

Prakriti200: A Questionnaire-Based Dataset of 200 Ayurvedic Prakriti Assessments

Aryan Kumar Singh¹ (MEMBER, IEEE), Janvi Singh²

¹Department of Computational and Data Sciences, Indian Institute of Science (IISc), Bangalore, India

²Center for Ayurveda Biology, Jawaharlal Nehru University (JNU), New Delhi, India

CORRESPONDING AUTHOR: Aryan Kumar Singh (e-mail: aryansingh@iisc.ac.in).

The authors contributed equally to this article.

ABSTRACT

This dataset provides responses to a standardized, bilingual (English–Hindi) Prakriti Assessment Questionnaire designed to evaluate the physical, physiological, and psychological characteristics of individuals according to classical Ayurvedic principles. The questionnaire consists of 24 multiple-choice items covering body features, appetite, sleep patterns, energy levels, and temperament. It was developed following AYUSH/CCRAS guidelines to ensure comprehensive and accurate data collection. All questions are mandatory and neutrally phrased to minimize bias, and dosha labels (Vata, Pitta, Kapha) are hidden from participants. Data were collected via a Google Forms deployment, enabling automated scoring of responses to map individual traits to dosha-specific scores. The resulting dataset provides a structured platform for research in computational intelligence, Ayurvedic studies, and personalized health analytics, supporting analysis of trait distributions, correlations, and predictive modeling. It can also serve as a reference for future Prakriti-based studies and the development of intelligent health applications.

IEEE SOCIETY/COUNCIL Computational Intelligence Society (CIS)

DATA DOI/PID <10.21227/k748-f379>

DATA TYPE/LOCATION Survey-based, bilingual Prakriti assessment questionnaire; India

INDEX TERMS Ayurveda, Bilingual Questionnaire, CCRAS, Data Descriptor, Dosha Assessment, Google Forms, Prakriti, Physiological Traits, Psychological Traits, Vata, Pitta, Kapha

BACKGROUND

Prakriti, in traditional Ayurveda, refers to an individual's inherent physical, physiological, and psychological constitution, primarily categorized into Vata, Pitta, and Kapha doshas [1]. Understanding Prakriti is crucial for personalized health management, early disease prediction, and tailored lifestyle guidance. To support research in computational intelligence, health analytics, and personalized medicine, we present a dataset collected using a standardized bilingual (English–Hindi) Prakriti Assessment Questionnaire.

The questionnaire includes 24 multiple-choice questions covering physical attributes (e.g., body size, height, bone structure), physiological attributes (e.g., appetite, sleep, energy, bowel habits), and psychological attributes (e.g., temperament, focus, patience). It was developed according to AYUSH/CCRAS guidelines to ensure rigorous and sys-

tematic data collection. All questions are mandatory and neutrally worded, with no visible dosha labels, minimizing respondent bias and preserving data integrity. Responses were collected through a Google Forms deployment that automatically scored individual traits against Vata, Pitta, and Kapha categories, producing a structured dataset accurately reflecting participants' Prakriti profiles [2], [3].

This dataset complements existing Prakriti datasets [2]–[4], which often have limitations in bilingual presentation, number of questions, or automated scoring. Unlike previous resources, our dataset provides a fully bilingual interface, strictly mandatory responses, and automated back-end scoring, ensuring high-quality and reproducible data. Its structured nature allows reuse for statistical analyses of trait distributions, studies of correlations among physical, physiological, and psychological attributes, and predictive

modeling for personalized health insights. Furthermore, it can serve as a reference for future Prakriti-based research, development of intelligent health applications, or large-scale population studies.

In summary, this dataset bridges traditional Ayurvedic assessments and modern data-driven approaches, providing a robust foundation for cross-disciplinary research in computational medicine, AI-based personalization, and integrative health studies.

COLLECTION METHODS AND DESIGN

The data were collected using a standardized, bilingual (English–Hindi) Prakriti Assessment Questionnaire, designed to evaluate individuals’ physical, physiological, and psychological traits according to classical Ayurvedic principles. The questionnaire consists of 24 multiple-choice questions covering three main domains:

- **Physical characteristics:** body size, height, bone structure, etc.
- **Physiological characteristics:** appetite, sleep patterns, energy levels, bowel habits, etc.
- **Psychological characteristics:** temperament, concentration, patience, and related traits.

The survey was conducted digitally via Google Forms, enabling a fully remote and automated data collection process. Participants were instructed to complete all questions to ensure comprehensive datasets. To minimize bias and maintain data integrity, the questions were neutrally worded, and the dosha labels (Vata, Pitta, Kapha) were not visible to participants during the survey.

