Phase-Specific Dietary Guidance through Predictive Modeling of Menstrual Phases

Anagha Bhavaraju[1], Krishna Tambatkar[1], Shambhavi Singh[1], Vagisha Prasad[1], S Lalitha[1]

[1] Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
anaghabhavaraju@gmail.com, ktams2530@gmail.com, shambhavi9229@gmail.com, vagisha0705@gmail.com, s_lalitha@blr.amrita.edu

Abstract. Nutrition for women largely changes based on their current menstrual phase due to differences in energy levels, hormones, and other health factors. To address this, the proposed work aims to help the menstruator accurately predict their menstrual phases and have comprehensive access to the required and recommended nutrition for it. This work presents a machine learning approach for predicting menstrual cycle phases and providing personalized dietary recommendations, with a focus on utilizing the XGBoost algorithm with SMOTE and hyperparameter tuning. The model processes user-input data based on their menstrual cycle to accurately predict phases of the cycle with an average accuracy of 99.39%. XGBoost was selected for its superior performance in handling imbalanced datasets, which is crucial for effectively predicting less frequent phases such as Ovulation. By leveraging a phase-specific approach, the system ensures that nutritional advice is aligned with the user's unique cycle patterns, improving overall well-being. The unique approach of combining accurate phase prediction with personalized nutrition recommendations ensures actionable insights for users. This work demonstrates the potential for real-world applications in personalized healthcare, particularly in managing menstrual health and nutrition more effectively.

Keywords: Menstruation, Menstrual Phases, Nutrition, Health Informatics, Machine Learning.

1 Introduction

The menstrual cycle, influenced by hormonal changes, is critical to women's health, affecting physical well-being, psychological balance, and dietary needs. Research shows that understanding these fluctuations can improve health management and lifestyle choices. Nutrition and exercise recommendations vary by cycle phase, influencing energy, mood, and metabolism [1][2]. The Cleveland Clinic notes that

adapting nutrition and exercise according to menstrual phases can optimize health outcomes [3].
Moreover, aligning dietary intake with the different phases of the cycle can enhance nutrient absorption and metabolic efficiency, potentially leading to better overall health.

However, many women struggle with menstrual health issues such as PMS and irregular cycles, which can diminish quality of life and increase healthcare costs. Tailored nutritional strategies may alleviate these symptoms [1][2] and reduce healthcare interventions.

Despite this, there is a gap in applying nutritional knowledge in practical frameworks, particularly regarding personalized dietary recommendations. While many smartphone apps track menstrual cycles, few offer personalized dietary advice based on specific phases [4]. Advanced machine learning techniques, such as artificial neural networks, have potential for improving cycle predictions [5], but their application for integrating dietary recommendations is limited [6]. Furthermore, existing models often overlook the complexity of individual dietary needs and preferences, which can vary significantly among women.

Hybrid models that combine traditional methods with machine learning show potential for comprehensive women's health management, especially regarding nutrition during the menstrual cycle [7]. Research from the NIH emphasizes the importance of personalized health interventions tailored to menstrual fluctuations [8].

The menstrual cycle consists of four phases: menstrual, follicular, ovulatory, and luteal. In this work, "menses" refers specifically to the menstrual phase, characterized by active bleeding at the beginning of the cycle. This work aims to bridge gaps by developing a predictive model that improves phase-specific cycle predictions and integrates dietary recommendations, ultimately enhancing women's health management throughout the menstrual cycle.

2 Related Works

Recent advancements in machine learning and data analytics have significantly improved the ability to predict menstrual cycle-related information and even identify possible menstrual disorders. While several studies have explored different datasets and techniques to provide insights on these fronts, a notable scarcity of work on menstrual cycle phase-based nutritional guidance tools is observed. Studies indicate that foodrelated requirements vary across different phases of the menstrual cycle, suggesting that women could benefit from dietary recommendations tailored to their hormonal changes [1]. These works highlight the importance of specific nutritional practices in alleviating menstrual symptoms, demonstrating a positive correlation between certain diets and symptom relief [2].

