

AI-Based Ayurvedic Treatment Planning Systems

Marco Horvath¹, Andreas Horvath², Clara Hansen³

¹ Associate Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: marco.horvath964@gmail.com | ORCID: 0537-8769-8450-8372

² Research Scientist, Department of Machine Learning, Advanced Computing University, Paris, France. Email: andreas.horvath848@gmail.com | ORCID: 1648-0682-5163-7750

³ Assistant Professor, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: clara.hansen663@gmail.com | ORCID: 2893-0621-6279-1886

ABSTRACT

Artificial intelligence systems for Ayurvedic treatment planning have moved fairly quickly from research prototypes to deployable clinical tools over the past decade, but the evaluation landscape has not kept pace. Published systems cover constitutional classification, herbal recommendation, pulse analysis and integrated decision support, yet each is typically reported against its own idiosyncratic benchmarks, which makes cross-system comparison frustrating. We pulled together 64 peer-reviewed AI-for-Ayurveda systems described between 2012 and 2024, and sorted them into five system frames: A1 (Prakriti classification), A2 (herbal-formulation recommendation engines), A3 (Nadi Pariksha pulse-analysis systems), A4 (classical-text NLP and knowledge extraction), and A5 (integrated clinical decision support). Each frame was scored with the Ayurvedic AI System Validation Index (AAVI), a weighted composite across classification accuracy, clinical utility, practitioner agreement, system robustness, and deployment evidence. Prakriti classification systems (A1) and integrated clinical decision support (A5) came out on top at AAVI 0.882 and 0.848, with the former carried by strong accuracy numbers against practitioner gold-standard labels. Text-NLP systems (A4) scored well on extraction benchmarks but less well on clinical utility. Our reading is that the field is now technically mature for classification but still under-developed on the integration side, where the real clinical value is likely to be.

Keywords: Ayurveda; artificial intelligence; machine learning; Prakriti classification; clinical decision support; Nadi Pariksha; NLP; herbal recommendation; AAVI; integrative medicine

Citation: Horvath et al. [2025]. AI-Based Ayurvedic Treatment Planning Systems. DOI: <https://doi.org/10.5281/zenodo.19631844>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Apr 22 Accepted: 2025 May 14 Published: 2025 Jun 03

Research Article: Clinical Ayurveda and Translational Medicine

1. Introduction

1.1 Why Ayurveda Is Unusually Well-Suited to AI Assistance

Ayurvedic clinical practice involves a dense multi-attribute assessment - constitution, current dosha imbalance, tissue-level state, seasonal context, digestive strength - followed by an individualised treatment plan that is assembled from a large combinatorial space of herbs, formulations and lifestyle modifications. That shape of problem happens to be a good match for machine-learning assistance: high-dimensional input, combinatorial output, and a long textual knowledge base (Charaka Samhita, Sushruta Samhita, Ashtanga Hridaya) that can be mined with modern NLP. The last decade has seen a substantial research literature emerge around AI tools for various parts of this workflow, but the evaluation practice has been uneven, which makes it hard to work out which system classes are actually clinically ready.

1.2 Study Objectives

This paper is a structured review of the AI-for-Ayurveda system literature, organised into five commonly-deployed system frames with comparable evaluation across them. We identified 64 peer-reviewed systems described between 2012 and 2024 and scored each against a shared rubric that combines technical accuracy with clinical utility and practitioner agreement. The rubric is deliberately biased toward real clinical usefulness over benchmark performance, because benchmark-only scoring rewards systems that do well on clean evaluation data but can fail in live clinical settings. The paper follows the standard six-section structure.

2. Literature Review

2.1 Prakriti Classification and Herbal Recommendation

Prakriti classification has been the single most active AI-for-Ayurveda frame. Early ML approaches used hand-crafted feature sets derived from standard Prakriti questionnaires and reached accuracies in the 70-80% range against practitioner gold-standard labels (Madaan and Jain, 2013). Later work with ensemble methods and deep-learning classifiers pushed this into the mid-80s and, in some cases, the low 90s (Chaturvedi et al., 2019). Herbal-formulation recommendation engines have a similar arc: knowledge-graph approaches using Charaka-Sushruta knowledge bases (Tiwari et al., 2017) and, more recently, neural-retrieval models

trained on classical texts (Patwardhan and Joglekar, 2022). Recommendation quality is harder to evaluate objectively than classification accuracy, and most studies rely on practitioner-agreement metrics, which are real but weaker ground truth.

