An Endocrinology Predictive Modeling Technology Based on Artificial Intelligence

Yizhen Fu[1,a,*]

[1]Shenzhen Middle School, Shenzhen, China

[a]13660862343@163.com

Keywords: Menstrual cycle prediction; Women's health; Hormone fluctuations; LSTM; LoRA fine-tuning

Abstract: The menstrual cycle is a key indicator of women's reproductive and endocrine health, influenced by hormones, lifestyle, and exercise. Accurate prediction is crucial for clinical decision-making and personal health management. Existing studies, however, often rely on small samples focus on single populations, or depend on noisy self-reported app data, lacking integration of multimodal features and medical interpretability. To address these gaps, we propose a multilevel AI framework that combines statistical analysis of exercise-hormone relationships with an LSTM-based time-series prediction model, incorporating physical, lifestyle, and hormona indicators. Furthermore, we fine-tune LLaMA3 using LoRA to build a domain-specific language model for menstrual health question answering and prediction interpretation. Experiments demonstrate a significant negative correlation between daily steps and progesterone, while estradiol shows no strong association. Our LSTM model outperforms ARIMA, GRU, and other baselines on MSE, MAE, RMSE, and R^2, while the fine-tuned LLaMA3 achieves superior PPL BLEU,ROUGE-L, and BERTScore compared to the original model. Overall, this study reveals exercise-hormone associations, proposes a robust cycle prediction method, and develops an intelligent health-oriented QA system, offering new tools for women's health management.

1. Introduction

The menstrual cycle is one of the most fundamental physiological processes in women of reproductive age, characterized by cyclic changes in ovarian and uterine function under the regulation of sex hormones such as estradiol (E2) and progesterone (P4). These hormones not only determine reproductive function but also exert systemic effects on metabolism, bone health, cardiovascular function, and psychological state. However, the cycle is highly dynamic and is influenced by both intrinsic and extrinsic factors, including age, body composition, psychological stress, and lifestyle. Among these, exercise, as a modifiable lifestyle variable, has been considered closely related to endocrine function and menstrual health. Previous studies have suggested that exercise may alter circulating E2 and P4 levels, thereby affecting cycle length ovulation, and menstrual regularity, although the mechanisms remain insufficiently understood. Moreover, irregular cycles and hormonal imbalances can significantly affect women's quality of life, fertility planning, and disease risk, underscoring the importance of predictive and intelligent tools for better monitoring and understanding of menstrual health.

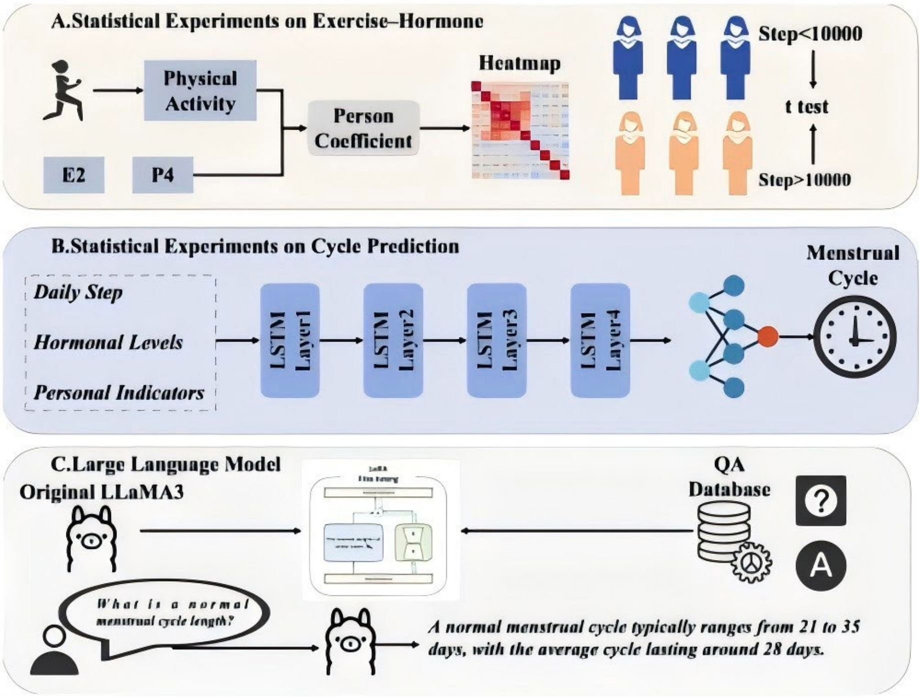
In recent years, studies have attempted to reveal the impact of exercise (e.g., step count and exercise intensity) on sex hormone fluctuations. A systematic review reported that moderate-tovigorous physical activity can lead to small but significant reductions in sex hormones, effects not entirely explained by weight loss, and highlighted the lack of mechanistic studies in adolescent and reproductive-age women [1]. Another study found that women with more than10.000 daily steps had significantly lower progesterone levels than less active women, while no significant effect was observed for estradiol; the study also pointed out the limitations of wearable devices such as Fitbit in accurately capturing physical activity [2]. These findings suggest a potential negative correlation between walking and progesterone, but existing studies have mostly focused on adults and face

