Application of machine learning in menstrual cycle assessment based on data from iYoni mobile app

Dominik Kossiński Katarzyna Goch Krzysztof Łukaszuk [1]

[1]Medical University of Gdansk

Introdution

According to WHO, infertility affects $10-16\%$ of people of reproductive age. In Poland it affects ca. 1 million couples.

One of the most important issues when trying to conceive (TTC) is to correctly determine the fertile window. This knowledge allows one to maximize chances of getting pregnant in the current cycle since fertile days fall $+/-2$ days before/after ovulation. The most popular method of ovulation prediction uses basal body temperature (BBT), but its accuracy can be effectively disrupted by incorrect measurements or an active infection. There is also a method based on ovulation tests detecting LH peak, but a significant bias linked with low quality of kits should be considered.

Another method of ovulation prediction is based on menstrual cycle length, but first one has to determine the length of the current cycle. Nowadays there are many applications with period tracking options. Most of them show predictions of the length of upcoming cycles based on standard algorithm. Except from data on the length of bleeding or cycle, the apps allow users to register additional information i.e. symptoms, mood, BBT, cervical mucus, which can help indicate the current phase of the cycle.

In the literature research groups proposed many extensions of baseline prediction algorithms. The prediction in example were made on BBT and heart rate (HR) record [3], where the model estimated probability for a day if it's in a fertile window. Another groups marked prediction accuracy improvements by applying autoregresive models to BBT and vector of previous cycle lengts [1] or a generative model, that predicts user's next cycle length and estimates likehood to skip tracking cycle in the application [2].

Using the database of menstrual cycles of users from the Polish application iYoni (over 300K registered cycles by over 80K users), we prepared datasets containing menstrual cycles and basic users' attributes (e.g. age, cycle regularity, months of TTC). The datasets were used to develop and train machine learning models (including regression models and deep models) that predict the length of the menstrual cycle. The application of machine learning improved the quality of cycle length prediction relative to baseline algorithms (returning average/median/last cycle length).

We will discuss current results and opportunities to develop models that make predictions based also on the aforementioned daily measurements provided by users and medical interviews.

Data preprocessing

For this study, we used a database of self-tracked menstrual cycles from the iYoni App by Lifebite. The database contains over 300 thousand menstrual cycles from about 80 thousand users. Before applying machine learning algorithms we needed to deidentify and clean the data.

In the first step, we removed menstrual cycles that had *NaN* on any of its attributes (end of cycle, end of period, end of cycle). Next, we filter out data tracked by our test accounts and menstrual cycles inserted as historical records before app publication - cycles that start before 01/01/2019.

In the next step, we filtered out users that didn't have at least cycles as the threshold. We defined 3 different thresholds t $(4, 6, 11)$ that determined the dataset that was used during training and evaluation. The next steps were applied to every version of the dataset. We removed unconfirmed cycles - current (not finished) cycles. Then based on medical criteria we removed cycles that were shorter than 10 days or longer than 100 days.

After initial filtering, we prepared an uninterrupted series of length *n* of menstrual cycles for every user. We removed the series in that:

- The day of the end of the previous cycle is one day before the beginning of the next cycle.
- Cycles that are the beginning of a pregnancy or occur after a pregnancy.

Every threshold for every threshold the *n*-parameter was different, $n = t - 1$, because we need one cycle to be the expected output.

Dataset

After data preprocessing, we had to split data into 3 datasets (training, validation, and testing). In order to avoid *data leackage* we decided to prepare sets by dividing users into different sets. We divided users into 3 groups:

- training - ca. 60% of users,
- validation - ca. 20% of users,
- testing - ca. 20% of users,

Every prepared dataset had two sets of attributes:

- Menstrual Cycles Lengths (MCL) - vector of n previous menstrual cycle lengths.
- User Attributes (U) - vector of attributes describing the user (age, boolean flag - *isCycleRegular*, standard deviation of previous cycles, current cycle day).

Table 1 presents basic information about prepared training datasets.

Dataset name	No. Users	No. Menstrual cycles	μ_{MCL}	σ_{MCL}
MCLU-3MCL	9 582	107 572	29.42	6.80
MCLU-5MCL	6 875	94 068	29.29	6.49
MCLU-10MCL	3 629	69 182	28.93	5.99

Table 1. Basic training datasets information.

