2018

Open Cycle: Forecasting Ovulation for Family Planning

Karen Clark  
*Southern Methodist University*, krclark@smu.edu

Mridul Jain  
*Southern Methodist University*, mridulj@smu.edu

Araya Messa  
*Southern Methodist University*, amessa@smu.edu

Vinh Le  
*Southern Methodist University*, vtle@smu.edu

Eric C. Larson  
*Southern Methodist University*, eclarson@lyle.smu.edu

Follow this and additional works at: [https://scholar.smu.edu/datasciencereview](https://scholar.smu.edu/datasciencereview)

Part of the [Investigative Techniques Commons](https://scholar.smu.edu/datasciencereview), and the [Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons](https://scholar.smu.edu/datasciencereview)

**Recommended Citation**  
Available at: [https://scholar.smu.edu/datasciencereview/vol1/iss1/2](https://scholar.smu.edu/datasciencereview/vol1/iss1/2)

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit [http://digitalrepository.smu.edu](http://digitalrepository.smu.edu).
Open Cycle: Forecasting Ovulation for Family Planning

Karen Clark¹, Mridul Jain¹, Vinh Le¹, Araya Messa¹, and Eric C. Larson¹

¹ Southern Methodist University, Master of Science in Data Science
Dallas, Texas 75275
(krclark@mail.smu.edu, mridulj@smu.edu, vtle@smu.edu, amessa@smu.edu, eclarson@lyle.smu.edu)

1 Introduction

Historically, biological knowledge of a woman’s menstruation cycle has been a black box. It was not until the 19th Century that doctors realized it was linked to ovulation and a woman’s fecundity. Motherhood is often a positive and exciting time, but unanticipated pregnancies and struggles of getting pregnant are often the most stressful issues dealt within a relationship [1]. Family planning enables a couple to space and limit pregnancies that has a direct impact on a couple’s well-being and, often, directly impacts the health of the mother [2]. Predicting ovulation can be valuable for couples because it allows planning around times of increased fertility. However, it is difficult to predict ovulation because not all women follow a normal cycle. There are many factors such as age, health, and nutritional status that can have a direct impact to women’s menstruation cycle. There have been many techniques used by women to predict ovulation, but each technique has its own issues and no method has been found to be universally applicable for all women.

2 A Primer on Basal Body Temperature and Family Planning

Pregnancy, especially unintended pregnancies, are also associated with many negative health and economic consequences that could be avoided through family planning. From an economic perspective, family planning can increase women’s labor-force participation, increase wages, and increase lifetime earnings. From a health perspective, women that understand the benefits of family planning can space pregnancies out at least 2 years apart to allow time to care for the new baby and to recuperate after the birth of a child. Frequent and numerous pregnancies are more likely to lead to maternal death from hemorrhage, toxemia, or septicemia [3]. During pregnancy, a woman may be faced with severe or chronic malnutrition due to sickness. Thus, her defenses to fight infection or illness will decrease. Couples who plan their births for the times when the mother is best prepared avoid high-risk pregnancies [3].

Over time knowledge about the female reproductive physiology has grown and natural methods of family planning have started to gain popularity [4]. A difference in
a woman’s body temperature was first documented in 1868 by Dr. W. Squires, a physician at a Tuberculosis Sanatorium. He noted there were two different temperature levels recorded in female patient charts, but he had no explanation for these differences. Later in 1904, Dr. Van der Velde Holland concluded that the temperature changes observed were related to ovulation [5]. It wasn’t until the 1930’s and early 1940’s that the temperature change was found to correspond to the hormonal and endometrial changes that occur with ovulation. Dr. Zuck, an American physician, established that a woman was at her highest fecundity during these temperature changes [6]. Finally, in 1947, a Belgium physician recommended that women should monitor their basal body temperatures to help with family planning. This sounded easy, but physicians found that the basal body temperature may only increase 0.5 to 1.0-degree Fahrenheit, and in most instances, it could take four to five days for the temperature to stabilize. In 1963, Dr. John Marshall, a neurologist conducted a prospective trial of fertility regulation by monitoring a woman’s basal body temperature method from a data set that included temperature readings from 1,798 women who took part in the natural family planning clinic provided by the Catholic Marriage Advisory Council of England and Wales in 1950 [7].