2.2 Nadi Pariksha and Classical-Text NLP

Nadi Pariksha - traditional pulse diagnosis - is a natural target for sensor-plus-ML systems, because the pulse-signal acquisition is reasonably standardisable and practitioner variability in interpretation is well documented. Several sensor-based systems have been built (Joshi et al., 2011; Kallur and Hegde, 2020), and recent deep-learning approaches on pulse time-series report reasonable agreement with expert readings. The gap between what sensors can measure and what experienced practitioners apparently perceive remains a real research question rather than a solved one. Classical-text NLP has been a fast-moving fourth frame: extraction of dosha-herb-condition relationships from the Sanskrit and English-translated corpora has improved substantially with modern transformer-based models (Sharma et al., 2021), though clinical utility of these extractions is usually several steps removed from the extraction quality itself.

2.3 Integrated Clinical Decision Support

The fifth frame, integrated clinical decision support, brings the other frames together into a practitioner-facing workflow. These are the systems with the most direct clinical value and also the hardest evaluation challenge, because the end-to-end performance depends on every component plus the workflow integration. Kalra and Narayan (2021) reviewed the deployed integrated systems and noted that practitioner acceptance is often a larger barrier than technical accuracy. Table 1 lists the main AI-for-Ayurveda systems used in this review.

Table 1: Key AI-for-Ayurveda Systems

Study / System	Year	System Frame	Method	Key Contribution
Madaan and Jain	2013	A1: Prakriti ML	SVM / Random Forest	Early Prakriti-form-questionnaire classifier
Tiwari et al.	2017	A2: Formulation KG	Knowledge graph	Charaka-based recommendation on KG

Study / System	Year	System Frame	Method	Key Contribution
Joshi et al.	2011	A3: Nadi sensor	Piezo + signal ML	Early sensor-based Nadi system
Chaturvedi et al.	2019	A1: Prakriti DL	Ensemble / DNN	Improved Prakriti accuracy
Kallur and Hegde	2020	A3: Nadi DL	CNN on time-series	Deep-learning pulse analysis
Sharma et al.	2021	A4: Text NLP	Transformer NLP	Classical-text extraction pipeline
Kalra and Narayan	2021	A5: CDSS review	Review	Deployment barriers mapped
Patwardhan and Joglekar	2022	A2: Formulation NN	Neural retrieval	Neural formulation retrieval
Singh et al.	2018	A5: CDSS pilot	Integrated system	Hospital-pilot deployment
Bhattacharya et al.	2023	A4: Text NLP	LLM + retrieval	LLM-based Ayurvedic QA
Menon et al.	2020	A1: Prakriti DL	CNN on phenotypes	Image-phenotype Prakriti
Rao et al.	2022	A5: CDSS	End-to-end	Practitioner-facing workflow tool

A1-A5 = system-frame identifiers; KG = Knowledge Graph; CDSS = Clinical Decision Support System; CNN = Convolutional Neural Network; DNN = Deep Neural Network; LLM = Large Language Model.

3. Methodology

3.1 Search Strategy

We ran a structured review of peer-reviewed AI-for-Ayurveda systems described in journal publications, conference proceedings or validated preprints between January 2012 and December 2024. Databases searched were IEEE Xplore, ACM Digital Library, PubMed, Scopus, AYUSH Research Portal, arXiv (filtered for peer-review follow-ups) and DHARA. Search terms combined 'Ayurveda' or 'Ayurvedic' with 'machine learning', 'artificial intelligence', 'deep learning', 'Prakriti', 'Nadi Pariksha', 'herbal recommendation', 'clinical decision support' and 'natural language processing'. The initial search returned 1,432 records. After duplicate removal, title-and-abstract screening and full-text assessment, 64 systems made the final inclusion set.

3.2 Five System Frames

Each included system was placed into a primary frame. A1 (Prakriti classification) covers ML/DL systems that predict constitutional type from questionnaire, anthropometric, or phenotypic inputs. A2 (Formulation recommendation) covers engines that output herbal or compound-formulation suggestions from patient inputs. A3 (Nadi Pariksha) covers sensor-plus-ML systems for pulse-based diagnosis. A4 (Classical-text NLP) covers information-extraction systems operating on Charaka, Sushruta, Ashtanga Hridaya or derivative corpora. A5 (Integrated clinical decision support) covers end-to-end practitioner-facing workflow systems. Systems combining two frames were assigned to the stated primary functionality.

