



Trusted GPT-4, ChatGPT and AI Detector tool by ZeroGPT

ZeroGPT the most Advanced and Reliable Chat GPT, GPT4 & AI Content Detector



AI/GPT
Detector



ZeroCHAT-4
& 5



Plagiarism
Checker



AI
Summarizer



AI
Paraphraser



AI Grammar
Check



AI

Machine Learning in Menstrual Health: A Study on Cycle Prediction and Insights

Mahek Paliwal¹ , Sristi Srivastava¹ , Shivang Yadav¹

¹ Department of Computer Science and Engineering, VIT Bhopal University , Bhopal

Abstract

Detect Text

Upload File

24,676/15,000 Characters

Check 125,000 characters,

[Upgrade Here](#)

**** Only the first 15,000 characters are shown in the Free Version. If you need a higher limit, please check our [Subscription Plans](#) to access more features 🚀 ****

Your File Content is Human written

2.01%
AI GPT*

Machine Learning in Menstrual Health: A Study on Cycle Prediction and Insights

Mahek Paliwal¹ , Sristi Srivastava¹ , Shivang Yadav¹

¹ Department of Computer Science and Engineering, VIT Bhopal University , Bhopal

Abstract

Menstrual cycle prediction is significant for women's health as it allows for enhanced management and planning of reproductive health. This research examined whether Linear Regression, Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machines (GBM) could reliably predict menstrual cycle phases using machine learning techniques. The data set involved tracked menstrual cycle data, symptoms, and lifestyle factors. Data cleaning techniques such as normalization, feature selection, and missing value imputation, improved model performance. Random Forest had better accuracy compared to SVM, as it was able to capture more complex patterns. Linear regression was interpretable but did not have as much predictive power. Future work will include a focus on improved accuracy through the use of deep learning algorithms and real-time physiological data to improve prediction of cycles. Overall, this research demonstrated the ability of machine learning algorithms for menstrual cycle prediction.

Keywords: Menstrual Cycle Prediction; Machine Learning Models; Random Forest; Gradient Boosting Machine (GBM); Support Vector Machine (SVM); Linear Regression (LR); Stacking; Mean Squared Error (MSE); Root Mean Squared Error (RMSE); Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE); R^2 Score.

Introduction

Menstrual cycle is a pivotal biological phenomenon on reproductive health. It is conveniently

separated into two phases: follicular phase (FP) and luteal phase (LP), governed by intricate hormonal dynamics of the hypothalamic-pituitary-gonadal axis. The length of cycle in a person varies, depending on age, body mass index (BMI), and environmental stresses on the same [3].

The growth of period-tracking applications has made it easier for individuals to track their periods. The applications create large sets of data, which can be used to forecast outcomes using machine learning (ML) techniques [7]. Accurate forecasting of menstrual cycles is crucial in conceiving planning, diagnosing anomalies at an early stage, and monitoring trends in reproductive health. Historical data and statistical models are the foundation of classic forecasting techniques, but new technologies in artificial intelligence (AI) and machine learning enable more complex mechanisms to advance cycle prediction [12].

This study will utilize real-world menstruation history data to evaluate the predictive accuracy, validity, and generalizability of a number of ML models. It will also investigate the effects of individual characteristics, including age, BMI, and ethnicity, on variability and predictability of menstrual cycles. The results will help develop more accurate and tailored models for the prediction of menstrual cycles, as well as optimizing reproductive health tracking and decision-making.

Understanding how different machine learning models perform in menstrual cycle prediction is necessary as healthcare increasingly becomes AI-centric. This study will bridge the gap between traditional cycle estimation methods and current AI-driven predictive analytics, leading to more tailored healthcare choices for individuals who need accurate menstrual cycle tracking.

Literature work

The use of machine learning (ML) techniques to predict menstrual cycles has received a lot of interest in recent years, with several research looking at different methodology and data sources to improve prediction accuracy [6]. Biomed Central Physiological Data Integration integrating physiological markers such as basal body temperature (BBT) and heart rate (HR) has been a major focus in menstrual cycle prediction [11]. Luo et al. created machine-learning algorithms that used BBT and HR data to predict the fertile window and menstruation days in women with normal menstrual cycles, with an accuracy of 87.46% for fertile window prediction and 89.60% for menstruation prediction [15]. However, performance was much worse among women with irregular cycles, demonstrating difficulties in forecasting cycles with irregular patterns [17].