3 The Catholic Marriage Advisory Council Data Set

The data used in our study is based on the same data set that Professor John Marshall (Department of Statistical Sciences of the University of Padua) used in his study of women’s menstrual cycle and predicting ovulation by monitoring a woman’s basal body temperature. The data set is registered information from 1,798 women, each having at least one sequence of six cycles. A cycle is defined as a menstrual cycle starting from the first day of the onset of menstruation up to and including the day before the next cycle begins. The total number of segments were 2,397 with each woman having a maximum of eight distinct recorded segments. The statistical information from the charts sent by women was converted into the electronic data to perform analysis. The women who took part in this study were not charged for services or discriminated against based on race, nationality, or religious affiliation [7].

Data was captured when the family planning service began in 1950, initially using calendar method and later superseded by Basal Body Temperature (BBT) method. This method used the temperature shift associated with ovulation to determine the fertile and infertile phases of the menstrual cycle. The analysis of the BBT was used on a single spike among the last six lower temperature levels. The spike was defined by an increase in BBT of 0.2 degrees Fahrenheit above the corresponding temperatures. The predicted ovulation was then recorded for that cycle. The original prediction of ovulation was done manually as the data was analyzed after the participant sent their temperature recordings back to the family planning service. The complete data set is comprised of 36,140 instances (this number excludes the 662 cases where charts or data were missing) and 112 features. Age was captured for only 1,074 participants and ranged between 18 to 50 years of age. The assumption made was that the temperature spike seen could be used and ovulation could be calculated. The participants included women that were considered to have no fertility issues,
therefore, there was no reason to believe their cycles are abnormal. The data set is a time series with only BBT recordings. It does not provide any other data such as cervical mucus observations or hormone levels. Additional information was included in the 14,231 cycles, in which the BBT was recorded every day, from the first to the last day of the cycle. In 14,520 cycles, only the first two temperatures were missing and in 14,564 cycles only the last two temperatures were missing [7]. This information is important because the data set used provided us with a robust set of usable BBT records to use in our predictive analysis.

The data was split hierarchically into levels.
1. Unique identifier for each participant.
2. Group identifier for each set of consecutive cycles per participant.
3. The number of cycles within the group.
4. Cycle identifier
5. Total number of cycles for that participant.
6. Date of the first menstrual cycle for the participant.
7. Time span of the cycle.
8. Manually calculated preovulation day for the participant.
9. Length of menstruation cycle in days for the participant.
10. Temperature readings ranging from 1 to 99 individual recordings.

Each participant’s menstrual cycles were collected within “groups” of consecutive cycles. The consecutive cycles were then grouped together for a single woman (“Donna” in Italian) [7].

3.1. Data Characteristics
Rudimentary visual analysis of the sequences of cycle length and preovulation phase length provides some common features that were observed [8]:

1. The data was recorded in days, and observed length of cycles are discrete.
2. The length of a woman’s cycle did show a trend downward as a woman’s age increased.
3. There were several instances where women recorded long cycles and there were other instances when women recorded a change in the mean level. We can only assume this is representative of a change in life style, health, physical, or behavior since we were not provided any ancillary information.
4. We did see that long cycles in the data set were commonly followed by a shorter cycle. This did impact our ability to predict ovulation for the second cycle when the temperature set was not long enough for the algorithm.
5. Data from some subjects, included either small or large observations, can only be interpreted as outliers when compared to the woman’s normal pattern. Some anomalous cycles could be related to biological changes or sickness but there is no supporting data for this conclusion.
6. Heterogeneity or abnormal readings were observed across all participants that were included in the research analysis.
4 Data Analysis

Each subject provided a sequence of at least 6 consecutive menstruation cycles, defined from the first day of menstruation through the start of the next cycle. Some women supplied menstruation dates and temperatures multiple times in a sequence that were greater than or equal to 6 cycles [7]. A segment is a sequence of 6 consecutive cycles.

![Recorded cycles - Qualification](image1)

![Data recorded for segments - Qualified Cycles](image2)

Figure 1: Reported Cycles by Qualification and Segments. The top chart above provides a visualization of the breakdown of the records in the data set. Only 1.82% of participants were missing chart data, which is a very small percentage and did not have an impact on our analysis. The data set was comprised of 86.58% where participants provided complete information. The one aspect that had the greatest impact on our analysis was the 8.89% of records where critical temperature readings were missing. The bottom chart provides a visual representation of the number of cycles that were submitted by participants (this is further explained in Figure 2).