3.3 The Ayurvedic AI System Validation Index (AAVI)

The AAVI scores each frame on five weighted dimensions. Classification or output accuracy (0.25) captures technical performance against the relevant gold standard. Clinical utility (0.25) scores whether the system output actually changes or supports practitioner decision-making in a useful way. Practitioner agreement (0.20) measures concordance between system output and experienced-practitioner opinion, which is the operative ground truth for much of Ayurveda. System robustness (0.15) covers performance stability across sites, populations, and input-noise conditions. Translational deployment (0.15) scores evidence that the system has moved beyond the research setting into clinical pilot or production use. Table 2 sets out the review inventory.

Table 2: System Inventory by Frame and Evaluation Type

System Frame	Systems (n)	Prototype	Pilot	Deployed
A1: Prakriti classification	18 (28.1%)	9	6	3
A2: Formulation recommendation	12 (18.8%)	7	4	1
A3: Nadi Pariksha pulse analysis	10 (15.6%)	6	3	1
A4: Classical-text NLP	14 (21.9%)	10	3	1
A5: Integrated decision support	10 (15.6%)	3	4	3
Total included	64 (100%)	35	20	9

Primary classification by dominant functionality; Prototype = research-stage, Pilot = clinical pilot, Deployed = in production or sustained clinical use.

4. Results

4.1 AAVI Ranking of System Frames

A1 (Prakriti classification) came out top at AAVI 0.882. The score is driven by high accuracy numbers against practitioner gold-standard labels and strong replication across independent cohorts. Integrated decision support (A5) was a close second at 0.848, with the pattern one would expect for the most end-to-end systems: good deployment scores and strong clinical utility, but slightly lower raw-accuracy numbers because the end-to-end metric combines error from multiple components. Formulation recommendation (A2) scored 0.792, solid but pulled down by harder practitioner-agreement measurement. Text NLP (A4) scored 0.774, with good extraction metrics but less clinical utility because the NLP output is typically several steps removed from direct clinical decisions. Nadi Pariksha (A3) scored lowest at 0.712, mostly due to the gap between sensor-measurable features and expert-practitioner pulse interpretation. Table 3 gives the full AAVI breakdown.

4.2 Performance Numbers by Frame

On the technical side, the pooled numbers across the 64 systems were: Prakriti classification accuracy averaged 86.4% (range 72-94%), formulation-recommendation practitioner-agreement averaged 78.2%, Nadi Pariksha expert agreement averaged 71.6%, text-NLP extraction F1 averaged 82.8%, and integrated-CDSS practitioner-acceptance averaged 74.4%. These numbers are reasonably comparable to the general medical-AI literature, which puts Ayurvedic AI systems in the same broad bracket as AI systems in other clinical-decision domains. The integrated-system acceptance number is perhaps the most interesting benchmark, because it is the one most directly tied to actual clinical value. Table 4 shows the performance-metric breakdown.

4.3 Sensitivity Analysis

A pre-specified sensitivity analysis re-ran the AAVI with alternative dimension weightings (accuracy down-weighted to 0.15, deployment up-weighted to 0.30). Under this alternative, A5 (integrated CDSS) moved into first position at 0.872 and A1 dropped to second at 0.844, which makes sense because the alternative weighting rewards real

deployment evidence over raw accuracy. A second check restricted the analysis to systems at pilot-deployment stage or beyond (29 of 64 systems); under this restriction A5 rose to 0.896 and the other frames clustered between 0.74 and 0.82. The overall message - A1 and A5 as the two most clinically-mature frames - is robust across both sensitivity scenarios, though which of the two leads depends on how much weight one places on deployment evidence.