DOI: 10.25236/iwmecs.2025.062

limitations in measurement tools and individual variability

Furthermore, menstrual cycle prediction models have mostly relied on self-tracking data from mobile apps. For example, Li et al. developed a generative model capable of online updating with cycle progression, distinguishing true physiological behaviors from interruptions in user tracking. However, this approach still depends on long-term adherence, and cannot address missing or low-quality data [3]. Urteaga et al. proposed a more flexible generalized Poisson model to handle irregularities in cycle data, yet it lacks integration of hormonal fluctuations lifestyle, and other multimodal features [4]. Finally, in the application of large language models (LLMs) or intelligent Q&A systems for menstrual health, there is still a lack of customized models grounded in clinical endocrinology, physiological cycles, and lifestyle behaviors. This gap limits medical interpretability and the ability to provide personalized, interactive health guidance.

To address these gaps, this study combines statistical analysis, machine learning, and large language models to build an AI-based framework for menstrual health prediction and interaction. Our framework is demonstrated in Fig. 1. First, independent-sample t-tests and Pearson correlation analysis are employed to quantitatively evaluate the relationships between physical activity (e.g., daily steps) and sex hormones (E2 and P4). Second, a Long Short-Term Memory (LSTM) network is constructed for time-series prediction to capture the dynamic patterns of cycle length and provide personalized predictions. Finally, Low-Rank Adaptation (LoRA) is applied to fine-tune the LLaMA language model, creating a domain-specific model for menstrual health scenarios, enabling both medically interpretable predictions and user-friendly interactive Q&A.



A normal menstrual cycle typically ranges from 21 to 35 days, with the average cycle lasting around 28 days.

Figure 1 AI-based framework for menstrual health prediction and interactior.

Our contributions are summarized as follows:

We quantitatively analyze the relationship between physical activity (e.g., daily step counts) and sex hormones (E2, P4) using Pearson correlation and independent-sample t -tests, providing new evidence on exercise-endocrine interactions. We develop an LSTM-based time-series prediction model that integrates physical, lifestyle and hormonal indicators, achieving more accurate and personalized menstrual cycle predictions than traditional statistical and baseline machine learning models. ·We fine-tune the LLaMA3 language model with Low-Rank Adaptation (LoRA)to build a domain-specific system for menstrual health, enabling medically interpretable predictions and interactive Q&A functionality for end users. To the best of our knowledge, this is the first integrated AI framework combining statistical analysis, sequence modeling, and customized large language models for menstrual health monitoring and guidance.

2. Method

2.1. Statistical Analysis of the Relationship between Exercise and Sex Hormones

In the statistical analysis stage, we first employ the Pearson correlation coefficient to quantify the linear association between physical activity level, represented by daily step counts, and two major sex hormones: estradiol (E2) and progesterone (P4). The Pearson correlation coefficient r^* ranges from $-1\leq r\leq 1$, where $r>0$ indicates a positive correlation, $r<0$ indicates a negative correlation, and $r=0$ suggests no significant linear relationship. The formula is given as:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

Here, X_i denotes the daily step count, y_i represents the corresponding hormone level, and $\bar{x}$ and y are the mean values of steps and hormone concentrations, respectively. By calculating both 7 and its associated P-value, we can determine whether a statistically significant lineat dependence exists between exercise and hormone levels.