Models

In our work we proposed deep learning models based on fully connected layers. We trained two groups of models: the first group had as input only previous cycles lengths (MCL), the second one processed also attributes describing user (U). The outputs of our models were single number - next cycle length or a normal distribution.

In order evaluate our models we defined 3 baseline models:

- Last Cycle Length Model (LCLM) - returns previous cycle length as a prediction of next cycle length.
- Mean Cycle Length Model (MeanCLM) - returns mean of previous cycles lengths as a prediction.
- Median Cycle Length Model (MedianCLM) - returns median of previos cycles lengths as a prediction.

In the study we compared prepared models to other machine learning algorithms such as Bayessian Regression, XGBoost, Stochastic Gradient Descend Regressor.

Results

Table 2 presents results achieved on *MCLU-10MCL*. XGBoost, marked in bold obtained the smallest errors on Root Mean Squared Error (RMSE) and Measn Absolute Error (MAE), but it was highly overfitted to the training dataset. Marked with underline - *FC-REG-MCLU* achieved second results on presented metrics. All defined baseline methods made larger errors than methods using machine learning.

Model Name	RMSE-train	RMSE-val	RMSE-test	MAE-train	MAE-val	MAE-test
FC-REG-MCLU	<u>4.59</u>	<u>4.66</u>	<u>**4.90**</u>	<u>2.59</u>	<u>2.61</u>	<u>2.74</u>
FC-NORMAL-MCLU	5.86	5.82	6.32	3.03	3.05	3.22
FC-WEIBULL-MCLU	5.02	4.98	5.25	2.99	2.96	3.12
SGD-MCL	5.96	5.96	6.44	3.01	3.02	3.20
SGD-MCLU	4.80	4.86	5.10	2.72	2.75	2.87
BayesianReg-MCLU	4.77	4.84	5.06	2.71	2.75	2.86
XGBoost-MCLU	**3.58**	**4.65**	**4.90**	**2.15**	**2.57**	**2.69**
LCLM	7.70	7.90	8.33	3.94	4.02	4.22
MeanCLM	6.03	6.00	6.52	3.07	3.28	3.08
MedianCLM	6.04	6.00	6.56	2.86	2.87	3.06

Table 2. Evaluation results on MCLU-10MCL.

Results observed on the two remaining datasets were similar. In most cases, model *XGBoost-MCLU* was the best but also over-fitted, the second place took *FC-REG-MCLU*.

Table 3 presents results of *FC-REG-MCLU* on all datasets compared to the best baseline model.

Model Name	RMSE-train	RMSE-val	RMSE-test	MAE-train	MAE-val	MAE-test
			MCLU-3MCL			
FC-REG-MCLU	5.25	5.40	5.24	3.02	3.08	3.03
MedianCLM	7.28	7.47	7.41	3.59	3.67	3.63
			MCLU-5MCL			
FC-REG-MCLU	5.12	5.10	5.02	2.87	2.87	2.83
MedianCLM	6.87	6.73	6.70	3.30	3.25	3.22
			MCLU-10MCL			
FC-REG-MCLU	4.59	4.66	4.90	2.59	2.61	2.74
MedianCLM	6.04	6.00	6.56	2.86	2.87	3.06

Table 3. *FC-REG-MCLU* evaluation results compared to best baseline model on all datasets.

We defined two regularity intervals SHORT (26-30 cycle days) and LONG (25-35 cycle days). Based on intervals we divided users into two groups with regular cycles - all previous cycles have length in specific interval, and with irregular cycles - at least one cycle length is out of interval. Then we evaluated models on all groups. For regular-cycle users *FC-REG-MCLU* achieved about 93% on OneDayMiss (ODM) - accuracy $+/-1$ day, when the baseline methods were LCLM 67%, MeanCLM 89%, MedianCLM 75% (results on *MCLU-3mcl*, *SHORT* interval).

Conclusions

In our work, we proposed a deep learning model that improved the quality of next cycle length prediction. The trained neural network achieved better results in comparison to baseline methods. The designed model performed better than other machine learning algorithms and unlike XGBoost didn't overfit the traning dataset.

Future work

The improvement of prediction quality is not the only goal to be achieved by machine learning models. We want to interpret models' predictions. This step of improvement could provide valuable information not only for research purposes but also improve the user's experience when using the application.