Figure 1 provides a visualization of the breakdown of the data collected. A total of 656 (1.82% of the dataset) charts lack any information about them and were excluded...
from the study. A total of 31,290 (86.58% of the dataset) charts were complete. A total of 3,211 (8.89% of the dataset) has information but missed critical temperature readings. After comprehensive analysis and performance comparisons, we finally ended up replacing the missing critical temperatures with the median values of each cycle temperatures. The breakdown in Figure 1 provides proof that even though the data set was collected in 1950s manually, it still provided robust and well-rounded data for our analysis.

To make sure the data set provided enough participants that had at least 6 qualified consecutive menstruation cycles, we examined the total number of consecutive cycles per participant to evaluate the impact this might have on our analysis. It was important to have enough qualified consecutive cycles (qualified in the graph represents cycles that are greater or equal than six consecutive cycles) to validate the prediction of ovulation. Figure 2 provides a breakdown of subjects with complete charts that have qualified consecutive cycles which means that have more 7-10 consecutive analysis which provides an adequate representation for the analysis.

![Figure 2: Number of Consecutive Cycles by Subject and Segment](image)

The length of the menstrual cycle for a participant was also important to analyze and evaluate any odd length cycles to decide if they were true in length or if unknown or human mistakes could have contributed to their extended length. The Women’s Health and Service Department reports that a typical menstrual cycle will average around 28 days [9], so any cycle that was over this value warranted a closer look.
Figure 3: Record Menstruation Cycle based on expected ovulation day. In our data set, we see that for 95% of the qualified cycles that the cycle length is less than or equal to 34 days. 80% of these cycles are under 30 days in length and the average and median cycle length is approximately 28 days.

We are confident that the data set will provide us a true representation of a “normal” cycle based on our analysis of the cycle length. This is an important statistic to consider in our prediction analysis because women will ovulate anywhere between day 11 to day 21 of their menstrual cycle. This is known as the most fertile time of a woman’s cycle [9]. This information also provides a means to validate our findings because the day of ovulation will determine the length of a woman’s cycle [9]. Where we are able to accurately predict ovulation we can also validate the length of the cycle recorded.
Figure 4: Length of Menstruation Period in Days. Above is visual representation of the average menstruation period in days. We found that 96% of the recorded qualified cycles recorded less than or equal to 7 days of menstruation period. Out of these 88% had less than 6 days. Overall the mean and median period length was close to 5 days. The graphs also show that 1,179 cycles have a menstruation period length of zero.

Knowing the menstrual period length is important in predicting ovulation because the first day of the last menstrual period is known as the follicular phase and this phase continues until ovulation occurs [10]. The follicular phase is when the egg follicle on an ovary is ready to release an egg. The first 5 days of the follicular phase to ovulation is the most fertile window [10].

The egg can be released any time from day 7 to day 22 of a normal menstrual cycle, and the woman’s basal body temperature should show a spike 12 to 24 hours after [10].
Figure 5: Pre Ovulation days i.e. “Number of days before BBT rise”. The figure tells us how many days were recorded with no rise in BBT before the ovulation occurred and temperatures were seen to be increased. 95% of the cases, the pre-ovulation duration is less than or equal to 22 days, and in 83% of the cases the pre-ovulation duration is less than or equal to 18 days. The mean days before BBT rise is close to 13 days and the median number of days is 15 days.

There are 1,992 cases where the pre-ovulation days are 0, and nearly 10% of the cases where the pre-ovulation days are less than 10 days. The data set overall validates that ovulation is occurring during day 7 to day 22 of a woman’s menstrual period which provides a sound base for our prediction model.

Our basic data analysis has provided some fundamental knowledge of the data, but the recordings of the basal body temperatures are the most important component of the prediction model. We randomly selected several participants with 6 consecutive qualified cycles to evaluate the temperature recordings. The prediction model is based around the “three-over-six” rule, so we manually reviewed the temperature readings provided to see if we could visually see the spike in temperature. We found that one participant had a cycle length of 16 days but the manually predicted pre-ovulation was 0 in the data set. Further analysis revealed that the BBT (“three-over-six”) was being calculated after the 16th day of the menstrual period but no temperature reading was recorded for this day.
Figure 5: Basal Body Temperature (BBT). Above is a visualization of the 3 over 6 rule. Temp 1 shows a dip in the 3 temperatures and then a gradual increase in temperature. Temp 2 doesn’t show as drastic decrease in temperature as the others but remember the change in temperature can be as small as 0.2-degree Fahrenheit. The other temp sequences continue to display what we expect to see in a normal cycle.