To further examine differences in hormone levels across activity groups, participants are divided based on whether their average daily step count reaches 10,000 steps (≥10000 steps as the high-activity group, <10000 steps as the low-activity group). An independent-samples t -test is then conducted to compare progesterone and estradiol levels between the two groups. The null hypothesis H_0 states that there is no significant difference inmean hormone levels between the groups. If the calculated P -value is less than the significance threshold (commonly $\alpha=0.05$), the null hypothesis is rejected, indicating a statistically significant group difference.

The test statistic of the independent-samples t -test is defined as:

$$t=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \tag{2}$$

where X_1 and X_2 denote the sample means of the two groups, s_1^2 and s_2^2 are the sample variances, and n_1 and m_2 are the sample sizes. By comparing the computed t value with the critical value corresponding to the degrees of freedom and evaluating the associated P -value, we can determine whether significant differences in hormone concentrations exist between high and low-activity participants.

Through the combination of correlation analysis and group-level comparisons, this study not only explores potential linear relationships between exercise and endocrine indicators but also provides quantitative evidence of activity-related hormonal differences. These analyses contribute to a deeper understanding of how physical activity may influence female reproductive endocrinology and support the development of exercise-informed health management strategies.

2.2. Menstrual Cycle Prediction Using LSTM

The length and regularity of the menstrual cycle serve as critical indicators of women's reproductive health and endocrine balance [5]. Irregular cycles are often associated with polycystic ovary syndrome (PCOs), ovarian dysfunction, or psychological stress [6], which makes accurate prediction of the next cycle essential for clinical diagnosis, individual health management, and lifestyle planning. However, cycle variation is not solely determined by endocrine hormone regulation; it is also influenced by age, body mass index (BMI), lifestyle factors such as physical activity, and other extrinsic variables [2, 3]. These complex interactions give rise to highly dynamic temporal patterns that traditional linear statistical approaches are often insufficient to capture

To address these challenges, we adopt a Long Short-Term Memory (LSTM) network, which extends the recurrent neural network (RNN) framework by introducing gating mechanisms that enable the model to retain or discard information over long temporal spans. This makes LSTM particularly well-suited for modeling sequential data with nonlinear dependencies, such as menstrual cycles. The architecture of the LSTM network is shown in Figure 2.

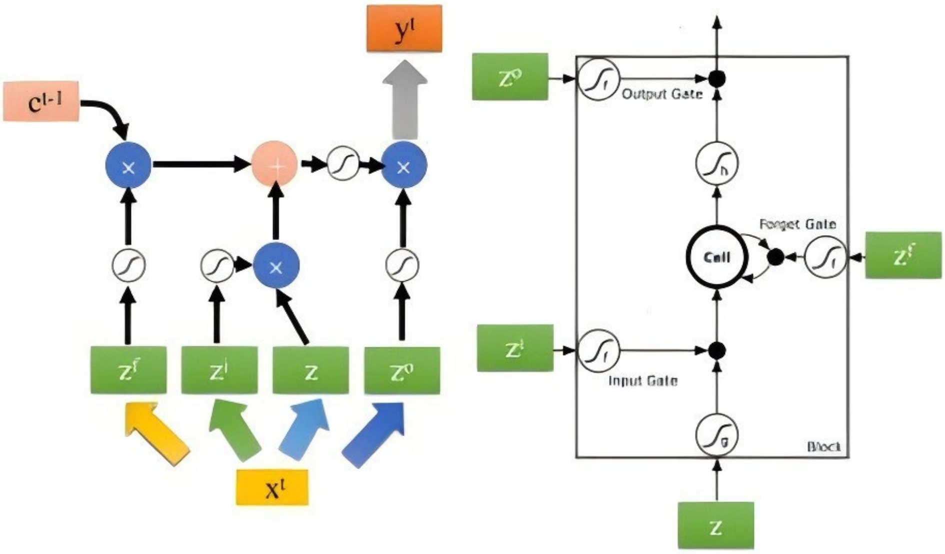


Figure 2 Illustration of the LSTM architecture, including input, forget, and output gates, which regulate information flow and enable long-term dependency modeling.

