paper / I / we / project -
the proposed work.

Predictive Nutritional Modeling for Phases of Menstruation

Anagha Bhavaraju[1], Vagisha Prasad[1], Krishna Tambatkar[1], Shambhavi Singh[1], Dr. S. Lalitha[2]

[1] *Department of Electronics and Computer Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, Karnataka, 560035, India*

[2] *Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, Karnataka, 560035, India*

anaghabhavaraju@gmail.com, vagisha0705@gmail.com, ktams2530@gmail.com, shambhavi9229@gmail.com, s_lalitha@blr.amrita.edu

unlike the existing works that

uniquely

***Abstract*—Nutrition for women largely changes based on their current menstrual phase due to difference in energy level, hormones and other health factors. To address this, this paper aims to help the menstruator accurately predict their menstrual phases and be able to have a comprehensive access to the required and recommended nutrition for it. This paper presents a machine learning approach for predicting menstrual cycle phases and providing personalized dietary recommendations, with a focus on utilizing the XGBoost algorithm. The model processes user-input data about 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 applying SMOTE to balance the dataset and incorporating advanced feature selection techniques, the model successfully reduces errors and enhances prediction accuracy. The personalized dietary recommendations are then generated based on the predicted phases, providing actionable insights for users. This model demonstrates the potential for real-world applications in personalized healthcare, particularly in managing menstrual health and nutrition more effectively.**

how much

Keywords: Menstruation, Phase, Diet, Health Informatics, ML

I. Introduction

The menstrual cycle, influenced by hormonal changes, is crucial to women's health, affecting physical and psychological well-being and dietary choices. Nutrition and exercise recommendations vary by cycle phase, impacting energy, mood, and metabolism [1][2]. The Cleveland Clinic notes that adapting nutrition and exercise according to menstrual phases can optimize health outcomes [3]. However, many women struggle with menstrual health issues like PMS and irregular cycles, which can diminish quality of life and increase healthcare costs. Research suggests that tailored nutritional strategies may alleviate these symptoms [1][2].

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].

Hybrid models that combine traditional methods with machine learning show promise 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]. This study aims to bridge these gaps by proposing a predictive model that improves menstrual cycle phase predictions and integrates personalized dietary recommendations, ultimately enhancing women's health management during the menstrual cycle.

author et al
suggests

II. Related Works

This table summarizes various studies focused on menstrual cycle prediction, phase identification, and related health recommendations. It compares their techniques, datasets, performance metrics, and results, providing insight into the strengths and weaknesses of each approach. Recent advancements in machine learning and data analytics have significantly improved the ability to predict menstrual cycle related information and

In another
work
author.

author et al

TABLE I

COMPARISON OF MENSTRUAL CYCLE PREDICTION AND RELATED RESEARCH

Year

~~Reference~~	Author(s) and Year	Dataset Type	Technique	Performance	Results
[13]	Ankita Karia et al. (2023) (15)	Cycle and Symptom Tracker Mobile App	Random Forest	90.44%	Period prediction, early PCOS detection
[7]	Logapriya E., Surendran R. (2024) (1	User data, self-reported symptoms, health data	Hybrid recommendation system (Collaborative and Content Based Filtering)	94%	Recommending foods that align with specific menstrual symptoms
[14]	K. Li, I. Urteaga, et al. (2019)	Mobile health data, self-tracking data	Hierarchical, generative model	MAE of 1.6 days	Menstrual cycle lengths prediction
[15]	Odirichukwu J. C.I, Njoku O. A., et al. (2023)	Menstrual Cycle Data (same dataset)	Decision Tree	R^2 of 0.9864	Ovulation day prediction
[16]	L. Symul, S. Holme (2019)	Temperature, Cervical Mucus, Menstruation, and Luteinizing hormone tracking	Hidden Semi-Markov Models	80-85%	Prediction of next cycle
[5]	Kriti N., Priyanshi G., et al. (2018)	Previous 11 tracked cycle lengths	Artificial Neural Networks (ANN)	1.92% error possibility and 98.08% accuracy	Next cycle length prediction
[17]	Prof. Dr. Rosana Rego (2018)	Menstrual cycle length data (time series)	Time series forecasting (ARIMA)	MAPE = 5%	Provided effective time-series prediction model for cycle lengths
[18]	Deokule (2018)	Physiological signals: acceleration, EDA, blood pressure	Stacked regression with ARIMA	RMSE of 0.04 days during menstruation	Improved prediction accuracy as cycle progresses

