Human Body Prakriti Identification using Tridosha's

---

1. Abstract

The Prakriti identification project leverages machine learning to classify human body constitutions (Prakriti) into three primary types—Vata, Pitta, and Kapha—based on principles from Ayurveda. By combining Ayurveda's insights with Random Forest classification, this project achieves a high accuracy of 98%, suggesting a promising method for integrating traditional health practices with modern technology. This approach to Prakriti estimation can provide personalized health insights, making it a valuable tool for wellness and preventive care. The project's use of feature extraction and selection, downscaling from over 120 attributes to 13 core characteristics, significantly enhances the model's efficiency and interpretability.

---

2. Introduction

Ayurveda, one of the oldest holistic health systems, emphasizes the significance of body constitution (Prakriti) in personalized healthcare. According to Ayurveda, Prakriti influences various physiological and psychological attributes of an individual, making it essential for understanding predispositions to certain health conditions. The three primary Prakriti types—Vata, Pitta, and Kapha—are based on a balance of Tridoshas, or fundamental energies, which regulate biological functions and define an individual's physical and emotional tendencies.

In this project, Prakriti identification is approached using machine learning, specifically focusing on Random Forest-based classification. By automating Prakriti classification, the project provides a scalable and objective method, allowing for consistent classifications without the subjective bias that often accompanies traditional assessments. Machine learning methods are particularly suitable for this task due to their ability to analyze large datasets and extract meaningful patterns, offering a modern perspective on Ayurveda.

---

3. Novelty

This project introduces a unique interdisciplinary approach by blending Ayurveda's body constitution classification with machine learning. Some key innovations include:

- **Feature Selection using Immutable Characteristics**: The model narrows down over 120 initial attributes to 13 immutable features. These core attributes were identified based on their stability over time, making them ideal for consistent Prakriti classification.
- **Ensemble Approach with Random Forest**: Ensemble learning, particularly Random Forest, was applied due to its effectiveness in handling large datasets and reducing overfitting. This approach provides a high level of accuracy and interpretability, allowing for a more reliable classification model.
- **Application of Technology to Traditional Medicine**: Integrating machine learning with Ayurveda opens up possibilities for personalized healthcare, with Prakriti-based insights being used to inform dietary, lifestyle, and healthcare decisions tailored to an individual's constitution.

This combination of traditional and modern techniques enhances the novelty of this project, as machine learning applications in Ayurveda are still an emerging field.

---

4. Dataset

The dataset used in this project consists of records that capture characteristics corresponding to Vata, Pitta, and Kapha Prakriti classifications. Key details include:

- **Source and Structure**: The dataset was compiled from multiple sources and was refined to balance class representation across Vata, Pitta, and Kapha types. This balance was essential to ensure the model could learn without bias toward any specific Prakriti.
- **Initial Feature Count and Preprocessing**: Initially, the dataset contained over 120 features, capturing various physiological, psychological, and lifestyle attributes associated with Prakriti classification. Data preprocessing included handling missing values, normalizing feature scales, and encoding categorical features for compatibility with the Random Forest model.
- **Class Balance and Quality Assurance**: To ensure robust model training, the data was balanced across Prakriti classes. Quality checks were conducted to ensure data consistency, and the data was further prepared through normalization to enhance the classifier's performance.

---

5. Feature Extraction and Dimensionality Reduction

Feature extraction is a critical component of this project, as the original dataset had over 120 features, which were scaled down to 13 essential characteristics. This process was guided by two main factors:

- **Identification of Immutable Characteristics**: Prakriti traits are typically stable and unchanging in an individual, which led to the focus on immutable characteristics—traits that remain constant over a person's life. These characteristics included physical, emotional, and metabolic attributes that align with Ayurvedic principles.
- **Dimensionality Reduction for Model Optimization**: Dimensionality reduction techniques, such as statistical analysis and domain expertise, helped in selecting features that contributed most to the classification without compromising accuracy. This scaling down not only simplified the model but also enhanced its interpretability, as each of the 13 features has clear relevance to Prakriti classification.

This focused feature extraction approach ensures that the model remains accurate and efficient, minimizing computational overhead while retaining high prediction quality.

---

6. Proposed Method

The proposed method is built around the Random Forest Prakriti Classification (RFPC) algorithm, which is designed to capture the nuanced patterns in the Prakriti dataset. The steps involved include:

- **Model Selection and Justification**: Random Forest was chosen for its ensemble nature, which combines multiple decision trees to enhance stability and accuracy. This ensemble approach is especially effective in handling complex datasets with numerous features, making it ideal for Prakriti classification.
- **Model Architecture and Configuration**: The Random Forest model was configured with specific hyperparameters, including the number of decision trees, maximum tree depth, and minimum samples per split, optimized through cross-validation. These configurations ensured the model could learn effectively from the data while preventing overfitting.
- **Ensemble Learning and Tuning**: In addition to Random Forest, experimentation with other ensemble methods, such as boosting, was conducted to identify the optimal approach for this dataset. Hyperparameter

tuning was performed to maximize the model’s performance across the three Prakriti types.

This approach provides a highly accurate, interpretable model that leverages the strength of ensemble learning to classify Prakriti effectively.

---

7. Results and Analysis

The Random Forest classifier achieved an accuracy of 98% in identifying the correct Prakriti type, underscoring the model's effectiveness. Key points in the results analysis include:

- **Performance Metrics**: The model’s performance was evaluated using metrics such as precision, recall, F1-score, and accuracy. These metrics provided insights into the model's reliability across different Prakriti classes.
- **Feature Importance and Interpretability**: Feature importance analysis highlighted which of the 13 features contributed most to the classification. This analysis aligns with Ayurveda principles, further validating the model’s choices of key features.
- **Visualization and Comparative Analysis**: Confusion matrices and ROC curves were used to visualize classification performance, providing a clear understanding of the model’s strengths and limitations. Comparisons with other classification techniques, such as Support Vector Machines and Decision Trees, were conducted, showcasing the Random Forest model’s superior accuracy and robustness.

This analysis validates the effectiveness of the feature selection and ensemble method, indicating that the model is both accurate and reliable for Prakriti identification.

---

8. Conclusion

In conclusion, the project successfully demonstrates the potential of integrating machine learning with Ayurvedic Prakriti classification. By automating Prakriti identification, this project brings objectivity and consistency to a traditionally subjective field, opening new possibilities for personalized healthcare. The model’s accuracy, feature interpretability, and reliability affirm its effectiveness.

Future directions for this research include expanding the dataset to incorporate more diversity, exploring additional immutable features, and experimenting with advanced ensemble methods like gradient boosting to enhance model accuracy further. This project sets the foundation for a technology-driven approach to Ayurveda, making ancient health practices accessible and actionable for modern applications.