Let the input sequence be denoted as $\{X_t\}_{t=1}^{T}$, where $X_t \in \mathbb{R}^d$ represents the feature vector at time step t. These features may include age, BMI, daily step count, estradiol (E2) and progesterone (P4) levels, and past cycle lengths. The LSTM computation at each time step involves the following gating operations

$$
\begin{aligned}
f_t &= \sigma\left(W_f\,[h_{t-1}, X_t] + b_f\right),\\
i_t &= \sigma(W_i[h_{t-1}, X_t] + b_i),\\
\tilde{C}_t &= \tanh(W_C[h_{t-1}, X_t] + b_C)\\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t,\\
o_t &= \sigma\left(W_o\,[h_{t-1}, X_t] + b_o\right),\\
h_t &= o_t \odot \tanh(C_t).
\end{aligned}
\tag{3}
$$

Here, f_t is the forget gate that determines which past information should be discarded, i_t and $\tilde{C}_t$ together determine the contribution of new input to memory, C_t denotes the updated cell state serving as long-term memory, o_t controls the proportion of information passed to the hidden state, and h_t represents the hidden representation at time step t.

Based on the hidden state h_t , the model produces a prediction of the next cycle length $\hat{y}_{t+1}$

$$\hat{y}_{t+1} = W_y h_t + b_y. \tag{4}$$

The training objective is to minimize the mean squared error (MSE)between the predicted cycle length $\hat{y}_i$ and the ground truth y_i

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{5}$$

By optimizing this objective, the LSTM model is able to effectively integrate historical features of menstrual cycles with individual physiological indicators, thereby capturing temporal dependencies and providing personalized predictions. Compared with traditional approaches, this framework not only accounts for nonlinear sequential dynamics but also enables individualized forecasting aligned with women's health management.

2.3. Domain-Specific Language Modeling with LoRA-Fine-Tuned LLaMA3

With the rapid development of large language models (LLMs), their potential applications in medical question answering and health consultation have attracted increasing attention [7, 8]. Although general-purpose LLMs possess strong natural language processing capabilities, they often lack domain-specific expertise in reproductive endocrinology and menstrual health, making it difficult to provide reliable and personalized explanations. To address this limitation, our study introduces domain adaptation techniques in the third stage, aiming to construct an intelligent Q&A system for menstrual health. This system integrates clinical knowledge statistical predictions, and user input, thereby enabling interpretable and customized support.

As the backbone, we adopt LLaMA3 (Large Language Model Meta AI) [9], a Transformer based autoregressive language model. Its core architecture consists of multi-head self-attention (MHSA) and feed-forward networks (FFN), enabling effective long-range sequence modeling To further improve domain specialization with limited computational overhead, we employ Low-Rank Adaptation (LoRA)[10],which has proven to be an efficient fine-tuning strategy for adapting LLMs to specialized tasks. The structure of LLaMA3 is illustrated in Figure 3

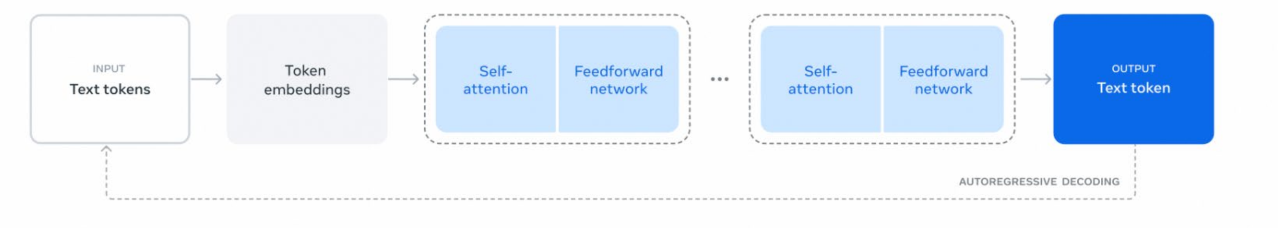


Figure 3 Basic architecture of LLaMA3, consisting of stacked Transformer layers with multi head self-attention and feed-forward components for autoregressive text generation.

The fundamental idea of LLaMA3 is to predict the next token y_t in a sequencegiven the preceding context $y_{<t}$, by minimizing the negative log-likelihood loss

$$\mathcal{L}_{\text{LLaMA}} = -\sum_{t=1}^{T} \log P(y_t \mid y_{<t}; \theta), \tag{6}$$

where θ represents the model parameters. Through stacked Transformer blocks, the model captures both syntactic and semantic structures of language sequences, enabling coherent text generation. In LLaMA3, Meta has further expanded the model’s context length and optimized positional encoding, thereby improving the ability to capture dependencies across long temporal spans, which is particularly important for modeling longitudinal health data and interactive dialogue.