even identify possible menstrual disorders. While several studies have explored different datasets and techniques to provide insights on these fronts, a notable scarceness of work on lack of menstrual cycle phase based nutritional guidance tools is noticeable. Studies indicate that food-related behaviours vary across different phases of the menstrual cycle, suggesting that women could benefit from dietary recommendations tailored to their hormonal changes [1]. They have highlighted the importance of specific nutritional practices in relieving menstrual symptoms, showing a positive link between certain diets and symptom relief [2]. The use of Artificial Neural Networks (ANNs) to predict menstrual cycle lengths based on previously self-tracked cycles has been explored in [5]. Their model achieved a 98.08% accuracy rate with a 1.92% error margin, showcasing the strength of machine techniques in capturing complex patterns in menstrual data. Another usage of self-tracked data was demonstrated in [14] 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 of these works have made use of self-tracked data which continues to be ambiguous as it's accuracy cannot be determined perfectly in the sense of menstrual data. [17] made use of 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. Stacked Regression combined with ARIMA to predict menstrual cycles in [?], reporting an RMSE of 0.04 days during menstruation. Their model demonstrates improved accuracy as more cycle data becomes available over time. These models also targeted cycle length and symptom prediction but no phase based prediction for further nutritional recommendations was observed. Leveraging a Decision Tree model, estimated ovulation days were predicted achieving a high correlation (R^2 = 0.9864) in predicting ovulation days in [6]. Their study further shows the effectiveness of decision trees in fertility tracking and cycle prediction based on menstrual cycle datasets. Multiple sophisticated algorithms were tested and Decision Tree model was determined to give the best results. The work done in this paper acts as a precursor to working on better dietary advice based on individual menstrual cycle phases. This work showed significant work for ovulation days prediction but did not investigate other menstrual cycle phases. The use of Random Forest algorithm on data collected from a cycle and symptom tracker mobile app was demonstrated in [13] to predict menstrual periods and detecting early signs of poly-cystic 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. Crucial work on PCOS detection through the symptoms and correlation with the luteal and follicular phases was observed but a deeper dive into the affect of diet choices during these phases was not discussed. The utilization of Hidden Semi-Markov Models to track menstrual health by labelling menstrual data was done in [16] and a 90% accuracy was achieved on data with realistic missingness. Their model successfully predicted the next cycle's length, demonstrating the benefits of incorporating biological markers into menstrual cycle prediction. While this model worked on prediction of the length of the next period, it did not explore the other menstrual phases. A Hybrid Recommendation System based on collaborative and content-based filtering was designed in [7]. Using user data, self-reported symptoms, and health records, their model achieved 94% accuracy in recommending foods that align with specific menstrual challenges like hair fall, stomach pain and back pain. Although this work made significant strides in the domain of relation of diet with menstrual troubles, it still did not address general well-being of a regular menstruation individual and their nutritional needs.

The present approach put forth offers a more inclusive model that predicts all four phases of the menstrual cycle: menstrual, follicular, ovulatory, and luteal. This provides a more personalized approach, through the inclusion of diet recommendations, which are tailored to each phase. This has been done through research into phase-specific dietary and their importance.[19] [illegible]

What sets the present approach apart is addressing a gap in prior literature by offering a thorough prediction model that also includes nutritional recommendations. The ability to integrate phase prediction with its respective nutritional requirements makes the model more valuable, providing a personalized and scientifically grounded approach to supporting women's health throughout the menstrual cycle.