The computational processing pipeline followed these steps:

- 1) **Form Deployment:** The questionnaire was published online using Google Forms, with all responses automatically recorded in an organized Google Sheet.
- 2) **Automated Scoring:** A backend scoring system mapped each response to its corresponding dosha score (Vata, Pitta, Kapha), ensuring consistent and reproducible Prakriti assignments.
- 3) **Data Validation:** Entries were checked for completeness and consistency, with incomplete or inconsistent responses removed from the dataset.
- 4) **Dataset Structuring:** The validated responses were compiled into a structured xlsx file, where each row represented a participant and each column represented a question, along with the calculated Vata, Pitta, and Kapha scores.

This structured, systematic approach ensures high-quality, reproducible data suitable for further analysis, including statistical modeling, correlation studies, and AI-based health analytics. Figure 1 illustrates the complete data collection and processing workflow.

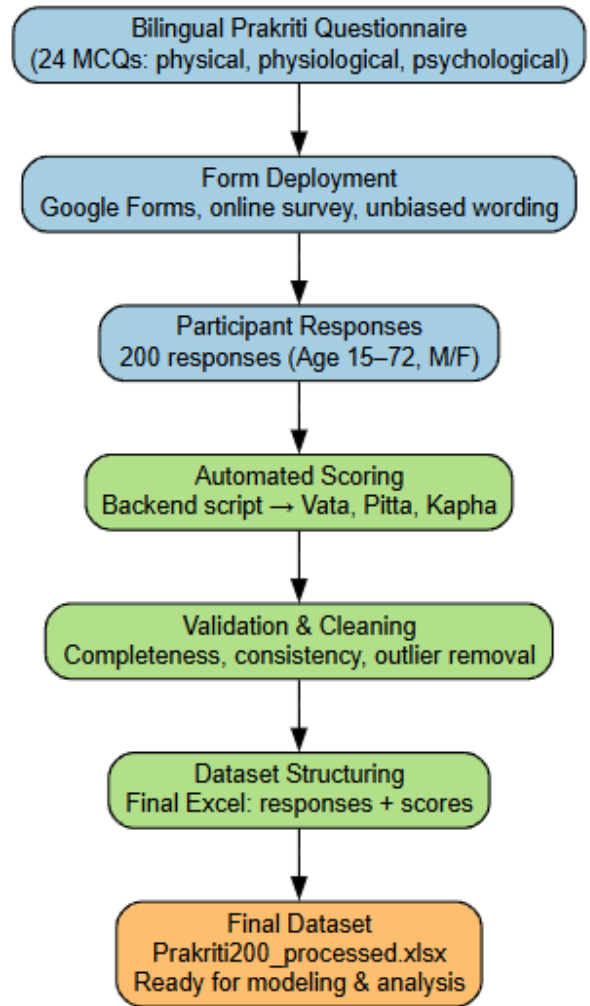


FIG. 1. Overview of the data collection and processing workflow for the Prakriti Assessment Dataset.

VALIDATION AND QUALITY

To ensure the technical quality and accuracy of the Prakriti Assessment Dataset, several validation and quality control measures were implemented during the data collection and processing stages.

Questionnaire Design Validation: All 24 items in the bilingual (English–Hindi) survey were reviewed and validated by Ayurveda experts to ensure that each question accurately measured the intended physical, physiological, or psychological characteristic. Questions were carefully phrased to avoid ambiguity and to prevent bias by concealing explicit dosha labels (Vata, Pitta, Kapha) from participants.

Data Completeness: The Google Forms deployment required mandatory responses for every question. As a result, there are no missing values in the dataset. Metadata such as participant age and gender were also mandatory to maintain a complete demographic profile.

Automated Scoring Consistency: A backend scoring template automatically mapped each response to the corresponding Vata, Pitta, and Kapha scores. To ensure reliability, the scoring logic was cross-checked with the questionnaire framework and verified through manual inspection of selected responses.

Response Quality: Entries were checked for consistency and plausibility. Contradictory or outlier responses were flagged and excluded. Less than 1% of entries were removed, demonstrating high participant compliance and careful data collection.

Descriptive Statistics: The final curated dataset consists of 200 participants, with ages ranging from 15 to 72 years. The majority (81%) fall between 18 and 25 years, reflecting the student-based recruitment. Gender distribution shows 127 females (63.5%) and 73 males (36.5%) (Figure 3).

