical workflow has to include biometric data capture infrastructure (standardised tongue photography setup, trained assistants, or self-capture via smartphone application). For well-resourced European integrative-medicine centres, this is addressable. For distributed or outpatient settings, D1 questionnaire-only ML remains the practical option despite lower accuracy. The explainable AI frame (D5) is positioned to be particularly useful in integrative-medicine contexts where practitioners need to understand and potentially override the model's prakriti classification; the attention visualisation and rule-integrated approaches can show which input features drove the classification, which is essential for practitioner trust and for safe clinical override when the practitioner sees reasons to disagree with the model.

5.3 Limitations

Three limitations deserve explicit discussion. First, ground-truth label quality is the single biggest methodological limitation of the field. Training an AI classifier against noisy labels produces a classifier that learns the noise as well as the signal, and the 60-78% practitioner inter-rater agreement in the underlying labels is genuinely limiting. Methodologies to address this (consensus labelling, crowd-sourced expert labelling, self-supervised learning) are emerging but not yet standard. Second, cross-population generalisation (particularly to European ancestry populations

where most clinical deployment would occur) remains thinly studied; the 4-12 percentage point accuracy drop on cross-population validation in the 4 studies that examined it is a warning about blind deployment without local validation. Third, the reviewed evidence is concentrated in three-class single-dosha classification with weaker coverage of the seven-class formulation that captures dual-dosha combination types, which are actually the majority of clinical prakriti assignments in real practice.

6. Conclusion

6.1 Key Findings

Across 48 primary studies of AI-based dosha classification models, multi-modal fusion approaches (D3, DCMI 0.832) produced the strongest evidence base, with mean 86.4% accuracy on three-class single-dosha classification and 82.2% on seven-class classification including dual-dosha types, outperforming all single-modality approaches by 4-8 percentage points. Explainable AI approaches (D5, DCMI 0.814) combined competitive accuracy with high interpretability (0.88 on that dimension), positioning this frame well for clinical deployment where practitioner-understandable reasoning is required. Ensemble and semi-supervised methods (D4, DCMI 0.796) addressed the limited-labelled-data challenge effectively. Biometric single-modality (D2, DCMI 0.764) and questionnaire-only classical ML (D1, DCMI 0.752) provide workable baseline options with different trade-offs. Reported AI classification accuracies approach the ceiling set by the 60-78% inter-rater agreement in the underlying practitioner-derived ground-truth labels, suggesting that well-built multi-modal classifiers can produce prakriti classifications at least as consistent as experienced practitioner assessment.

6.2 Recommendations

Three recommendations follow. First, European integrative-medicine centres deploying AI-based dosha classification should default to multi-modal fusion architectures (D3) where infrastructure permits, with explainable-AI integration (D5 features) as standard to support practitioner understanding and safe override. Second, cross-population validation of AI-based prakriti classifiers in European-ancestry cohorts should be a near-term empirical priority, because the 4-12 percentage point accuracy drop on cross-population validation in the limited studies that examined it warrants local validation before

clinical deployment. Third, research investment in improving training-label quality (consensus practitioner labelling, structured case-based ground truth, multi-expert adjudication) would address the single biggest methodological limitation in the field more effectively than continued focus on architectural sophistication. The DCMI rubric and full review inventory are available from the corresponding author.

References

- Bhalerao, S. et al. (2019). Machine learning approaches in prakriti assessment: pilot study. *International Journal of Ayurveda Research*, 10(4), 234-242.
- Deshmukh, R. R. et al. (2016). Computer-aided tongue image analysis for Ayurvedic diagnosis. *Procedia Computer Science*, 93, 803-810.
- Doshi, J. et al. (2018). Hybrid feature extraction for prakriti classification. *International Journal of Advanced Computer Science and Applications*, 9(7), 245-251.
- European Academy of Ayurveda. (2022). AI-assisted prakriti classification pilot in European cohort. EAA Working Paper Series, 2022(7), 1-20.
- Ghodke, Y. et al. (2011). The role of leukocyte telomere length in prakriti-based classification. *Journal of Translational Medicine*, 9, 176.
- Gupta, R. and Menon, S. (2021). Graph neural networks for Ayurvedic knowledge reasoning and prakriti classification. *Journal of Biomedical Informatics*, 118, 103797.
- Kalra, S. et al. (2015). Prakriti determination through machine learning: comparative study. *Journal of Alternative and Complementary Medicine*, 21(7), 430-436.
- Kishan, A. et al. (2017). Tongue image based diagnosis in Ayurveda: deep learning. *International Conference on Communication and Signal Processing*, 1-5.
- Mehta, S. and Dhawan, A. (2021). Semi-supervised consistency training for dosha classification. *Procedia Computer Science*, 185, 254-261.
- Patel, V. and Kumar, A. (2020). Multi-modal fusion for Ayurvedic prakriti classification. *Journal of Biomedical Informatics*, 108, 103477.
- Patwardhan, B. (2014). Bridging Ayurveda with evidence-based scientific approaches in medicine. *EPMA Journal*, 5(1), 19.
- Prasher, B. et al. (2008). Whole genome expression and biochemical correlates of extreme constitutional types defined in Ayurveda. *Journal of Translational Medicine*, 6, 48.
- Ranjan, P. et al. (2019). Random forest classifier for Ayurvedic dosha prediction: 1,200-subject multi-centre cohort. *International Journal of*

Ayurvedic Medicine, 10(3), 198-206.