III. METHODOLOGY

One fig

A. Preprocessing

1) Feature Selection and Data Cleaning

The Menstrual Cycle Dataset consisted 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 [9].

The key features used from this dataset included 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.

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 former two parameters 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.

Outliers in the dataset were identified from data plot visualization. As seen in the fig. 2, certain

Small para

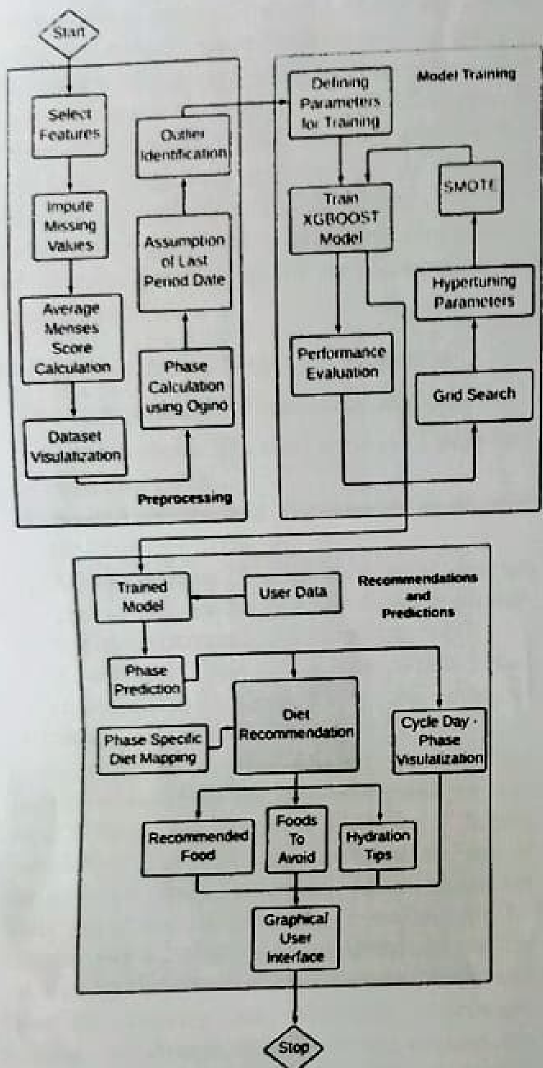

Fig. 1. Flowchart illustrating the methodology used for predicting menstrual cycle phases and providing dietary recommendations.

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.

2) 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

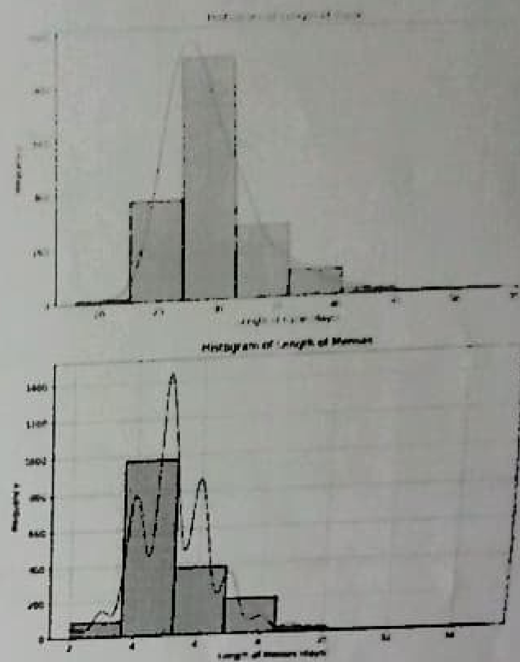

Fig. 2. Histograms showing the distribution of cycle lengths (top) and menses lengths (bottom) among users. These visualizations help identify common ranges and any outliers in the dataset.

occur on a single day, typically around 14 days before the onset of the next menstrual period. [10]

- **Ovulation day:** Estimated to occur 14 days before the next expected period
- **Fertile Window:** To increase accuracy, a window was 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 this study, a 6-day fertile window was 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) can protect eggs from oxidative stress, optimize hormonal balance, and improve egg quality during this critical time, as highlighted in recent studies [11][12]. 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 count for the 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, as illustrated in the fig. 3. Alternative methods were employed to address this imbalance, and their implementation will be discussed in the following section.