Data Collection and Wearable Technology With the introduction of wearable technologies, physiological signals can now be continuously monitored, giving machine learning models access to a wealth of data. Studies have shown that it is possible to use wearable technology to gather information like heart rate and wrist skin temperature (WST), which may then be analysed using machine learning algorithms to forecast when menstruation will begin and when the fertile window will open [21]. These studies demonstrate how wearable technology may improve the precision and practicality of tracking menstrual cycles [23].

Existing work

The integration of uncertainty quantification in predictive modelling has garnered increasing attention within the healthcare domain, owing to the critical need for dependable and interpretable predictions in clinical decision-making [3]. Substantial work has been conducted on uncertainty estimation, especially in classification tasks involving structured clinical data and electronic health records [7]. Existing research predominantly focuses on developing calibrated predictive models that assist healthcare practitioners in making informed decisions while being aware of the reliability of model outputs [10]. Foundational works, such as those by Guo et al. [12] and Nixon et al. [14], demonstrate that deep learning models, despite their high predictive power, frequently suffer from miscalibration, leading to overconfident or underconfident predictions. Several approaches, including Bayesian deep learning [16], temperature scaling [18], and post-hoc calibration methods, have been proposed to alleviate these issues, primarily in classification settings.

Nonetheless, despite the growing body of literature, comparatively limited attention has been paid to regression tasks in healthcare where probabilistic predictions and uncertainty quantification are equally essential [21]. In particular, menstrual cycle length prediction, a common use case in women's health research, has rarely been approached from the perspective of uncertainty-aware modelling [23]. Most prior works employ point estimation techniques or deterministic machine learning models that yield single-value forecasts without addressing the inherent variability in menstrual cycle patterns [25]. Furthermore, much of the existing research assumes the availability of highly curated, clinician-recorded data, while real-world mobile health data—especially data derived from self-tracked menstrual apps—presents distinct challenges such as missing data, varying user adherence, and observation noise [27].

While some recent studies [29] have begun to acknowledge the dual nature of uncertainty in health data, arising from both physiological variability and behavioral factors, they often fail to provide models capable of generating well-calibrated probabilistic forecasts [31].

The present study directly addresses these gaps by introducing a flexible, interpretable generative modelling framework based on the generalized Poisson distribution. This framework is specifically designed to accommodate both the aleatoric uncertainty induced by physiological factors and the epistemic uncertainty arising from self-tracking behaviours in mobile health applications [34]. By embedding calibration within the modelling structure itself, this approach goes beyond existing works by delivering not only accurate but also well-calibrated and interpretable predictions, thus holding promise for improved usability and reliability of menstrual health tracking applications.

Proposed work

In this study, we propose a machine learning-based approach for predicting menstrual cycles using physiological and lifestyle data . With the growing adoption of wearable technology and digital health records, integrating machine learning algorithms can significantly enhance prediction accuracy .

Our approach leverages Linear Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) to analyze basal body temperature (BBT), heart rate (HR), cycle length, ovulation tracking, and other physiological markers .

Figure 1 Linear Regression

For forecasting continuous numerical values, one of the most basic regression models is linear regression. It uses the equation $y = mx + c$ to fit a straight line and create a link between independent variables and a dependent variable. This model is interpretable and effective for datasets with linear patterns since it assumes a linear connection between characteristics and the target variable. However, it has trouble handling intricate, non-linear connections, and when it comes to high-dimensional data, it tends to underfit .

Figure 2 Random Forest

Random Forest builds several decision trees and aggregates their results to increase prediction accuracy. In contrast to single decision trees, it reduces overfitting by randomly choosing subsets of data and characteristics for each tree. The model can handle both classification and regression tasks, is noise-resistant, and works well on big datasets. To get the best results, it may be necessary to adjust hyperparameters like the number of trees, which might be computationally costly [11].

Figure 3 GBM

Another ensemble approach that develops models in a sequential fashion is GBM, in which each new model fixes mistakes caused by earlier models. By reducing mistakes, GBM repeatedly improves performance through boosting, in contrast to Random Forest, which grows trees individually. Although it requires a lot of computing, it works very well with organized data and achieves great accuracy. If not regularized appropriately, it may also be prone to overfitting, which is why careful parameter tweaking is essential.