Due to the inconsistent nature of the temperature readings, we scaled them down between 0 to 1 for our analysis. Scaling the temperatures helped by smoothing the variable. This allowed us to see the “saddle pattern” of the temperatures as the cycle moved through the different phases.
Figure 7: Basal Body Temperature graph for Inconsistent Temperature records. Subject ID 39 and cycle number 4 is an example of some of the data issues examined in the data set. The pre-ovulation days are 0 but the cycle length is 16 days. When we applied the 3 over 6 rule we saw that the cycle is really starting on the 15th day, but there are no temperatures available after the 16th day.

Missing temperature readings was just one from the many issues we had to deal with while working on this data set. We also found that there were 253 instances where the ovulation day was greater than 30. Each of these scenarios are abnormal and would greatly affect our ability to forecast the ovulation day.

Figure 8: Basal Body Temperature readings with menstruation cycle phases [10]. The figure above shows the different phases of the menstrual cycle. The Follicular phase, in which an egg is released and is considered a woman’s most fertile period. Ovulation indicates the actual
release of the egg and then the Luteal Phase occurs after ovulation but before a woman’s menstrual period begins.

The reason why temperature readings are important is because each temperature marks a point at the beginning, middle, and end of a woman’s menstrual cycle. If a temperature reading is missing, then a point in that phase is missing and the algorithm could miss a vital change in the temperature. With this information, we applied data cleansing techniques described in the next section.

5 Dataset Preparation and Pre-processing

In our dataset, the most important quantity is the BBT recorded by the women for every valid cycle. The issue with this data is that there are many missing values and sometimes the recorded data is not correct.

To make sure we maintain data integrity, we examine various options. The best one was to standardize or normalize the temperatures so that they scaled 0 to 1. This would help with the Long Short-Term Memory model because it is known to be sensitive to the scale of input data.

We removed all records where the pre-ovulation was greater than 30 days. This is an unrealistic value and one that the model would not predict correctly.

One key observation made during the data set analysis is that some temperatures were unrealistic. What we mean by this statement is that some temperatures were either higher or lower than what would be humanly acceptable. However, scaling the temperatures did help with this issue because we were still able to keep them between 0 and 1.

We created a new data set with all valid records including the subject id, cycle id, and pre-ovulation day as a rolling and reducing window. The idea was to create small incremental temperature values that could be fed into the selected models as shown below in Figure 8.

![Figure 9](https://example.com/figure9.png)

Figure 9: Training dataset has the temperature readings entered into the dataset one at a time and we keep total 9 temperature readings at any given time. The older temperature readings are moved to the right and new temperature readings are entered into the dataset. The target pre-ovulation date is our target variable, and it reduces by one as the new temperature value is entered.
The reason for this approach is that subjects will keep adding additional temperature values that will need to be fed to the model incrementally, so it can predict ovulation. The final data set was expanded into a convolved data set in which a specific number of temperatures, we chose 6, are kept in a rolling window. We first added 6 temperatures in one record for that subject and cycle ID. Then we continued inserting temperatures in this manner until all temperatures were included. The model would try to predict ovulation each time a new set of temperatures were read. We added a new temperature variable to reduce the days to ovulation by one as we build the convolved data set.

Another variable was added to capture the mean value of ovulation from the previous month. The idea behind this, as more data is available, is that the model will use the mean value of the previous month to predict the current month's ovulation. Additionally, we also used previous months' cycle length as another variable. The reason we use this variable is to make sure we can capture any unusual cases where the Ovulation date is abnormally higher and has correlation with the previous months' cycle length. The last month rolling pre-ovulation days were also taken as inputs to make sure that we have previous months' rolling ovulation memory in the model as shown below in Figure 10.