B. *Predictive Model Setup and Evaluation*

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 aforementioned parameters. XGBoost has several mechanisms to handle imbalanced data effectively, including the use of the scale pos weight parameter, which adjusts for class imbalance. Additionally, it implements L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting, ensuring that the model does not overly favor the majority class. 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 the fig. 3, a bias remains between the classes in the dataset, particularly concerning the menstrual and ovulation phase. To address the data imbalance present, SMOTE (Synthetic Minority Over-sampling Technique) was 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.

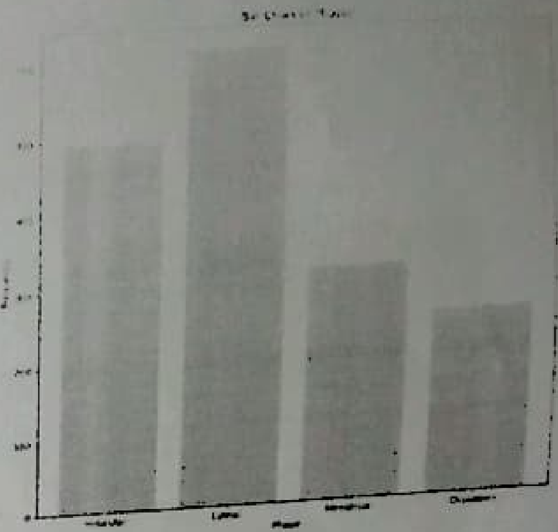

Fig. 3. Bar plot showing the distribution of phases calculated using the Ogino method.

- **Hyperparameter Tuning:**
  The performance metrics post-SMOTE indicated that precision, recall, and F1 scores were high for the follicular, luteal, and menstrual phases, but the ovulation phase still exhibited comparatively lower performance. Despite the dataset enrichment through SMOTE, which improved class balance, the unique characteristics of ovulation, such as its smaller data size and complex hormonal patterns, presented challenges. Grid search played a critical role in optimizing the model's hyperparameters, ensuring that it adapted to the diverse features of each menstrual phase. While high performance was achieved for most phases, further fine-tuning, especially for the ovulation phase, led to more accurate predictions and better overall performance.

C. *Personalized Recommendations* 5

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-veg, veg, or vegan). For each phase, a list of recommended foods, foods to avoid, and hydration tips was given.

IV. RESULTS & Analysis

This section presents the findings to predict menstrual cycle phases and provide personalized dietary recommendations based on those predictions.

A. *Key Results*

1) **Model Performance:** The application of SMOTE balanced the dataset, allowing the model to learn from a more comprehensive range of ovulation data, while hyperparameter tuning optimized its predictive power specifically for ovulation. This led to improved precision and recall, achieving a noteworthy F1-score.

TABLE II

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

Final XGBoost Test Accuracy: 0.9939

2) **K-Fold Stratified Cross-Validation with SMOTE:** The performance of the XGBoost classifier was evaluated using a 5-fold stratified cross-validation approach. Table III shows the results. The average accuracy obtained from this evaluation was **0.9862**, indicating that the model performed exceptionally well across different folds.

TABLE III

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

3) **Visualization of Complete Cycle:** The GUI fig. 5 visualizes the complete menstrual cycle, illustrating the predicted phases over time, for the input values cycle length = 34, menses length = 5, mean menses score = 2.