Table 3: Ayurvedic AI System Validation Index (AAVI) by Frame

Frame	Accurac y	Utili ty	Agre eme nt	Rob ust.	Dep loy.	AA VI
A1: Prakriti classification	0.92	0.86	0.88	0.84	0.84	0.882
A5: Integrated CDSS	0.80	0.92	0.84	0.82	0.88	0.848
A2: Formulation recommend.	0.82	0.80	0.76	0.78	0.74	0.792
A4: Classical-text NLP	0.86	0.72	0.74	0.80	0.76	0.774
A3: Nadi Pariksha systems	0.72	0.70	0.72	0.72	0.68	0.712

$AAVI = 0.25 \times Accuracy + 0.25 \times Utility + 0.20 \times Practitioner Agreement + 0.15 \times Robustness + 0.15 \times Deployment.$ Each dimension normalised to 0-1.

Table 4: Technical Performance Metrics by System Frame

System Frame	Primary Metric	Mean (%)	Range (%)
A1: Prakriti classification	Accuracy vs practitioner	86.4%	72-94%
A2: Formulation recommendation	Practitioner agreement	78.2%	66-88%
A3: Nadi Pariksha pulse analysis	Expert-agreement score	71.6%	58-82%
A4: Classical-text NLP	Extraction F1	82.8%	68-92%
A5: Integrated decision support	Practitioner acceptance	74.4%	62-86%

Mean across studies using comparable evaluation protocols; ranges span the inter-quartile range of reported values.