Kriti N. et al. [5] explored the use of Artificial Neural Networks (ANNs) to predict menstrual cycle lengths based on previously self-tracked cycles. The model achieved a 98.08% accuracy rate with a 1.92% error margin, showcasing the strength of machine learning techniques in capturing complex patterns in menstrual data.

Another application of self-tracked data was demonstrated by K. Li et al. [9], wherein a Hierarchical Generative Model using mobile health data and self-tracking data achieved a mean absolute error (MAE) of 1.6 days in predicting menstrual cycle lengths. Both studies utilized self-tracked data, which remains ambiguous, as its accuracy cannot be perfectly determined in the context of menstrual data.

Prof. Dr. Rosana Rego et al. [10] utilized Time-Series Forecasting (ARIMA) to model menstrual cycle length based on historical data, achieving a Mean Absolute Percentage Error (MAPE) of 5%. This approach illustrates the efficacy of time-series models in predicting menstrual cycle patterns, particularly when sufficient historical data is available.

Deokule et al. [11] employed Stacked Regression combined with ARIMA to predict menstrual cycles, reporting an RMSE of 0.04 days during menstruation. The model demonstrates improved accuracy as more cycle data becomes available over time. While these models primarily focused on cycle length and symptom prediction, there is a lack of phase-based predictions for dietary recommendations.

Leveraging a Decision Tree model, Odichukwu J.C. et al. [6] estimated ovulation days, achieving a high correlation ($R^2 = 0.9864$) in predicting ovulation days. This work further demonstrates the effectiveness of decision trees in fertility tracking and cycle prediction based on menstrual cycle datasets. Multiple sophisticated algorithms were tested, and a Decision Tree model was determined to yield the best results. Although significant work has been done on ovulation day prediction, other menstrual cycle phases have not been investigated thoroughly.

Ankita Karia et al. [12] utilized a Random Forest algorithm on data collected from a cycle and symptom tracker mobile app to predict menstrual periods and detect early signs of polycystic ovary syndrome (PCOS) with a 90.44% accuracy rate. This model emphasizes the utility of mobile health applications in collecting real-time user data for accurate health predictions. While crucial work on PCOS detection has been observed, a deeper exploration of the impact of dietary choices during different menstrual phases has not been discussed.

The utilization of Hidden Semi-Markov Models to track menstrual health by labeling menstrual data was performed by L. Symul et al. [13], achieving 90% accuracy on data with realistic missingness. The model successfully predicted the next cycle's length, demonstrating the benefits of incorporating biological markers into menstrual cycle prediction. While this model worked on predicting the length of the next period, it did not explore the other menstrual phases.

Logapriya E. et al. [7] designed a Hybrid Recommendation System based on collaborative and content-based filtering. Using user data, self-reported symptoms, and health records, the model achieved 94% accuracy in recommending foods that align with specific menstrual challenges, such as hair fall, stomach pain, and back pain. Although this work made significant strides in relating diet to menstrual troubles, it did not address the overall well-being of individuals with regular menstruation and their nutritional needs. The present approach offers a more inclusive model that predicts all four phases of the menstrual cycle: menstrual, follicular, ovulatory, and luteal. This provides a personalized approach through the inclusion of dietary recommendations tailored to each phase, based on research into phase-specific diets and their importance [14]. What sets the present approach apart is its focus on addressing gaps in the existing literature by offering a thorough prediction model that also includes nutritional recommendations.

The gaps identified and addressed are categorized as follows:

- Lack of focus on phase-based nutritional guidance throughout the menstrual cycle.

- Limited work on predicting phases such as menstrual, follicular, and luteal, despite the exploration of ovulation prediction.
- Many models focus on symptom prediction (e.g., PCOS detection) but do not offer dietary advice for regular menstrual phases.

Integrating phase prediction with nutritional needs enhances the value of the proposed work, offering a personalized and scientific approach to support women's health during the menstrual cycle.

3 Methodology

This section outlines the approach used to predict menstrual cycle phases and provide dietary recommendations. This involves utilizing phase prediction models based on cycle data and mapping it to specific diet plans tailored for each phase, ensuring personalized nutritional advice for users.