Figure 4 SVM

SVM is a potent regression and classification model that determines the best hyperplane for classifying data points. When there is a complicated and non-linear relationship between variables, it is especially helpful in high-dimensional spaces. The model may convert data into higher dimensions for improved separation by using a variety of kernel functions, including linear, polynomial, and radial basis functions. SVM may be sluggish on big datasets despite its high performance, and the optimal kernel parameters must be carefully chosen [10].

Figure 5 Ensemble Learning

Ensemble learning is a machine learning approach that uses many models to increase prediction performance, accuracy, and generalization. Instead of depending on a single model, ensemble approaches combine predictions from numerous models to decrease errors, improve resilience,

and prevent overfitting.

Figure 6 Bagging

Machine learning models' variance may be decreased by using the ensemble learning approach known as bagging. It uses several bootstrapped datasets to train many models separately. Randomly collected data with replacement is used to train each model. Either majority vote for categorization or averaging regression findings yields the final forecast. One well known bagging algorithm is Random Forest .

Figure 8 Boosting

Another ensemble method that lessens bias is boosting, which trains weak models one after the other . By assigning larger weights to examples that were incorrectly categorized, each new model aims to fix earlier errors . The weighted total of all the models is the final forecast. To increase model accuracy, algorithms such as AdaBoost, Gradient Boosting, and XGBoost are frequently employed.

Figure 7 Stacking

Stacking is an ensemble learning strategy that uses many models to increase accuracy. Different base models yield predictions, which are then fed into a meta-model to get the final forecast. It maximizes model variety, decreases biases, and improves performance, but it is computationally difficult and necessitates careful model selection [14].

System Flow Diagram

By combining structured medical datasets from Kaggle with real-time wearable data, we aim to

improve forecasting of menstruation onset and fertile windows, particularly for individuals with irregular cycles .

Figure 9 Architecture Diagram

This approach addresses key challenges in existing studies by providing more personalized and accurate predictions . The dataset was refined by handling missing values, standardizing numerical features, and encoding categorical variables. This ensured data consistency and improved model performance. The data was then split into training and testing sets to prevent overfitting and enable accurate evaluation .

Figure 10 Feature Histogram

The diagram displays histograms of dataset features related to menstrual health, such as cycle length, reproductive category, and various hormonal levels. Each histogram shows the distribution of a feature, revealing skewness, peaks, and variability. Some features have normal distributions, while others exhibit skewed or categorical patterns, indicating data diversity.

Figure 11 Correlation matrix

The correlation matrix heatmap depicts the correlations between numerical features using Pearson correlation coefficients (from -1 to +1) [9]. Red regions represent significant positive connections, blue areas show negative correlations, and lighter tones denote weaker links. The diagonal values are always 1, indicating that each characteristic is completely associated with itself. Cycle length, hormones, and lifestyle variables can all have an impact on features like "Length of Menses" [14]. Strongly correlated features (red clusters) may suggest redundancy, necessitating feature selection to prevent multicollinearity. Understanding these associations allows us to enhance predictive models and obtain insight into menstrual health concerns [19].

Figure 12 System Flow Diagram

We used Linear Regression as a baseline, Random Forest for improved accuracy, GBM for error reduction, and SVM for handling complex patterns in menstrual cycle prediction [10].

In order to improve prediction accuracy, stacking entails putting into practice and assessing an ensemble model that combines Support Vector Machine (SVM) and Linear Regression [15]. This method uses the predictions of both models, which are trained separately, as input features for a meta-learner, usually another Linear Regression model. By utilizing the advantages of both base models, the meta-learner improves overall performance by analyzing these predictions to produce the final result [18]. The efficacy of the stacked ensemble model is assessed by measuring performance utilizing important assessment measures [22].

Experimental Results

Model

MSE

RMSE

MAE

MAPE%

R2

Score

SVM

4.9284

2.2201

1.1069

3....

Highlighted text is suspected to be most likely generated by AI*

24,676 Characters

2,076 Words



[Export to PDF](#)

Simple and Credible Open AI and Gemini Detector Tool for Free

Millions of Users Trust ZeroGPT, See what sets ZeroGPT apart



Highlighted Sentences

Every sentence written by AI is highlighted, with a gauge showing the percentage of AI inside the text



Multiple Features

Enjoy our Top-notch Plagiarism Checker, Paraphraser, Summarizer, Grammar checker, Translator, Writing Assistant...