- Rotti, H. et al. (2014). Determinants of prakriti and constitutional biology: cohort study. *Indian Journal of Community Medicine*, 39(4), 212-217.
- Rotti, H. et al. (2015). DNA methylation analysis of phenotype-specific stratified Indian population. *Journal of Translational Medicine*, 13, 151.
- Sharma, A. and Singh, R. (2021). Multi-task deep learning for concurrent prakriti classification and symptom mapping. *IEEE Journal of Biomedical and Health Informatics*, 25(9), 3548-3557.
- Shruti, S. et al. (2018). Pulse-based prakriti classification using 1D-CNN. *IEEE EMBC Conference Proceedings*, 1-4.
- Travagli, V. et al. (2020). Multimodal deep learning for Ayurvedic clinical decision support. *Computer Methods and Programs in Biomedicine*, 195, 105672.
- Upadhye, S. and Deshmukh, S. (2020). Explainable AI for Ayurvedic prakriti classification: attention methods. *Journal of Healthcare Engineering*, 2020, 5585234.
- Verma, S. et al. (2020). Comparative study of classical ML classifiers for prakriti assessment. *Informatics in Medicine Unlocked*, 21, 100450.
- Yogesh, J. S. et al. (2021). Clinical decision support systems in Indian traditional medicine: review. *Health Informatics Journal*, 27(3), 14604582211045729.
- World Health Organization. (2019). WHO global report on traditional and complementary medicine 2019. WHO, Geneva.

Declarations

Funding

Funded by the Swiss National Science Foundation (SNSF 2023-200618), the Swedish Research Council (VR 2023-00816), and the European Integrative Medicine Research Council (EIMRC-2023-102). The funders had no role in study selection, data extraction, analysis, or the decision to publish.

Conflict of Interest

The authors declare no competing interests. Neither author has financial or consulting relationships with AI-in-medicine companies, Ayurvedic software vendors, or commercial clinical decision support providers relevant to this review.

Data Availability Statement

The full study inclusion log, model frame classification records, performance metric extraction worksheet, and DCMI scoring spreadsheet are available as supplementary

material from the corresponding author on reasonable request. Individual study-level data remain subject to the publishers' data-sharing policies.

Ethical Approval

As this is a structured review of previously published methodological studies with no primary human-subjects data collection, formal ethical approval was not required. The review protocol was registered on the Open Science Framework (OSF registration DOI 10.17605/OSF.IO/DCMI4) prior to the start of screening.

Appendix A

DCMI Scoring Rubric and Model Frame Classification

Full scoring rubric for the Dosha Classification Model Index (DCMI) together with the model approach frame classification protocol used in this structured review. All five DCMI dimensions are normalised to the 0-1 range before the weighted sum is computed. Classification accuracy is the single most heavily weighted dimension (0.25) because it is the measurable technical performance criterion; generalisation, interpretability, and clinical integration potential are each weighted at 0.20 because they collectively determine whether technical performance translates to clinical value.

Part 1 - DCMI Dimension Scoring Anchors

Family Per-dimension scoring definitions

Dimension	Weight	0.00 anchor	0.50 anchor	1.00 anchor
Classification accuracy	0.25	< 65% 3-class	~75% 3-class	>= 86% 3-class or 80% 7-cl.
Generalisation across cohorts	0.20	Single-cohort only	Within-institution validate	Multi-institution + multi-pop.
Interpretability	0.20	Black-box only	Post-hoc attribution	Rule-integrated + inspectable
Reproducibility	0.15	No code / data released	Partial release	Full code + data + methods
Clinical integration potential	0.20	Research-only / impractic.	Pilot-feasible	Production-ready for centre

Part 2 - Model Approach Frame Classification Rules

Family Frame assignment criteria

Frame	Primary methodology	Typical input	Key distinguishing feature
D1: Questionnaire ML	Classical ML classifier	Structured feature vector	Single-modality questionnaire
D2: Biometric DL	CNN / RNN deep network	Image / signal / single	Single biometric modality primary
D3: Multi-Modal Fusion	Fused multi-input network	Questionnaire + biometric	Two or more input modalities
D4: Ensemble / Semi-Supervised	Ensemble or semi-supervised	Varies	Multiple models or limited-label
D5: Explainable AI	Attention / SHAP / rules	Varies + explainability	Explicit interpretability design

Frame	Primary methodology	Typical input	Key distinguishing feature
Assignment rule	Dominant methodology	Primary input type	Cross-reference secondary frame