3.1 Flow of the Proposed Workflow

Fig. 1. illustrates the methodology used for predicting menstrual cycle phases and providing dietary recommendations. The diagram outlines the key components and processes involved, emphasizing the integration of predictive analytics with tailored nutritional strategies for women's health.

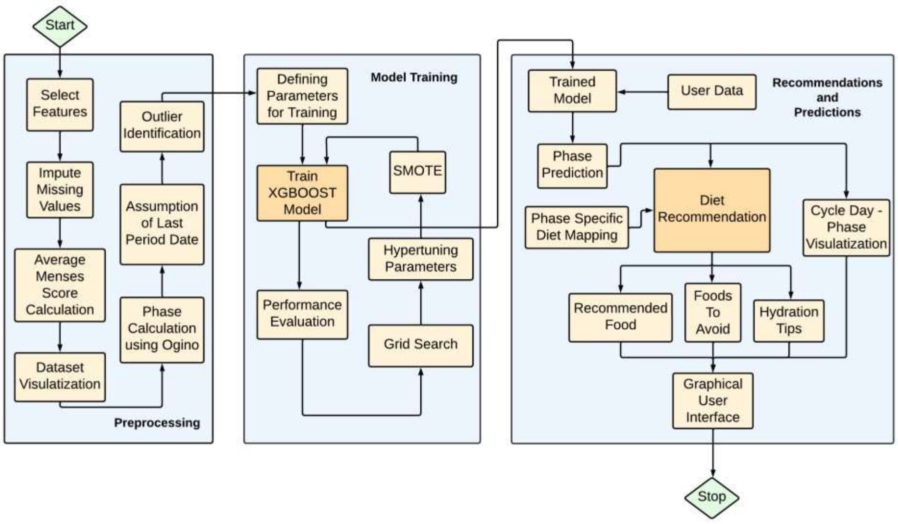

Fig. 1. Workflow diagram

3.2 Preprocessing

Feature Selection and Data Cleaning. The Menstrual Cycle Dataset consists of 1,408 instances and 80 features, with few of the attributes of instances containing missing values. Some columns, such as *Reproductive Category*, were removed due to their lack of relevance. Additionally, features with excessive empty values have been excluded. Although the dataset does not contain a specific target variable, it offered sufficient information to predict menstrual phases based on the available data [15].

Outliers in the dataset were identified from data plot visualization. As seen in the Fig. 2, certain users reported length of menses exceeding 10 days, which was considered an outlier. These outliers were removed from the dataset to maintain data integrity and improve the accuracy of the model.

The key features used from this dataset include *Length of Cycle*, *Menses Length*, *Cycle Day*, and *Mean Menses Score*, which was calculated using the menses scores for each bleeding day. The parameter *Last Period Start Date* was generated by selecting a random integer between 1 and 31 (representing the number days in a month) using a uniform distribution to simulate a sample cycle day for each instance.

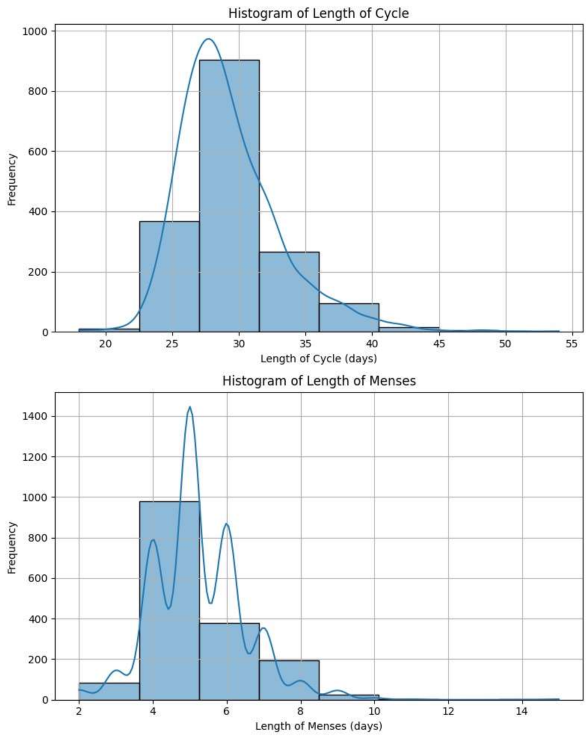

Fig. 2. Histograms of the **(a)** distribution of cycle lengths **(b)** menses lengths among users. These visualizations help identify common ranges and any outliers in the dataset.