In the next steps of development of our models, we want to extend datasets with data from *daily measurements*. Using the measurements could improve the quality of the next cycle prediction. In addition, it will also allow us to improve the interpretability of models.

The application iYoni allows users to fill out medical interviews and receive personalized advice. Users answer questions defined by clinicians about their health. The data from interviews could be used to analyze user health, detect fertility disorders, and return more accurate predictions.

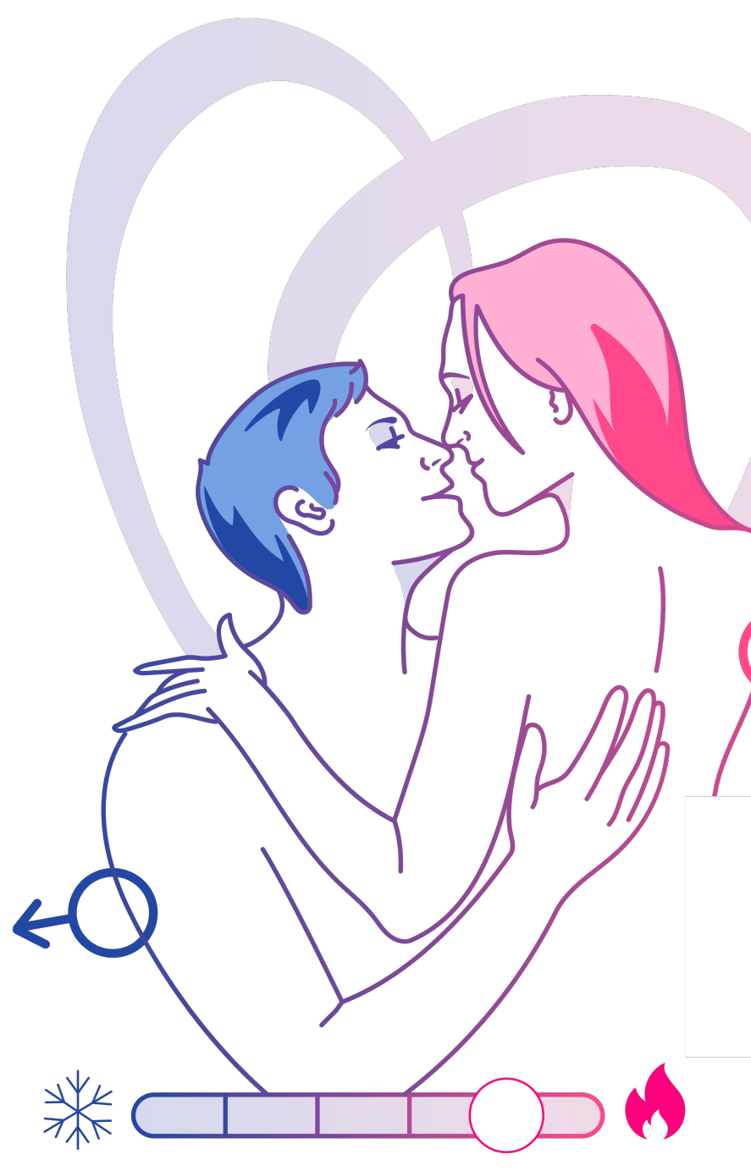
Currently, our models are based on fully connected layers. Application of recurrent neural network (RNN) may result in an improvement of next cycle prediction quality. The vector of previous menstrual cycles lengths has a time dimension, so the use of RNN could reduce prediction errors.

References

[1] Ai Kawamori, Keiichi Fukaya, Masumi Kitazawa, and Makio Ishiguro.
A self-excited threshold autoregressive state-space model for menstrual cycles: forecasting menstruation and identifying ovarian phases based on basal body temperature.
Statistics in Medicine, 38, 07 2017.

[2] Kathy Li, Iñigo Urteaga, Amanda Shea, Virginia Vitzthum, Chris Wiggins, and Noémie Elhadad.
A predictive model for next cycle start date that accounts for adherence in menstrual self-tracking.
Journal of the American Medical Informatics Association, 29, 09 2021.

[3] Jia-Le Yu, Yun-Fei Su, Chen Zhang, Li Jin, Xian-Hua Lin, Lu-Ting Chen, He-Feng Huang, and Yan-Ting Wu.
Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms.
20, 08 2022.



iYoni app