![Data Table](data-table.png)

Figure 10: Cycle and pre-ovulation lengths from previous cycles of subject in question

We also noticed that the 3-over-6 rule works best if all the data is available, but in many cases, due to practical limitations, we discovered that there are many cases when the data was either missing or there were abnormal temperature values. The reason for this could be due to high activity prior to recording the temperature, fever or other related issues. To make sure to capture all the data issues and make the data very robust against these abnormal temperature readings, the model calculated and used the following fields:

1. **Segment ID**: Identifies the subject or participant.
2. **Cycle ID**: Identifies the cycle within the subject.
3. **Pre_PrevOv**: Represents the previous ovulation date.
4. **Pre_1_CYCLE**: Represents the cycle length of the previous cycle.
5. **LastMonth_rolling_PREOVLATION**: Represents the rolling ovulation memory from the last month.
Calculating the temperature differences provided a set of temperature values that were stable and could be used in the model for training. This resulted in a very robust dataset, which our model could consume and produce desired results.

The final features that were used to train the CNN and MLP models are shown below:

1. Med3-Med6, which is the median value of the recent 3 temperature readings minus median value of the last 6 temperature values
2. Med3-Mean6, the median value of the recent 3 temperature readings minus mean value of the last 6 temperature values
3. Med3-Min6, the median value of the recent 3 temperature readings minus minimum value of the last 6 temperature values
4. Mean3-Med6, the mean value of the recent 3 temperature readings minus median value of the last 6 temperature values
5. Mean3-Mean6, the mean value of the recent 3 temperature readings minus mean value of the last 6 temperature values
6. Mean3-Min6, the mean value of the recent 3 temperature readings minus minimum value of the last 6 temperature values
7. Max3-Median6, the maximum value of the recent 3 temperature readings minus median value of the last 6 temperature values
8. Max3-Mean6, the maximum value of the recent 3 temperature readings minus mean value of the last 6 temperature values
9. Max3-Min6, the maximum value of the recent 3 temperature readings minus minimum value of the last 6 temperature values
10. Sum_Of_All - The sum of all the above differences calculated above
13. Pre_PreOvL1 - Pre-Ovulation days of the last cycle (if available, else median value of the temperatures for that cycle)
14. Pre_PreOvL2 - Pre-Ovulation days of the second last cycle (if available, else Pre_PreOvL1 value of the temperatures for that cycle)
15. Pre_L_CYCLE1 - Cycle length of the last cycle (if available, else median value of data)
16. Pre_L_CYCLE2 - Cycle length of the second last cycle (if available, else Pre_L_CYCLE1 value of data)
17. NumOfTempsEntered - number of temperatures entered for that cycle
18. LastMonth_rolling_PREOVULATION - Continuous counting of days from last cycle to new pre-ovulation

Our final data set consisted of 3,662 records and 25 variables as shown in figure 10.1 below. For training and validation purposes, the data set was further divided using the 80:20 rule, with 80% of the data being used for training and the remaining 20% used for testing/validation.

![Input features feed into 1D-CNN or MLP model. The above figure is just a representation of the input features that are used to build the Convolutions Neural Network and Multilayer Perceptron model](image)

### 6 Analysis of Results

Time series prediction problems is a common method of predictive modeling. The predictive complexity of this type of modeling is the result of a sequence of dependence that exists among the input variables. We examined several different types of predictive modeling algorithms to see which one provided the best results. Our analysis started with Convolutional Neural Network (CNNs) which is a type of network that has recently gain popularity with past classification problems.
Even though CNNs was originally designed for image classification, it has grown in popularity in time series predictions. CNNs are made up of convolutional layers that are constructed using one-dimensional kernels that move through the sequence. The kernels act as filters which are being learned during the training phase [11]. We used a Tensor flow backend for our first CNN model which only allowed us to predicted ± 2 days for ovulation across the training set. This means that ovulation could occur in 2 days in the future or has occurred 2 days prior. This model did not provide a close enough forecast for a woman to depend on it for family planning.

Since CNN has provided decent results with other time series forecasting, we tried it using the SoftMax classifier. The SoftMax classifier allows us to predict ovulation based on a categorical discrete variable but apply a linear algorithm for a continuous variable. We used this model to try to predict how far we were from our target ovulation day by using the prediction error in terms of absolute days.

![Figure 11](image-url)  
**Figure 11:** CNN with SoftMax classifier Mean Absolute Error Results. Clearly, the MAE of this model reduced initially, and the model achieve close to 2 days of absolute error. At the same time, the model only achieved a 30% accuracy in terms of the categorical variable value prediction. After a deeper evaluation of the results, we found that instead of accuracy we should have based the model’s success on the accuracy for days closed to the actual pre-ovulation dates.