Data preprocessing involved the removal of columns such as *Length of Luteal Phase*, *Estimated Day of Ovulation*, *Current Date*, and *Last Period Start Date*, etc. The first two parameters mentioned were excluded because such information is typically not provided by users. As a result, these features were not included to ensure the model could be used with more commonly available data.

Phase Calculation. The Ogino method, developed by a Japanese Gynaecologist, Kyusaku Ogino, is a calendar-based technique for estimating the phases of a menstrual cycle. In this method, ovulation is estimated to occur on a single day, typically around 14 days before the onset of the next menstrual period [16].

- Ovulation day: Estimated to occur 14 days before the next expected period.
- Fertile Window: To increase accuracy, a window is created around the ovulation day.

For dietary recommendations, focusing solely on the ovulation day is insufficient. Since fertility is typically highest in the days leading up to ovulation, it is more effective to implement a nutritional strategy throughout the entire fertile window. In the proposed work, a 6-day fertile window is considered for phase calculation. This approach ensures that the diet not only supports ovulation but also provides essential nourishment during the late follicular and early luteal phases.

Nutrient-rich foods like antioxidants (found in berries), for example, can protect eggs from oxidative stress, optimize hormonal balance, and improve egg quality during this critical time, as highlighted in recent studies [17][18]. Once the phases were calculated through this method, data visualization was applied to analyze the distribution of phases. By considering a fertile window around ovulation, the frequency of ovulation phase increased significantly, improving the model's efficiency. However, a bias remained in the dataset due to the uneven distribution of phases. Phases like menstrual and ovulation, which occur over fewer days, were less frequent compared to other phases. Alternative methods were employed to address this imbalance, and their implementation will be discussed in the following section.

3.3 Model Training

The features used for model training are:

- Cycle Length: The total number of days in a menstrual cycle
- Length of Menses: The number of days of bleeding (the menstrual phase)
- Mean Menses Score: Calculated average score, indicates the severity of symptoms during menstruation
- Cycle day: Day within the current cycle, calculated as (last period start date) - (current date)

An XGBoost classifier was utilized for phase prediction based on the parameters. XGBoost has several mechanisms to handle imbalanced data effectively, including the use of the scale pos weight parameter which ensures that the model does not overly favor the majority classes, namely the follicular and luteal phases. Moreover, it manages missing data through its tree-splitting process, ensuring that all available data contributes to predictions even when certain phases have less data.

Handling Class Imbalance. As seen in Fig. 3, a bias remains between the classes in the dataset, particularly concerning the ovulation phase. To address the data imbalance present, SMOTE (Synthetic Minority Over-sampling Technique) is employed. SMOTE generates synthetic instances for underrepresented phases by interpolating between existing data points. While XGBoost’s scale pos weight parameter adjusts for class imbalance, SMOTE further enhances the classifier's ability to generalize across all menstrual phases.

Hyperparameter Tuning. The performance metrics post-SMOTE indicated high precision, recall, and F1 scores for the follicular, luteal, and menstrual phases, but the ovulation phase still exhibited comparatively lower performance. Despite dataset enrichment through SMOTE improving class balance, the unique characteristics of ovulation, such as its smaller data size, presented challenges. Grid search played a crucial role in optimizing the model's hyperparameters, allowing it to adapt to the diverse features of each menstrual phase. While high performance was achieved for most phases, further fine-tuning, particularly for ovulation, enhanced prediction accuracy.