To adapt LLaMA3 to the menstrual health domain, we employ Low-Rank Adaptatior (LoRA), a parameter-efficient fine-tuning (PEFT) method. Traditional fine-tuning requires updating and storing all parameters of the model, which is computationally expensive and memory-intensive. In contrast, LoRA introduces low-rank decomposition into the weigh update process. Specifically, instead of directly updating the full weight matrix W∈R^(d×k) LoRA constrains the update ΔW to a low-rank form:

$$W' = W + \Delta W, \quad \Delta W = AB \tag{7}$$

where A$\in \mathbb{R}^{d\times r}$ and B$\in \mathbb{R}^{r\times k}$, with r≪min(d,k) . During fine-tuning, the original weights W remain frozen, while only the low-rank matrices A and B are optimized. This significantly reduces the number of trainable parameters, allowing efficient adaptation without sacrificing the original linguistic capability of LLaMA3.

By integrating LoRA fine-tuning with domain-specific corpora, including medical literature, clinical guidelines, and structured menstrual health records, our approach enables LLaMA3 to provide not only natural language understanding but also interpretable reasoning grounded in reproductive endocrinology. This design allows the model to generate personalized responses to user queries, link statistical predictions of cycle outcomes with explanatory knowledge, and ultimately serve as a hybrid system for intelligent health management.

In summary, through parameter-efficient adaptation, our customized LLaMA3 model achieves a balance between general language generation ability and specialized domain knowledge. This ensures that the model can support interpretable Q&A on menstrual health, providing actionable insights for both individuals and healthcare practitioners.

3. Experiment

3.1. Experimental Setup and Datasets

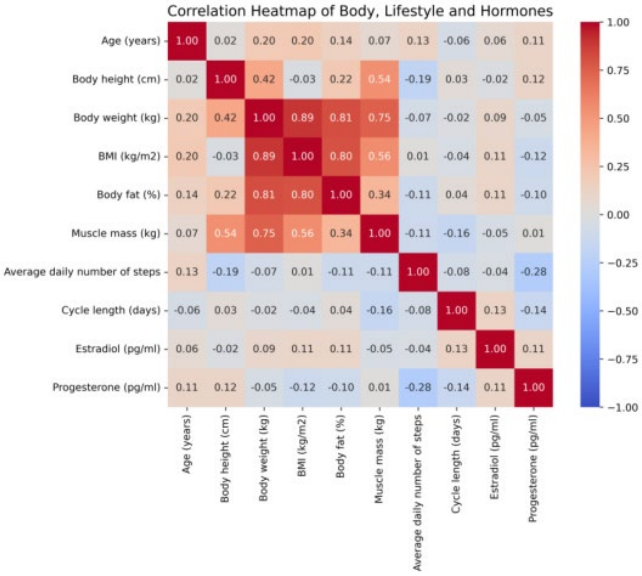
All experiments were conducted on a server with an NVIDIA A100 GPU(80 GB)running Ubuntu 20.04 LTS. The environment was based on Python 3.10 and PyTorch 2.1

Two datasets were used. The first, collected from 85 reproductive-age women, included anthropometric measures (age, height, weight, BMI, body fat, muscle mass), lifestyle indicators (average daily steps, wearable device tracking), menstrual features (cycle length, duration, luteal phase, bleeding volume, pain scores), and hormonal levels of estradiol (E2) and progesterone (P4). These data were analyzed with Pearson correlation and t -tests, and then structured as sequences for LSTM-based cycle prediction

The second dataset was a textual corpus for LoRA fine-tuning of LLaMA3, covering academic articles, clinical guidelines, patient-doctor dialogues, and community health discussions Additional lifestyle texts (diet, exercise, sleep, mental health) were included. After normalization and tokenization, this corpus enabled LLaMA3 to adapt to menstrual health tasks, combining predictive modeling with interpretable Q&A.