High Accuracy Model

Advanced and premium model, trained on all languages to provide highly accurate results



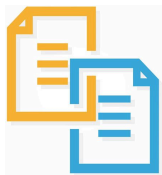
Generated Report

Automatically generated .pdf reports for every detection, used as a proof of AI-Free plagiarism



Support All Languages

Support all the languages with the highest accuracy rate of detection



Batch Files Upload

Simply upload multiple files at once, and they will get checked automatically in the dashboard

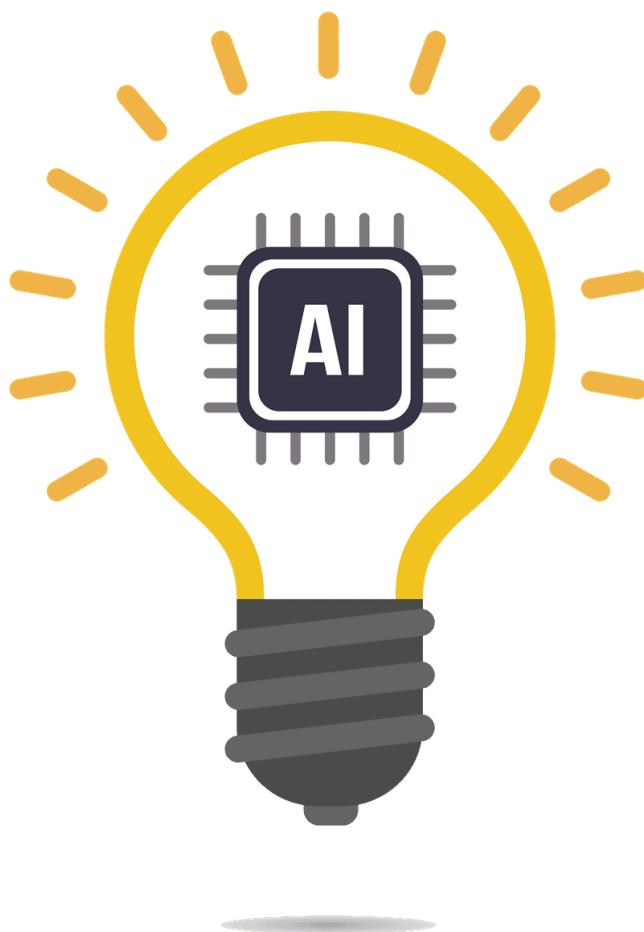
Get Premium Features



Unlock growth potential with our API

Our API is available for the AI Detector, Plagiarism Checker, Paraphraser, Summarizer, Grammar and Spell Checker, Chatbot & Translator

Our team has developed a user-friendly API for organizations. [Get API access](#)