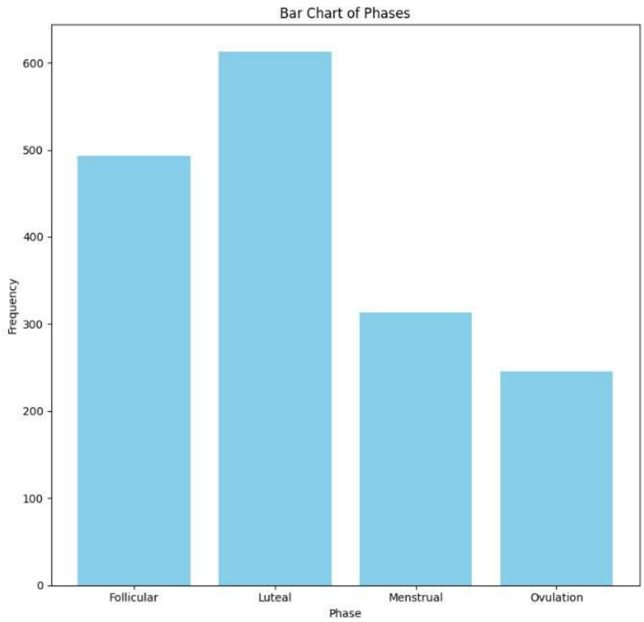

Fig. 3. Bar plot showing the distribution of phases calculated using the Ogino method.

3.4 Recommendations and Predictions

Once the menstrual phase was predicted, personalized dietary recommendations tailored to the user's current phase were provided. These recommendations were customized based on the user's dietary preference (non-vegetarian, vegetarian, or vegan). Table 1 presents the dietary recommendations associated with the proposed approach.

Table 1. Dietary Recommendations for the Follicular Phase.

Phase	Diet Type	Foods	Hydration	Foods to Avoid

Follicular	Vegetarian	Lentils, Zucchini, Bell Peppers, Pumpkin Seeds	Focus on hydrating with electrolyterich drinks such as coconut water and green juices.	Refined carbs, Excessive sugar, Heavy fats
	Non-Vegetarian	Turkey, Tuna, Lean Beef		
	Vegan	Chia Seeds, Oats, Flaxseeds, Almonds		

4 Results and Analysis

This section outlines key findings from training the XGBoost model on the pre-processed data. Major results include improved performance through data balancing and tuning, addressing class imbalances, particularly for the ovulation phase. The model's ability to accurately predict all cycle phases was assessed using precision, recall, and F1-score. A graphical user interface (GUI) was also developed to provide dietary recommendations based on predicted phases.

4.1 Key Results

Model Performance. The XGBoost model initially struggled with biased predictions due to the underrepresentation of the ovulation phase. To correct this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, generating additional samples for ovulation. This balanced the dataset, enhanced the model's generalization ability, and reduced the risk of overfitting on more common phases. Table 2 and Table 3 demonstrate the change in performance after applying the mentioned techniques.

Table 2. Before SMOTE and Hyperparameter Tuning.

Phase	Precision	Recall	F1-Score
Follicular	0.99	0.97	0.98
Luteal	0.99	0.96	0.97
Menstrual	0.99	1.00	0.99
Ovulation	0.88	0.98	0.93

This technique balanced the dataset and exposed the model to more diverse examples, improving its ability to handle data variability. Hyperparameter tuning further optimized performance, especially in precision and recall for the ovulation phase, enhancing the model's reliability in predicting menstrual cycle phases.

Table 3. After SMOTE and Hyperparameter Tuning.

Phase	Precision	Recall	F1-Score
Follicular	1.00	0.99	0.99
Luteal	1.00	0.99	1.00
Menstrual	1.00	1.00	1.00
Ovulation	0.96	1.00	0.98

The final XGBoost test accuracy increased from 97.29% to 99.39%, demonstrating the tangible benefits of applying SMOTE and tuning the model. This improvement suggests that the model's predictions became more consistent and reliable.

Fig. 4 illustrates the comparison of Precision, Recall, and F1-Score for the Ovulation phase, as well as the overall accuracy before and after applying SMOTE and hyperparameter tuning. The results indicate a notable improvement across all metrics, further validating the effectiveness of the applied techniques. These enhancements suggest that the applied techniques significantly boosted the model's performance for the Ovulation phase.