3.2. Statistical Experiments on Exercise-Hormone Relationships

To investigate the statistical associations between physical activity and hormonal fluctuations, we first analyzed the correlations between anthropometric variables, lifestyle indicators, and sex hormone levels. As shown in Figure 4, body height, weight, BMI, and body fat percent-age were found to be strongly and positively correlated with one another ($r>0.75$), indicating high internal consistency among these anthropometric indicators. In contrast, the correlation between average daily step count and progesterone (P4) levels was moderately negative ($r=-0.28$), suggesting that higher physical activity is associated with reduced progesterone. Meanwhile, the correlation between daily steps and estradiol (E2) levels was relatively weak ($r\approx-0.05$), implying a less consistent relationship. Furthermore, correlations between menstrual cycle characteristics (e.g., cycle length, luteal phase duration) and hormone levels were generally low, indicating that these features may be influenced by multiple interacting factors rather than a single predictor. Overall, the results suggest a relatively stable negative association between physical activity (steps) and progesterone, while the relationship between estradiol and cycle features appears less significant



	Age (years)	Body height (cm)	Body weight (kg)	BMI (kg/m2)	Body fat (%)	Muscle mass (kg)	Average daily number of steps	Cycle length (days)	Estradiol (pg/ml)	Progesterone (pg/ml)
Age (years)	1.00	0.02	0.20	0.20	0.14	0.07	0.13	-0.06	0.06	0.11
Body height (cm)	0.02	1.00	0.42	-0.03	0.22	0.54	-0.19	0.03	-0.02	0.12
Body weight (kg)	0.20	0.42	1.00	0.89	0.81	0.75	-0.07	-0.02	0.09	-0.05
BMI (kg/m2)	0.20	-0.03	0.89	1.00	0.80	0.56	0.01	-0.04	0.11	-0.12
Body fat (%)	0.14	0.22	0.81	0.80	1.00	0.34	-0.11	0.04	0.11	-0.10
Muscle mass (kg)	0.07	0.54	0.75	0.56	0.34	1.00	-0.11	-0.16	-0.05	0.01
Average daily number of steps	0.13	-0.19	-0.07	0.01	-0.11	-0.11	1.00	-0.08	-0.04	-0.28
Cycle length (days)	-0.06	0.03	-0.02	-0.04	0.04	-0.16	-0.08	1.00	0.13	-0.14
Estradiol (pg/ml)	0.06	-0.02	0.09	0.11	0.11	-0.05	-0.04	0.13	1.00	0.11
Progesterone (pg/ml)	0.11	0.12	-0.05	-0.12	-0.10	0.01	-0.28	-0.14	0.11	1.00

Figure 4 Correlation heatmap of anthropometric indicators, lifestyle factors, and sex hormones. Daily step count shows a negative correlation with progesterone (P4), while estradiol (E2) correlations remain weak.

We further compared hormone levels between two activity groups: women with high activity (≥10000 steps/day) and those with low activity <10000 steps/day). As illustrated in Figure 5, the progesterone distribution of the high-activity group was shifted downward, with the median value lower than that of the low-activity group, supporting the negative correlation identified in the Pearson analysis. This indicates that higher physical activity may be linked to reduced progesterone levels. In contrast, estradiol distributions between the two groups showed no significant differences, and t -test results confirmed the absence of meaningful group-level variation.