We stepped back to evaluate CNN model using the Tensor flow backend and the SoftMax classifier in predicting our target ovulation day. The CNN using the Tensor flow backend did not provide the prediction value of ± 2 days prior to ovulation. However, the CNN using the SoftMax classifier performed well in predicting ovulation the closer the algorithm came to the actual observed day. In the end, this model could reach approximately 90% accuracy in predicting the ovulation day at the time of ovulation.
Figure 12: CNN with SoftMax Classifier Confusion Matrix. The confusion matrix clearly shows that as it approached ovulation the model is able to accurately predict it. The yellow column (Day 0), then we can see that this is the date of the actual ovulation. The CNN with SoftMax classifier model is also able to predict approximately 60% of the time before ovulation occurs.

Since the temperatures in the data set are dependent on each other we decided to try the Multi-Layer Perceptron model. This model is based on a feedforward artificial neural network. The model provided the same accuracy level as the CNN with SoftMax classifier when evaluating the accuracy in terms of the categorical variable prediction.

Figure 13: Multi-Layer Perceptron (MLP) Model Confusion Matrix. The results were similar to what we saw with the confusion matrix from the CNN with SoftMax classifier. The accuracy of predicting ovulation on the day of ovulation is again roughly 90% with the model being able to predict around 61% of the time before ovulation. The MLP models shows a slight improvement over Day 1 by providing a slightly higher accuracy value.
The MLP model performed slightly better than expected. Figure 14 compares the Prediction Absolute Error on the day of ovulation versus the Prediction Absolute Error 1 day before ovulation to show how far off the model was predicting the actual ovulation day. Neither the MLP nor CNN using SoftMax provided a high enough accuracy value when predicting ovulation before it occurred.

![Figure 14: MLP Prediction Absolute Error Analysis](image)

**Figure 14**: MLP Prediction Absolute Error Analysis. The figure visually shows the predicted versus actual pre-ovulation day of few subjects and cycles along with the absolute error. The absolute error improves as the model approaches the pre-ovulation day. In some cases, the model still cannot predict accurately as seen above from Subject 946, Cycle-ld – 3. Other subjects and cycles the model predicted very close to actual observed ovulation day.

The last model we felt might give us the results we were looking for is the Long Short-Term Memory (LSTM) model which is a recurrent neural network model. The model provides the ability to classify, cluster, and make predictions on a data set, particularly a time series or text data set. This is a great benefit where classical linear method can be difficult to adapt to multivariate or multiple input forecasting problems. The model helps preserve the error that can back propagate through the time and layers.

The first step is to prepare a dataset for the LSTM model. This involves framing the dataset as a supervised learning problem to predict ovulation at time (t) given the feature predictors at the prior time step (t-1). This is shown in Figure 15 (predictors) and Figure 16 (target variable). The predictor values included the following:

1. The first 9 temperature values
2. The median of the most recent three temperatures minus median of the previous six temperatures values
3. The median of the most recent three temperatures minus mean of previous six temperatures values
4. The median of the most recent three temperatures minus minimum of previous six temperatures values
5. The mean of the most recent three temperatures minus the median of previous six temperatures values
6. The mean of the most recent three temperatures minus the mean of previous six temperatures values
7. The mean of the most recent three temperatures minus the minimum of previous six temperatures values
8. The maximum of the most recent three temperatures minus the median of previous six temperatures values
9. The maximum of the most recent three temperatures minus the mean of previous six temperatures values
10. The maximum of the most recent three temperatures minus the minimum of previous six temperatures values
11. The sum of all the differences
12. The number of temperatures entered for that cycle

The target value is the length of pre-ovulation day.

Figure 15: Feature Predictors. The figure above provides a list of the predictor values used in the model.

Figure 16: Target Variable. The figure above provides a list of the target values used in the model.
Figure 17: Mean Absolute Error on Training versus Test set using the LSTM model. We used both the Mean Absolute Error (MAE) and Root Mean Square error (RMSE) to evaluate the accuracy of the model. Both the MAE and RMSE provided average results in predicting the target ovulation day.

For accuracy metrics, we considered both Mean Absolute Error (MAE) (Figure 17) and Root Mean Square error (RMSE) as shown in the analysis. Lower values of MAE and RMSE are preferred for the model to perform well. The LSTM model provided the lowest MAE (to the degree 0.1) and RMSE (0.12) values than the other models evaluated.