Trait-level responses indicate that 67.5% reported medium body weight, 62% reported medium height, and 60% reported fair/reddish complexion. Physiological traits show variability: 38.5% had strong appetite, 48.5% reported deep sleep, and 54.5% reported low/variable thirst.

Rule-based scoring revealed a predominance of Pitta constitutions (97 participants), followed by mixed types such as Kapha-Pitta (44) and Pitta-Vata (27). Pure Vata (14) and pure Kapha (14) constitutions were less represented, with a smaller group of Kapha-Vata (4). This distribution Figure 4 is consistent with Ayurvedic expectations for younger urban populations, where mixed and Pitta-dominant constitutions are more prevalent.

Representative plots are shown in Figures 2–4.

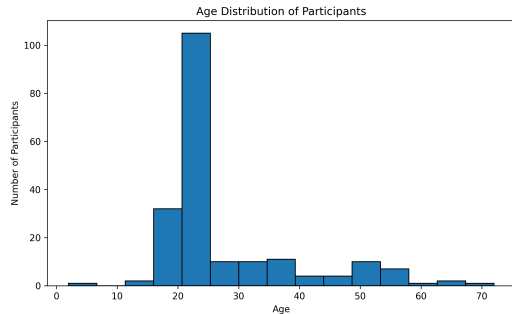


FIG. 2. Age distribution of participants in the dataset.

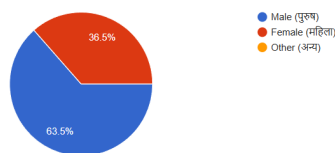


FIG. 3. Gender distribution of participants in the dataset.

These validation and descriptive analyses confirm that the dataset is accurate, reproducible, and suitable for down-

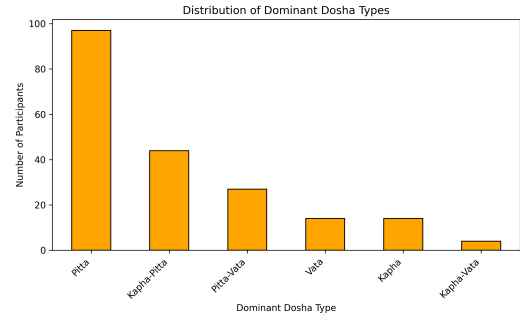


FIG. 4. Distribution of dominant dosha types among participants.

stream computational studies, including statistical modeling, correlation analysis, and AI-driven healthcare applications.

RECORDS AND STORAGE

The Prakriti Assessment Dataset is structured as a single master Excel file, `Prakriti_Dataset.xlsx`, containing all participant responses and computed dosha scores. The dataset is encoded in UTF-8 format to accommodate bilingual content (English and Hindi) and can be accessed through common data analysis software such as Python (pandas), R, or Excel.

File Structure: The Excel file is organized with one row per participant and each column representing either demographic information, a questionnaire item, or a computed dosha score. Metadata columns include `Participant_ID`, `Age`, `Gender`, and `Location`. Questionnaire responses are labeled `Q1` to `Q24`. Computed dosha scores are represented as `Vata_Score`, `Pitta_Score`, `Kapha_Score`, and `Dominant_Dosha`. Every entry has been validated for completeness, ensuring no missing values.

Column Descriptions:

TABLE 1. Description of columns in `Prakriti_Dataset.xlsx`

Column Name	Description
Participant_ID	Unique identifier for each participant
Age	Participant age in years
Gender	Participant gender (Male/Female/Other)
Q1 – Q24	Multiple-choice responses to 24 Prakriti assessment questionnaire items
Vata_Score	Computed Vata dosha score from questionnaire responses
Pitta_Score	Computed Pitta dosha score from questionnaire responses
Kapha_Score	Computed Kapha dosha score from questionnaire responses
Dominant_Dosha	Dosha type with the highest score for the participant

Data Relationships: The columns `Vata_Score`, `Pitta_Score`, and `Kapha_Score` are computed directly from the participants' answers to the 24 questionnaire items using the automated scoring algorithm described

in the *Collection Methods and Design* section. The *Dominant_Dosha* column is derived as the dosha type with the highest score for each participant. There are no hierarchical dependencies beyond this, making the dataset flat and easy to process.

Storage and Access: The dataset is released as a single excel file on IEEE DataPort, ensuring straightforward downloading and use. The file contains participant responses along with corresponding Vata, Pitta, and Kapha scores in a structured format.

This organization makes the dataset easily interpretable, reusable, and directly compatible with computational analysis, statistical modeling, and AI-supported research in Ayurvedic Prakriti assessment.