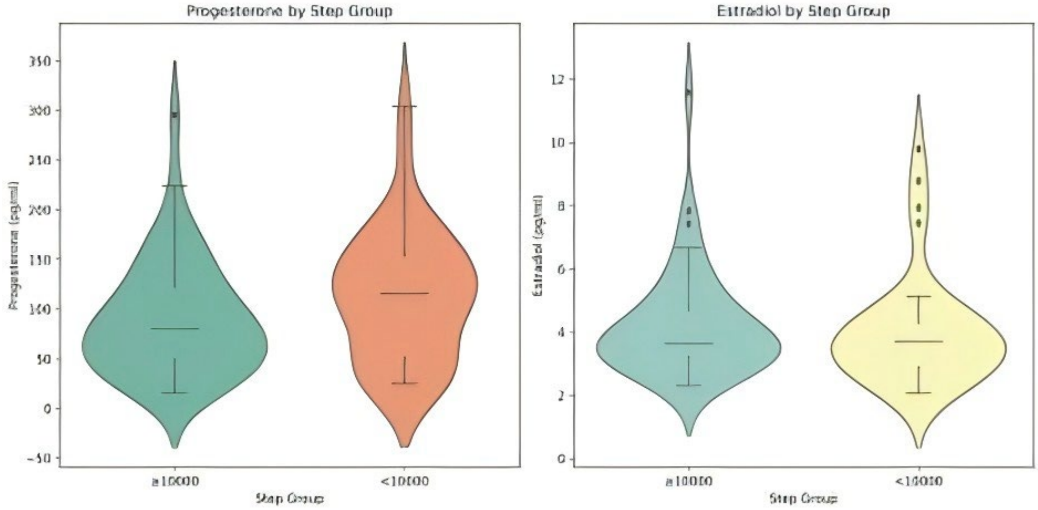


Figure 5 Comparison of progesterone (P4) and estradiol (E2) distributions between high activity (≧ 10000 steps) and low-activity (< 10000 steps) groups. Significant group-level differences were observed for progesterone but not for estradiol.

Taken together, these results highlight a consistent negative association between step count and progesterone, whereas estradiol levels appear unaffected by physical activity. The findings support the hypothesis that progesterone fluctuations may be more sensitive to lifestyle factors such as exercise, while estradiol regulation follows a more complex mechanism influenced by additional physiological variables.

3.3. Statistical Experiments on Cycle Prediction

In this study, we formulated menstrual cycle length as the target variable for time-series prediction and employed a Long Short-Term Memory (LSTM) model to capture sequential dependencies. The model inputs consisted of multimodal features including anthropometric indicators (e.g., age, BMI), lifestyle variables (e.g., daily step counts), and hormonal levels (estradiol and progesterone).The output was the predicted length of the next menstrual cycle.

To rigorously evaluate predictive performance, several regression metrics were adopted. including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The MSE loss function used for model optimization is defined as

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{8}$$

where y_i denotes the observed cycle length, $\hat{y}_i$ the predicted value, and N the number of samples. Minimizing this loss enables the model to reduce average squared prediction errors over the sequence. During training, the loss consistently decreased with increasing epochs and converged after approximately 50 iterations, indicating effective learning of sequential patterns. The resulting model demonstrated both stability and convergence, consistent with the statistical findings reported in Section 3.1

For comparative analysis, we benchmarked the LSTM model against several baselines: the classical autoregressive ARIMA model, the gated recurrent unit (GRU)model, and a naive historical average predictor. As summarized in Table 1, the LSTM consistently outperformed.

All baselines across evaluation metrics. In particular, it achieved the lowest RMSE and MAE while yielding the highest R^2 score, reflecting superior predictive accuracy and generalization. These results confirm that LSTM is more effective in capturing the nonlinear and temporal dependencies inherent in menstrual cycle data, compared to both traditional statistical methods and simpler recurrent models.

Table 1 Prediction results comparing different models on menstrual cycle length forecasting LSTM achieves the best performance across all evaluation metrics.

Model	MSE	MAE	RMSE	R^2
Average	42.15	5.73	6.49	0.21
ARIMA	35.84	5.12	5.98	0.34
GRU	28.41	4.36	5.33	0.51
LSTM	22.97	3.92	4.79	0.62

In summary, the LSTM model not only reduces prediction error but also provides more personalized forecasts of menstrual cycle length compared to ARIMA and GRU. These findings suggest that deep learning methods, particularly LSTM, offer a promising approach for individualized reproductive health management

3.4. LoRA-Fine-Tuned LLaMA3 Question Answering Experiments

In the final stage of our experiments, we further evaluated the performance of the LLaMA3 model fine-tuned with Low-Rank Adaptation (LoRA) on domain-specific question answering tasks in menstrual health. The objective was to examine whether parameter-eficient fine-tuning could enhance the model's ability to generate accurate and clinically meaningful responses. For this purpose, we constructed a domain-specific corpus covering clinical guidelines, patient-doctor dialogues, and community health discussions, and fine-tuned the LLaMA3 backbone using the LoRA approach

During fine-tuning, the LoRA rank was set to r=16 , the scaling factor α=32, and the number of training epochs was three. Optimization was performed with AdamW using a learning rate of 2×10^{-4} .This efficient training setup ensured that only a small number of additional parameters were introduced, while preserving most of the pretrained weights of LLaMA3.

To evaluate the model, we adopted multiple automatic metrics. First, perplexity (PPL) was used to measure the model's generative fluency, with lower values indicating better language modeling. In addition, BLEU and ROUGE-L were applied to assess lexical and structural overlap between model-generated responses and reference answers. Finally, semantic similarity was measured using BERTScore, which captures embedding-level alignment between generated and reference text. The perplexity loss function is defined as:

$$\mathcal{L}_{\mathrm{PPL}} = -\frac{1}{T}\sum_{t=1}^{T} \log P(y_t \mid y_{<t}; \theta), \tag{9}$$

where y_t denotes the ground-truth token at position t, $y_{<t}$ the preceding context, and θ the model parameters.