B. *GUI Component*

Additionally, a Graphical User Interface (GUI) was developed to facilitate user interaction with the model. The GUI allows users to input their menstrual cycle details and receive personalized dietary recommendations based on the predicted phases. Below are the key GUI windows that can be included as images:

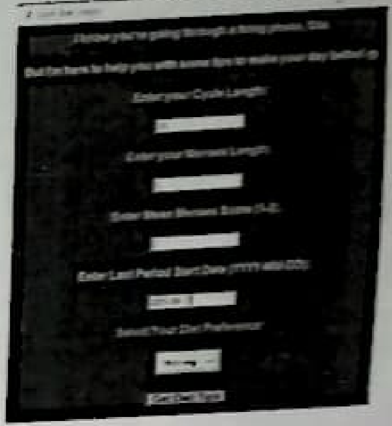

Fig. 4. An image of the GUI input window where users enter their cycle details.

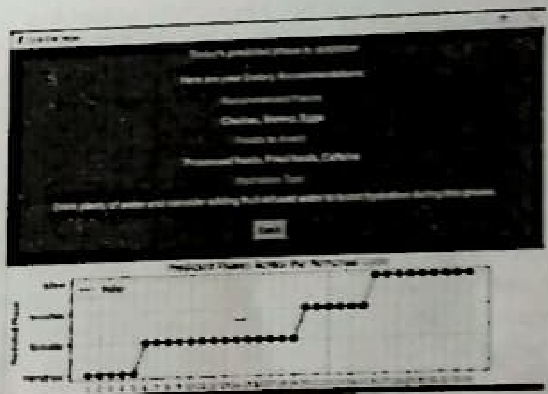

Fig. 5. An image of the results window displaying the predicted phase as ovulation and it's respective dietary recommendations. (Assuming today's date as 2024/10/08 and last period date as 2024/09/17)

V. DISCUSSION AND ANALYSIS

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

A. *Significance of Results*

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 hyper-parameters 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.

B. 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 generalized or based on symptoms 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.

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.

C. Limitations

While the model performed well, certain limitations were noted. The Ovulation phase posed some challenges due to its inherent variability, which affected precision and recall. Additionally, the use of the Ogino method for phase calculation, while functional for the dataset, is considered less accurate compared to newer methods. Integrating more advanced phase prediction techniques could improve precision and reliability.

The dataset, while diverse, lacked substantial irregularities, particularly those associated with PCOS and PCOD. This limits the model's generalizability. Furthermore, reliance on manually inputted data restricts scalability, as it could benefit from the integration of real-time data from wearable devices.

In summary, the model demonstrates strong performance across various phases of the menstrual cycle and shows promise for real-world personalized healthcare applications.

VI. Conclusion

This work presents a comprehensive approach to addressing challenges like that of scarceness of personalized dietary recommendations for the various menstrual phases through a machine learning techniques. The model used is XGBoost and it has been chosen because of the robust performance it displayed as compared to other models. The techniques SMOTE and hyperparameter tuning used with the model gave a significant accuracy of 99.39%. The high performance achieved demonstrates the model's capability to predict all four of the menstrual phases (follicular, ovulation, luteal and menstrual). The use of SMOTE to balance the dataset, lead to effective handling of class imbalances, which gave improved precision and recall, especially for the ovulation phase. The work has presented an easy-to-use and accessible tool to the users who can utilize it to receive nutritional recommendations along with their selected dietary preference in real-time. The accessibility aspect is heightened further through the provision of a simple but clear GUI. The model's ability to predict all four menstrual cycle phases, rather than focusing solely on ovulation or just the coming cycle lengths, defines the novel aspect of the work. The phase-specific approach to nutritional recommendations provides a tool that can help user's effectively plan out their nutrition for their current phase, an aspect that is often overlooked in menstrual health. Relevant future work can be done in the following fields:

- Prediction of early follicular, early luteal and late luteal phases to provide more narrowed down food recommendations.
- Recommendation of foods known to support fertility, if the user is trying to conceive.
- Taking into account allergies, dietary restrictions, nutritional deficiencies while recommending food.
- Prediction of phase accurately for irregular menstrual cycles.