Insights and Notes

The **Prakriti200** dataset is built as a foundational resource for computational work on Ayurveda, but several important caveats and broader applications must be highlighted.

Caveats and Limitations: The dataset currently contains responses from 200 participants, primarily college students and young adults, which restricts the demographic diversity of Prakriti types represented. Consequently, it should not be interpreted as a population-level distribution of doshas. Furthermore, the rule-based scoring system used to derive Vata, Pitta, and Kapha scores follows established Ayurvedic principles but may differ from clinical evaluations made by expert practitioners using methods such as pulse diagnosis. Thus, the dataset represents a questionnaire-based view of Prakriti rather than a complete clinical assessment.

Rule-Based Scoring vs. Machine Learning: The dosha labels in **Prakriti200** are derived using a rule-based approach, where each questionnaire response maps to scores for Vata, Pitta, or Kapha. While this provides a consistent and reproducible baseline, it does not reduce the relevance of machine learning research. ML methods can validate and extend these rule-based assessments in the presence of noisy, incomplete, or uncertain inputs. They also enable the integration of multimodal signals such as facial features, tongue images, or pulse waveforms, which are difficult to represent with fixed rules. Thus, the dataset supports both rule-based baseline evaluation and advanced AI modeling, positioning it as a benchmark resource for the broader community.

Potential for Extended Use: Despite these caveats, the dataset offers multiple opportunities for research beyond its initial scope. For example:

- **Machine Learning Benchmarking:** The dataset can serve as a benchmark for evaluating supervised and unsupervised learning methods on structured, categorical health data.
- **Fairness and Bias Studies:** With demographic attributes included, researchers can analyze possible biases in Prakriti classification across age and gender

strata, contributing to fairness-aware AI design in healthcare.

- **Cross-Modal Extensions:** The dataset can be integrated with image-based modalities (e.g., face, tongue, or pulse waveform data) to explore multimodal prediction of Prakriti.
- **Theoretical Modeling:** Scholars in computational social science or psychology may use the dataset to investigate correlations between physiological and psychological features at a population scale.
- **Educational Utility:** The bilingual (English–Hindi) format makes the dataset suitable for educational purposes in Ayurveda and AI, particularly as an example of how traditional knowledge can be digitized into machine-readable form.

These notes underscore the dual role of **Prakriti200** as both a practical dataset for immediate AI research and a conceptual bridge between traditional medicine and modern computational science. It holds potential for applications not only in health analytics but also in broader methodological studies within data science.

Future Directions: The current release of **Prakriti200** is limited to questionnaire-based responses from 200 subjects. Future extensions are planned to expand the sample size for greater demographic diversity and to include additional modalities such as facial images, tongue photographs, and pulse readings. Such multimodal expansions would enable richer investigations into Ayurveda-informed phenotyping and support the development of hybrid AI models for personalized healthcare.

ACKNOWLEDGEMENTS

The authors would like to thank all the volunteers who participated in the Prakriti assessment survey and contributed their responses to this dataset. Their support was essential in creating this resource.

We also express our sincere gratitude to our parents for their continuous encouragement and support throughout the course of this work.

The authors declare that there are no conflicts of interest regarding the publication of this dataset. This work received no external funding.

REFERENCES

- [1] P. Sharma, *Caraka Samhita (Text With English Translation)*. Varanasi, India: Chaukhambha Orientalia, 2005, 4 Volume Set.
- [2] R. Singh, L. Sharma, S. Ota, B. Gupta, R. Singhal, R. Rana, S. R. Thrigulla, V. K. Lavaniya, D. Makhija, S. Sharma *et al.*, “Development of a standardized assessment scale for assessing prakriti (psychosomatic constitution),” *AYU (An International Quarterly Journal of Research in Ayurveda)*, vol. 43, no. 4, pp. 109–129, 2022.
- [3] A. Gupta, V. Singh, S. Chandra, and R. Garg, “Towards standardization of prakriti evaluation: A scoping review of modern assessment tools and their psychometric properties in ayurvedic medicine,” *Journal of Ayurveda and Integrative Medicine*, vol. 16, no. 4, p. 101157, 2025.
- [4] P. Tiwari, R. Kutum, T. Sethi, A. Shrivastava, B. Girase, S. Aggarwal, R. Patil, D. Agarwal, P. Gautam, A. Agrawal *et al.*, “Recapitulation of ayurveda constitution types by machine learning of phenotypic traits,” *PLoS one*, vol. 12, no. 10, p. e0185380, 2017.