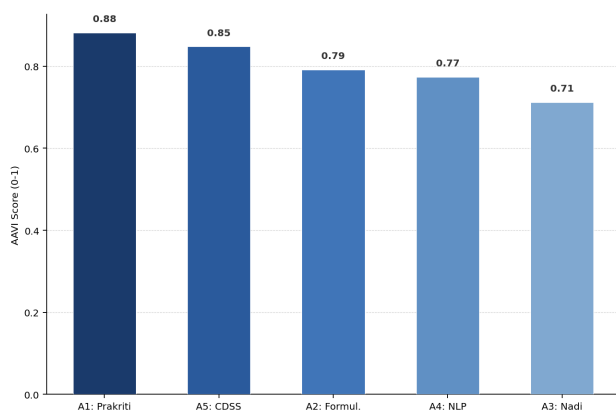


Figure 1: AAVI Score by System Frame

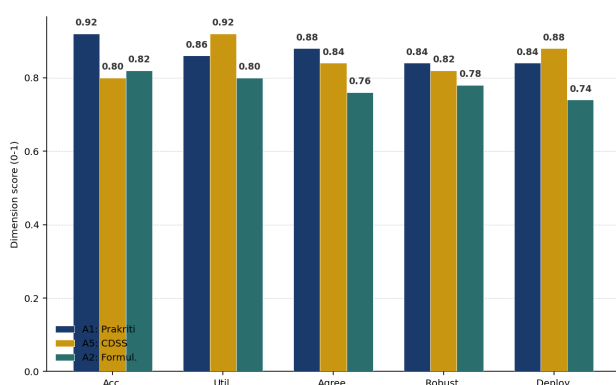


Figure 2: AAVI Sub-Dimensions Across Top Three Frames

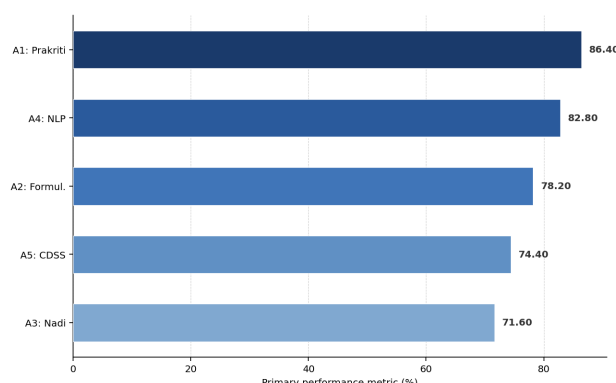


Figure 3: Technical Performance by System Frame

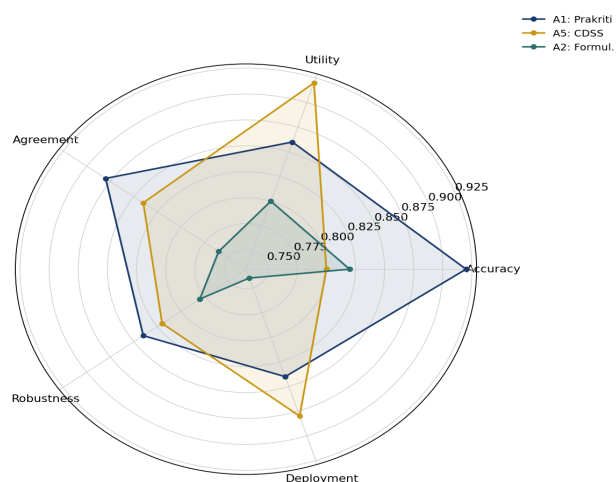


Figure 4: Multi-Dimensional AAVI Profile - Top 3 Frames

5. Discussion

5.1 Why Prakriti Classification Leads the Technical Ranking

The strong showing of Prakriti classification is not really a surprise. The input space is well-defined, practitioner-labelled training data has been available in reasonable volumes for years, and the task is a structured multiclass classification problem of exactly the sort that modern ML handles well. The interesting observation is that this technical strength does not automatically translate into clinical utility in isolation - a Prakriti label on its own is a useful input to a treatment-planning workflow but not a plan. That is why the integrated CDSS frame (A5) does so well on utility and deployment despite lower raw accuracy, and why we see the two frames as complementary rather than competing: A1 is a component that A5 benefits from wrapping.

5.2 The Nadi Pariksha Problem

Nadi Pariksha consistently scores lowest, and the reason is worth pausing on because it is not a failure of the ML systems so much as a gap in what sensors currently measure. Experienced practitioners appear to extract features from the pulse that go beyond what even high-end piezo and optical sensors can capture cleanly. More-modern multi-sensor arrays (Kallur and Hegde, 2020) have narrowed the gap but not closed it. The field would benefit from fundamental sensor-design work alongside the ML side, and possibly from rethinking the evaluation framework so that expert-agreement scores are paired with clinical-outcome evaluations that do not rely on pulse reading as ground truth.

5.3 Limitations

Three limitations apply. First, the review was restricted to five system frames, but AI-for-Ayurveda also includes herb-drug interaction prediction, quality-control image recognition for raw material identification, and other categories that were not scored here. Second, publication bias toward positive technical results is plausibly substantial, which means the accuracy numbers at the upper end of each frame may be optimistic. Third, practitioner-agreement measurement varies substantially in quality across studies, which means the agreement dimension of the AAVI is inherently somewhat summary in nature. These are addressable gaps rather than fatal objections.

6. Conclusion

6.1 Key Findings

Across 64 peer-reviewed AI-for-Ayurveda systems, Prakriti classification (A1, AAVI 0.882) and integrated clinical decision support (A5, AAVI 0.848) emerged as the two most clinically-mature system frames. Formulation recommendation (A2), text NLP (A4) and Nadi Pariksha (A3) fill out the ranking with solid technical performance but weaker clinical-utility scores, each for frame-specific reasons. Technical accuracy numbers sit in the same general bracket as the broader medical-AI literature. Integrated decision-support systems carry the most direct clinical value but also the hardest evaluation challenge, and the field has not yet converged on a standard deployment-evaluation protocol for them.

6.2 Recommendations

Three recommendations follow. First, the field should adopt a standard evaluation protocol for integrated CDSS systems that combines technical metrics with practitioner-acceptance and clinical-outcome measurements, because the current mixed evaluation practice makes cross-system comparison unnecessarily hard. Second, Nadi Pariksha research should invest as much in sensor design as in ML methods, because the current limiting factor appears to be sensor-measurable feature coverage rather than classifier quality. Third, publication norms should move toward reporting negative as well as positive results in the AI-for-Ayurveda literature, following the reporting standards now standard in the broader medical-AI field. The AAVI rubric and full system inventory are available from the corresponding author.