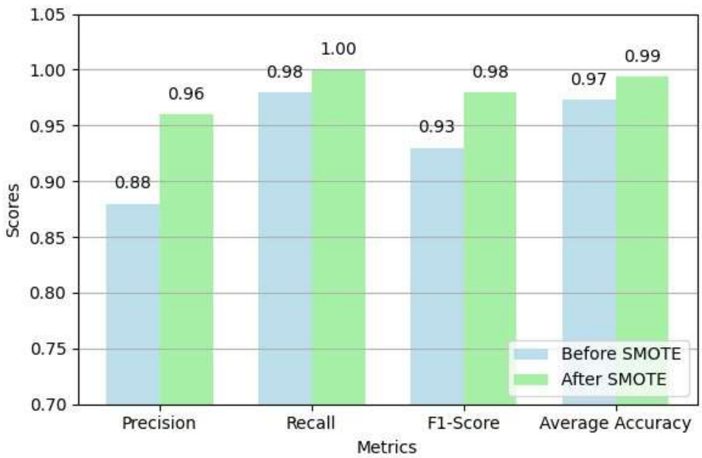

Fig. 4. Ovulation Phase Metrics Before and After Along with the Change in Overall Accuracy.

K-Fold Stratified Cross-Validation with SMOTE. A 5-fold stratified cross-validation was applied to assess the model's generalization. The average classification report presented in Table 4 demonstrates a well-balanced performance across different menstrual cycle phases, despite inherent class imbalances. The precision, recall, and F1scores are consistently high for all phases, indicating that the model effectively identifies each phase with minimal misclassification. Notably, the model performed well for the Follicular and Luteal phases, reflecting it's robustness in predicting these critical stages. The Ovulation phase, while showing slightly lower precision and recall, still maintains a commendable F1-score, suggesting that the model has managed to generalize well across various folds. An average accuracy of 98.62% was achieved.

Table 4. Average Classification Report

Phase	Precision	Recall	F1-Score
Follicular	0.9820	0.9858	0.9839
Luteal	0.9870	0.9886	0.9878
Menstrual	0.9936	0.9935	0.9936
Ovulation	0.9600	0.9470	0.9528

Visualization of Complete Cycle. The prediction timeline window Fig. 5 provides a comprehensive display of the predicted phases corresponding to the user input, including cycle and menses length.

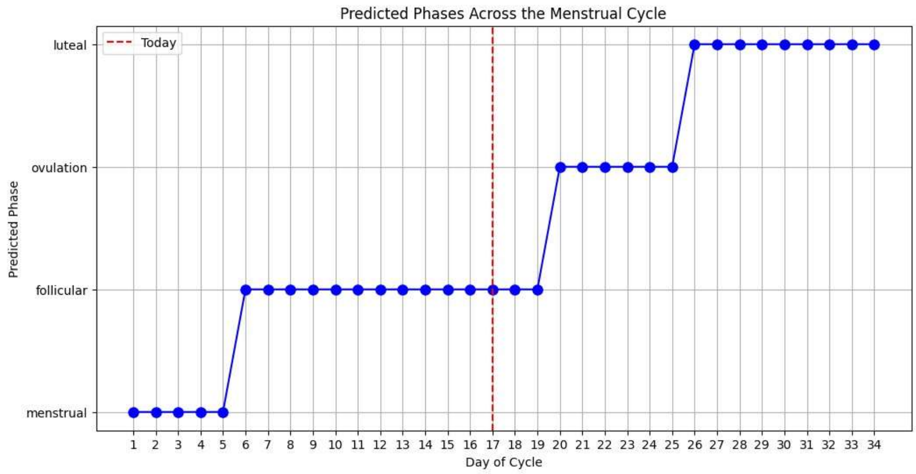

Fig. 5. Sample window with predicted phases across cycle.

4.2 GUI Component

The developed Graphical User Interface (GUI) allows users to input their menstrual cycle data, receive predictions for all phases, and access corresponding dietary advice. Fig. 6 shows the input window and Fig. 7 depicts the results window with the predicted phase and corresponding dietary recommendations along with the prediction timeline window.