In summary, this research provides a foundation for advancing personalized health recommendations, with scope for further refinement and integration into broader healthcare systems.

REFERENCES

[1] Nijboer, A.C.S., Sellitto, M., Ruitenberg, M.F.L., Kerkkanen, K.I.L., & Schomaker, J. (2023). "Food-related exploration across the menstrual cycle."

[2] Brown, N., Martin, D., Waldron, M., Bruinvels, G., Farrant, L., & Fairchild, R. (2023). "Nutritional practices to manage menstrual cycle related symptoms: a systematic review."

[3] Cleveland Clinic, "Nutrition and exercise throughout your menstrual cycle." Accessed: Oct. 07, 2024. [Online]. Available: https://health.clevelandclinic.org/nutrition-and-exercise-throughout-your-menstrual-cycle

[4] Trépanier L.C.M., Lamoureux E., Bjornson S.E., Mackie C., Alberts N.M., & Gagnon M.M. (2023). "Smartphone apps for menstrual pain and symptom management: A scoping review."

[5] Kriti N., Priyanshi G., Anushka P., & Rajiv K. (2023). "Menstrual Cycles Prediction using Artificial Neural Network."

[6] Odirichukwu J.C., Njoku O.A., Odirichukwu S.P.C., Ndigwe C., Nwachukwu D.C., Nworuka J.U., et al. (2023). "Improving Menstrual Cycle Prediction Accuracy using Advanced Machine Learning Model Methods."

[7] Logapriya E., & Surendran R. (2023). "Hybrid Recommendations System for Women Health Nutrition at Menstruation Cycle."

[8] NIH, "The impact of the menstrual cycle on women's health and nutrition," National Center for Biotechnology Information, Oct. 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10251302/

[9] Nikita B., "Menstrual Cycle Data," Kaggle.com, 2021. [Online]. Available: https://www.kaggle.com/datasets/nikitabisht/menstrual-cycle-data

[10] Ogino K. "Über den Konzeptionstermin des weibes und seine anwendung in der praxis" Zentralbl Gynakol

[11] Skoracka K., Alicja E. R., Anna M. R., Agnieszka D., & Iwona K. (2020). "Female Fertility and the Nutritional Approach: The Most Essential Aspects"

[12] Jurczewska J., & Szostak-Węgierek D. (2020). "The Influence of Diet on Ovulation Disorders in Women—A Narrative Review"

[13] Ankita Karna et al., "BeReady (Period Tracker & PCOS Di- agnosis)" Cycle and Symptom Tracker Mobile App 202[illegible] [illegible] period prediction, early PCOS de- Random Forest, 90.44% [illegible] tection

[14] K. Li, I. Urteaga, A. Shea, V. J. Vitzthum, C. H. Wiggins, N. Elhadad, "A Generative, Predictive Model for Menstrual Cycle Lengths," Year 2019. Mobile health data, self-tracking data

[15] Odirichukwu J. C. I., Njoku O. A., et al., "Improving Menstrual Cycle Prediction Accuracy using Advanced Machine Learning Model Methods." Menstrual Cycle Data 2023

[16] L. Symul, S. Holmes, "Labeling Self-Tracked Menstrual Health Records With Hidden Semi-Markov Models." 2019

[17] Prof. Dr. Rosana Rego, "Predictive Modeling of Menstrual Cycle Length: A Time Series Forecasting Approach." 2018

[18] Deokule, "Machine Learning Approaches for Menstrual Cycle Tracking."

[19] Heather Robinson,"Foods to eat for each stage of your menstrual cycle." London Clinic of Nutrition. [Online]. Available: https://londonclinicofnutrition.co.uk/nutrition-articles/foods-to-eat-for-each-stage-of-your-menstrual-cycle/