References

- Bhattacharya, A. et al. (2023). Large language models for Ayurvedic question answering: an evaluation study. *Journal of Ayurveda and Integrative Medicine*, 14(3), 100725.
- Chaturvedi, A. et al. (2019). An ensemble learning framework for constitutional (Prakriti) classification in Ayurveda. *IEEE Transactions on NanoBioscience*, 18(3), 377-384.
- Dey, S. and Pahwa, P. (2014). Prakriti and its associations with metabolism, chronic diseases, and genotypes: possibilities of new-born screening. *Journal of Ayurveda and Integrative Medicine*, 5(1), 15-24.
- Elangovan, R. and Kumar, K. S. (2018). A machine-learning framework for Ayurvedic Prakriti assessment. *Procedia Computer Science*, 132, 938-947.
- Hankey, A. (2010). Ayurvedic concepts of physiology and health promotion. *Journal of Ayurveda and Integrative Medicine*, 1(1), 28-35.
- Joshi, A. et al. (2011). Development of a three-channel piezoelectric sensor for Nadi Pariksha. *International Journal of Ayurveda Research*, 2(1), 38-43.
- Kallur, P. and Hegde, V. (2020). Deep learning for pulse diagnosis: a CNN approach to Nadi Pariksha. *Biomedical Signal Processing and Control*, 61, 102009.
- Kalra, S. and Narayan, S. (2021). Clinical decision support in integrative Ayurveda: review of deployed systems. *Journal of Ayurveda and Integrative Medicine*, 12(1), 1-11.
- Kumar, A. et al. (2017). An Ayurvedic diagnostic system using machine learning on clinical features. *International Journal of Advanced Computer Research*, 7(29), 1-9.
- Madaan, V. and Jain, S. (2013). Prakriti determination using SVM and decision-tree classifiers: comparative study. *International Journal of Computer Applications*, 72(13), 5-9.
- Menon, P. et al. (2020). Image-phenotype based Prakriti classification with convolutional neural networks. *Indian Journal of Traditional Knowledge*, 19(3), 526-533.
- Patwardhan, B. et al. (2015). Classification of human population based on HLA gene polymorphism and the concept of Prakriti in Ayurveda. *Journal of Alternative and Complementary Medicine*, 11(2), 349-353.
- Patwardhan, K. and Joglekar, R. (2022). Neural retrieval for Ayurvedic formulation recommendation from Charaka Samhita. *Journal of Biomedical Informatics*, 130, 104083.
- Rao, R. R. et al. (2022). End-to-end clinical decision support for Ayurvedic practice: design and pilot evaluation. *Journal of Healthcare Informatics Research*, 6(2), 217-238.
- Sharma, A. et al. (2021). Transformer-based information extraction from classical Ayurvedic texts. *Natural Language Engineering*, 27(6), 713-734.
- Singh, A. et al. (2018). Integrated Ayurvedic CDSS: design and hospital pilot. *International Journal of Medical Informatics*, 117, 39-47.
- Tiwari, P. et al. (2017). Ayurvedic knowledge-graph construction for formulation recommendation. *Procedia Computer Science*, 115, 558-566.
- Valiathan, M. S. (2009). *The Legacy of Caraka*. Orient Blackswan, Hyderabad.
- Verma, V. et al. (2019). A hybrid knowledge-based and machine-learning approach for herbal recommendation. *Computers in Biology and Medicine*, 110, 161-171.

- Wani, M. A. et al. (2020). Machine-learning classification of Ayurvedic Prakriti using lifestyle features. *Journal of Medical Systems*, 44(11), 198.
- World Health Organization. (2013). WHO Traditional Medicine Strategy 2014-2023. WHO, Geneva.
- Ministry of AYUSH, Government of India. (2020). AYUSH Grid: National Digital Health Infrastructure Framework. Ministry of AYUSH, New Delhi.

Declarations

Funding

Funded by the European Integrative Medicine Research Council (EIMRC-2024-041), the Swedish Research Council (VR-2023-00912), and the French Agence Nationale de la Recherche (ANR-23-CE45-0108). The funders had no role in study selection, extraction, analysis, or the decision to publish.

Conflict of Interest

The authors declare no competing interests. None of the authors holds equity in, or has received consulting fees from, any commercial developer of AI-for-Ayurveda systems covered in this review.

Data Availability Statement

The full system inclusion log, frame classification records, and the AAVI scoring spreadsheet are available as supplementary material from the corresponding author on reasonable request. System-level benchmark data remain subject to the original publishers' data-sharing policies.

Ethical Approval

As this is a structured review of previously published AI systems 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/AYUAI8) prior to the start of screening.

Appendix A

AAVI Scoring Rubric and System Classification

Full scoring rubric for the Ayurvedic AI System Validation Index (AAVI), together with the classification criteria used to assign systems to frames. All five AAVI dimensions are normalised to the 0-1 range before the weighted sum is calculated.