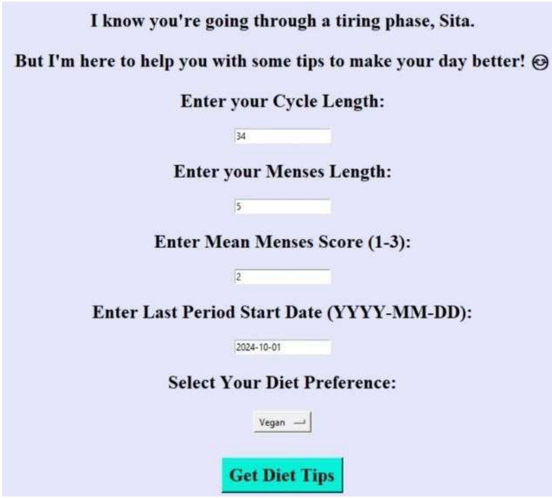

Fig. 6. GUI input window for user cycle details.

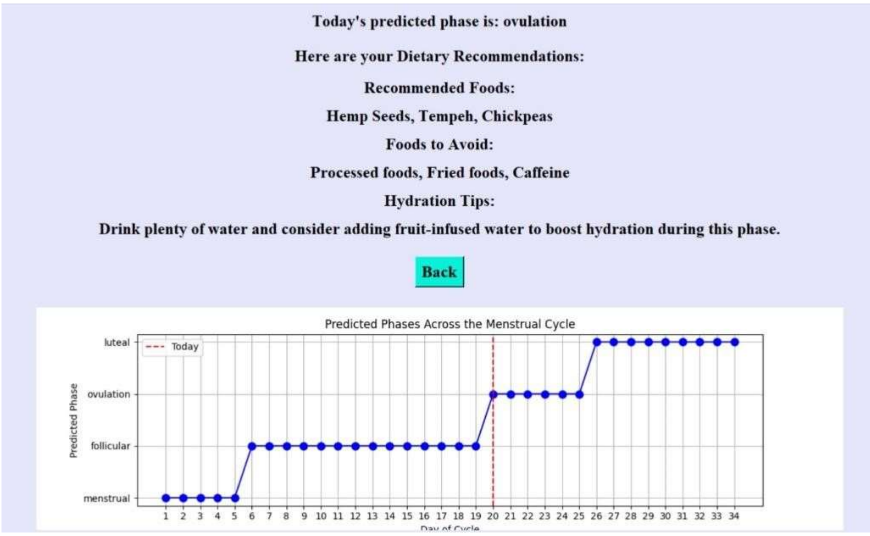

Fig. 7. Results window showing predicted ovulation phase and dietary recommendation.

5 Discussion

5.1 Significance of Results

The proposed model successfully predicts menstrual cycle phases with an average accuracy of 99.39%, demonstrating its effectiveness in real-world healthcare applications.

The high accuracy and balanced performance across all phases, including challenging phases like Menstrual and Ovulation, show that the model is effective for real-world applications in personalized healthcare. The use of SMOTE and hyperparameters for handling imbalanced data further improved predictions for minority classes. The multiclass log loss (mlogloss) value of **0.0624** indicates a robust model performance, as lower mlogloss values are desirable, suggesting that the predictions closely align with the actual outcomes.

5.2 Comparison with Prior Work

Prior research has largely concentrated on specific aspects of menstrual cycle prediction, such as cycle length, ovulation, or general symptoms. These studies have used machine learning techniques to predict individual phases or cycle timing, with limited scope in terms of the comprehensive phases of the cycle. While many works provide cycle length or ovulation predictions, they do not extend their focus to include all phases of the menstrual cycle. Additionally, some studies have integrated dietary recommendations, but these tend to be symptom-based rather than tailored to the distinct phases of the menstrual cycle [7].

In contrast, the present approach offers a more comprehensive model that predicts all four phases of the menstrual cycle: menstrual, follicular, ovulatory, and luteal. This expanded coverage enables a more personalized approach, particularly in the context of diet recommendations, which are tailored to each phase rather than being generalized or symptom-based alone. Such phase-specific dietary interventions mark a notable improvement over existing methods, which typically do not account for the complete range of physiological changes throughout the menstrual cycle and their influence on nutritional requirements. Moreover, the integration of phase-specific recommendations into a cohesive model not only enhances the predictability of the cycle but also ensures that women receive dietary guidance that aligns with their physiological needs at each stage.