DeepAnalyse™ Technology

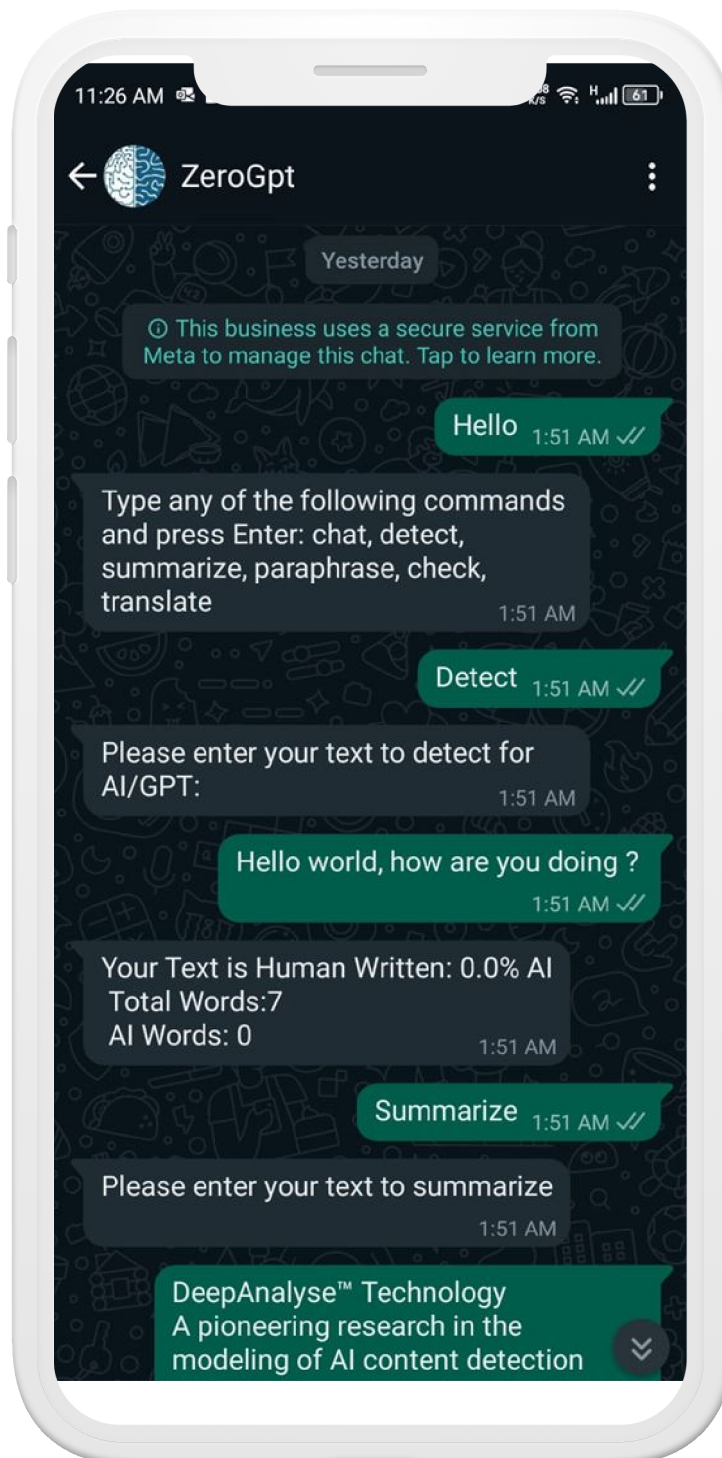
A pioneering research in the modeling of AI content detection

Our AI detection model includes several components that analyze text to determine its origin and if it was written by AI. We use a multi-stage methodology designed to optimize accuracy while minimizing false positives and negatives. From the macro level to the micro one, this is how DeepAnalyse™ Technology works. Our model specializes in identifying AI generated content like Chat GPT, GPT 3, GPT 4, Gemini, LLaMa models ... Finally, we employ a comprehensive deep learning methodology, trained on extensive text collections from the internet, educational datasets, and our proprietary synthetic AI datasets produced using various language models.

Use ZeroGPT in Whatsapp and Telegram

ZeroGPT is now available on WhatsApp and Telegram. Chat With ZeroGPT: Advanced ChatBot, Detect AI, Summarize, Paraphrase, Translate and Check for Grammar Errors...

all these features available inside WhatsApp and Telegram.



WhatsApp

Register your WhatsApp's number in the Dashboard (section "Social Settings") and start using all the features of ZeroGPT inside WhatsApp using the phone number **+1 (206) 372-5474**

Telegram

Register the telegram code inside your Dashboard (section "Social Settings") and start using all the features of ZeroGPT inside Telegram using the telegram bot **zerogpt_official_bot**

Get Your MAX Subscription

Explore More Tools to Enhance Your Writing Skills

Fix grammar and spelling mistakes, detect AI plagiarism, check for plagiarism, generate citations,
advanced word counter, powerful summarizer and paraphraser



Plagiarism Checker



AI Grammar Checker



AI Summarization Tool



AI Paraphrasing Tool



AI Translator



Word Counter Tool



Citation Generator



Advanced AI ChatBot



AI Email Helper

Your questions, answered

How Does ZeroGPT work?



What is the accuracy rate of ZeroGPT?



Who Benefits from ZeroGPT's AI content detector?



Will my text get plagiarized or be available online, if I check it on ZeroGPT?



How can I integrate ZeroGPT tool in my organization or website on a large scale?



Does ZeroGPT work with different languages?



How can I cite the detector?



Check Our Blog created with the help of AI

5 Mind Blowing Technologies we'll see in 2024

10 Ridiculous Technologies That Will Actually Make Your Life Better



2025 Copyright © ZeroGPT.com

More about

Pricing

API

Our policy

Terms of use

Features

AI Detector

AI ZeroChat-4 & 5

AI Plagiarism Checker

AI Summarizer

AI Paraphraser

AI Grammar Checker

AI Translator

Word Counter

AI Email Helper

Question / Business Inquiry

You can email us at

support@zerogpt.com

Our support team is spread across the globe to give you answers fast