The LSTM model also provided ~95% accuracy in predicting the day of ovulation. The actual versus predicted values of this model are shown in Figure 18 below. The figure shows how the model seamlessly forecasted ovulation on and before the observed ovulation day.

Figure 18: LSTM - Predicted versus Actual values. The plot above shows the actual ovulation day versus the predicted ovulation day. We can actually see that they are consistent results.

The LSTM provided the best model for our data set with the highest accuracy of predicting the ovulation day. The value loss is less than 1, which means we are predicting the actual day of ovulation. The good thing with this model is that we can supervise the given features to predict our variable of interest at any time (t).
Another important feature that we observed by utilizing advanced machine learning models is that the model provided resolution to the presence of inter-correlation between similar features (predictors).

7 Ethics

When evaluating the ethical issues around “Natural” family planning using the BBT method, presents a major issue in that it is not an exact science. The models take time to learn a woman’s normal cycle patterns so during the learning stage the actual day of ovulation may not be accurately predicted. If the model does not predict ovulation correctly the couple may be faced with an unplanned pregnancy. This could be financially, physically, and mentally challenging for a couple. Should a couple solely rely on machine learning to monitor their risk of pregnancy?

Religious adherents vary widely in their views on birth control. The women that were part of our study chose to use natural family planning technique because of religious beliefs of the Catholic church. The thought has been that controlled birth control are not true preventive measures but are abortifacients [13]. This belief on its own is an ethical topic but one our research did encounter because of the source of our data.

Family planning, contraceptives, and wanted and unwanted pregnancy will always be part of the ethical questions behind this research topic. Through this research, we could reveal that you can predict ovulation before it occurs, but there are other factors (age, economic status, and health) that need to be considered before solely relying on this method alone.

9 Further Considerations

Further considerations should include re-processing and evaluating the working data set used in this study to ensure that the assumptions and decisions made were the correct ones. Whenever working with a data set that has missing values or values that do not follow a consistent pattern, you run the risk of contaminating your validation or testing set. What this mean is that the validation or testing set may already have some knowledge of the training set based on how the data was divided into the two buckets.

Another concern is that there might be some confounding variable or factor issue introduced with the LSTM model. If this is the case, then the validity of our results could be considered questionable or bias. We did our best to eliminate this possibly in our analysis and provided sound measurable metrics to support the findings of the models evaluated.
10 Conclusions and Future Work

The overall results of the models evaluate (CNN, MLP and LSTM) provided insightful results which can be implemented very quickly. The prediction model did not perform well when there was little or poor input data, but with a robust set of input temperatures and an expanded set of engineered feature sets used by the models would provide a very powerful and accurate model for prediction of the complex and real-world problems.

Going forward our work will focus on tweaking current models and verifying the LSTM models more. The accuracy of LSTM models is much better and can be used for further analysis and practical implementation. Applications using the above-mentioned methodology and models can then be deployed on handheld devices. The main reason of using handheld devices for this type of prediction is that these new devices come with temperature sensors. The body temperature can be easily and automatically captured and sent to the application for prediction. Furthermore, with the devices, we can be confident that it will be able to record the temperature when the person is least active or sleeping. This will assure we get the best readings with no external factors like high activity, and hence, the prediction accuracy can increase tremendously.

Acknowledgments. The authors, Karen Clark, Mridul Jain, Araya Messa, and Vinh Le, would like to give many thanks to Dr. Eric Larson and Dr. Daniel Engels for their assistance, guidance, and feedback on our research and paper.

The paper is part of the work completed as part of the Southern Methodist University Data Science Program for Capstone.

Karen Clark is with the Southern Methodist University, Master of Science in Data Science Program, Dallas, TX 75205 USA (e-mail: kclark@mail.smu.edu).

Mridul Jain is with the Southern Methodist University, Master of Science in Data Science Program, Dallas, TX 75205 USA (e-mail: mjain@mail.smu.edu).

Araya Messa is with the Southern Methodist University, Master of Science in Data Science Program, Dallas, TX 75205 USA (e-mail: amessa@mail.smu.edu).

Vinh Le is with the Southern Methodist University, Master of Science in Data Science Program, Dallas, TX 75205 USA (e-mail: vle@smu.edu)

References