This distinction makes the present approach unique, addressing a gap in prior literature by offering a complete prediction model that also includes dietary recommendations. The ability to integrate phase-specific nutrition makes the model more valuable, providing a personalized and scientifically grounded approach to supporting women's health throughout the menstrual cycle. The Table 5 summarizes various studies focused on menstrual cycle prediction, phase identification, and related health recommendations. It compares their techniques, datasets, performance metrics, and results with the proposed work.

Table 5. Comparative Analysis of Previous Works with the Proposed Work

Year	Author(s), Reference	Dataset Type	Technique	Performance	Results
2018	Kriti N., et al., [5]	Previous 11 tracked cycle lengths	Artificial Neural Networks (ANN)	1.92% error possibility and 98.08% accuracy	Next cycle length prediction
2018	Prof. Dr. Rosana Rego, et al., [10]	Menstrual cycle length	Time series forecasting (ARIMA)	MAPE = ~5%	Provided effective time-series prediction

		data (time series)			model for cycle lengths
2018	Deokule, et al., [11]	Physiological signals: acceleration, EDA, blood pressure	Stacked regression with ARIMA	RMSE of 0.04 days during menstruation	Improved prediction accuracy as cycle progresses
2019	I. Urteaga, et al., [9]	Mobile health data, selftracking data	Hierarchical, generative model	MAE of 1.6 days	Menstrual cycle lengths prediction
2019	L. Symul, et al., [13]	Temperature, Cervical Mucus, Menstruation, and Luteinizing hormone tracking	Hidden Semi-Markov Mod-els	~80-85%	Prediction of next cycle
2023	Ankita Karia, et al., [12]	Cycle and Symptom Tracker Mobile App	Random Forest	90.44%	Period prediction, early PCOS detection
2023	Logapriya E., et al., [7]	User data, self-reported symptoms, health data	Hybrid recommendation system (Collaborative and Content Based Filtering)	94%	Recommending foods that align with specific menstrual symptoms
2023	Odirichukwu J., et al., [6]	Menstrual Cycle Data	Decision Tree	R^2 of 0.9864	Ovulation day prediction
2024	Proposed Work	Menstrual Cycle Data	XGBoost	99.39%	Prediction of all phases

5.3 Limitations

While the model performed well, certain limitations were noted. The Ovulation phase posed challenges due to its inherent variability, affecting precision and recall. Additionally, the Ogino method for phase calculation, while functional, is less accurate compared to newer methods. Integrating advanced phase prediction techniques could enhance precision and reliability. The dataset, although diverse, lacked substantial irregularities, particularly those related to PCOS and PCOD, limiting the model's generalizability. Moreover, reliance on manually entered data restricts scalability, which could be improved with real-time data from wearable devices.

In summary, the model demonstrates strong performance across phases and holds potential for real-world personalized healthcare applications.

6 Conclusion and Future Scope

This work presents a comprehensive approach to addressing the lack of personalized dietary recommendations for various menstrual phases using machine learning techniques. XGBoost was selected for its robust performance, achieving a significant accuracy of 99.39% with the help of SMOTE and hyperparameter tuning. The model effectively predicts all four menstrual phases—follicular, ovulation, luteal, and menstrual— demonstrating its capacity to handle class imbalances, especially in the ovulation phase.

The system offers users real-time nutritional recommendations based on their dietary preferences through a simple, accessible GUI. Unlike other works that focus on predicting cycle lengths or ovulation, this model's phase-specific approach supports planning nutrition for the current phase, providing a more comprehensive tool for menstrual health management.

Future work can enhance this model by:

- Predicting early follicular, early luteal, and late luteal phases for more precise recommendations.
- Offering fertility-supporting food recommendations.
- Considering allergies, dietary restrictions, and nutritional deficiencies.
- Improving phase prediction for irregular menstrual cycles.

In summary, this research lays the foundation for personalized health recommendations, with potential for further refinement and broader healthcare integration.