Experimental results are summarized in Table 2. Compared with the original pretrained LLaMA3, the LoRA-fine-tuned model achieved substantial improvements across all evaluation metrics. Specifically, PPL decreased by nearly 40% , while BLEU and ROUGE-L scores improved by 7-10 absolute points. BERTScore also showed a clear gain, indicating that the model generated responses that were semantically closer to expert references.

Table 2 Performance of original LLaMA3 versus LoRA-fine-tuned LLaMA3 on domain specific QA tasks. LoRA fine-tuning leads to significant improvements in fluency and semantic alignment.

Model	PPL↓	BLEU↑	ROUGE-L ↑	BERTScore ↑
Original LLaMA3	18.6	24.7	31.5	0.812
LLaMA3 + LoRA	11.2	32.5	41.8	0.871

4. Conclusion

In this study, we proposed an integrated AI framework for menstrual health monitoring that combines statistical analysis, time-series modeling, and domain-specific language modelling.

Through correlation analysis and independent-samples t-tests, we revealed a stable negative association between physical activity and progesterone levels, while estradiol correlations remained weak. An LSTM-based prediction model was then developed to forecast menstrual cycle length, achieving superior accuracy compared with ARIMA and GRU baselines across MSE, MAE, RMSE, and R^2 .Finally, we fine-tuned LLaMA3 with LoRA to construct a customized question answering system, which significantly reduced perplexity and improved BLEU, ROUGE-L, and BERTScore, thereby enhancing interpretability and user interaction Overall, our framework provides both quantitative predictions and personalized health guidance, offering a promising direction for intelligent reproductive health management.

References

[1] A. Bagga and B. Kulshreshtha, “Physical activity and reproductive health: a systematic review, Reproductive Health, vol. 17, no. 1, p. 158, 2020

[2] G. Mishra, K. Chan, A. K. Wesselink, K. J. Rothman, and L. A. Wise, “Feasibility of using afitbit device to monitor physical activity and its relationship with reproductive hormones in women,” BMC Women’s Health, vol.20, no.1,p. 215,2020

[3] L. Li, A. Haimovich, N. Vashisht, and H. M. Krumholz,“Predicting menstrual cycles with limited data,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 695-703.

[4] Urteaga, A. Mabe, and N. Elhadad, “Forecasting menstrual cycles with a generalized poisson model,” npj Digital Medicine, vol. 4, no. 1, p. 5, 2021

[5] J. R. Bull, S. P. Rowland, E. B. Scherwitzl, K. Scherwitzl, K. G. Danielsson, and J. Harper, Menstrual cycle characteristics: associations with fertility and reproductive health outcomes,’ Obstetrics & Gynecology, vol. 95, no. 6, pp. 1070-1076, 2000

[6] A. Gambineri, L. Patton, P. Altieri, U. Pagotto, C. Pizzi, L. Manzoli, and R. Pasquali, “Polycystic ovary syndrome: a common endocrine disorder with important health conse-. quences,” Human Reproduction Update, vol. 18, no. 4, pp. 485-503, 2012.

[7] A. Thirunavukarasu, D. S. Ting, P. Elangovan, L. Gutierrez, S. K. Tan, A. Alavi, J. Y. Lee T.Y. Wong, and H. Lin,“Large language models in medicine,” Nature Medicine, vol. 29 no. 8, pp. 1930-1940, 2023

[8] K. Singhal, S. Azizi, T. Tu, and et al., “Towards expert-level medical question answering with large language models,’ Nature, vol. 620, no. 7972, pp. 182-188, 2023

[9] H. Touvron, T. Lavril, G. Izacard, and et al.,“Llama: Open and efficient foundation language models,” arXiv preprint arXiv:2302.13971, 2023.

[10] E. J. Hu, Y. Shen, P. Wallis,Z. Allen-Zhu, Y. Li, S. Wang, L.Wang, and W. Chen,“Lora Low-rank adaptation of large language models,” in International Conference on Learning Representations (ICLR), 2022.