# Semi-Supervised Ovulation Detection Based on Multiple Properties

Amos Azaria
*Department of Computer Science*
*Ariel University, Israel*

Seagal Azaria
*Labor and Delivery*
*Mayanei Hayeshua Medical Center*
Benei-Brak, Israel

***Abstract*—Despite being a well-researched problem, ovulation detection in human female remains a difficult task. Most current methods for ovulation detection rely on measurements of a single property (e.g. morning body temperature) or at most on two properties (e.g. both salivary and vaginal electrical resistance). In this paper we present a machine learning based method for detecting the day in which ovulation occurs. Our method considered measurements of five different properties. We crawled a data-set from the web and showed that our method outperforms current state-of-the-art methods for ovulation detection. Our method performs well also when considering measurements of fewer properties. We show that our method's performance can be further improved by using unlabeled data, that is, mensuration cycles without a know ovulation date. Our resulted machine learning model can be very useful for women trying to conceive that have trouble in recognizing their ovulation period, especially when some measurements are missing.**

## I. Introduction

Couples who face difficulties when trying to conceive occasionally opt for methods of ovulation detection. Therefore, various methods have been proposed [22]. These ovulation detection methods may be based on measurements of different properties, including Basal Body Temperature (BBT) (that is, body temperature measured immediately on waking up in the morning), Salivary Electrical Resistance (SER), Vaginal Electrical Resistance (VER), Ovulation Prediction Kits (OPKs), which measure the Luteinizing Hormone (LH) present in subject's urine, and Cervical mucus. While all the above measurements have been shown to have some level of effectiveness in detecting ovulation [21], [16], [27], [15], [23], most existing methods rely only on measurements from a single property, or from two properties at most. Therefore, many of these methods do not reach very high accuracy.

In this paper we take a machine learning approach method for ovulation detection that takes into account *five* different measurements. We show that such an approach outperforms any currently used ovulation detection methods, which rely on measurements from fewer properties. The results of this method can be used to detect ovulation with higher accuracy for subjects who have performed all five measurements. Furthermore, due to the high accuracy, we show that such labeling can further be used to label additional unlabeled data, that is, menstrual cycles without a know ovulation date. This data, in turn, can be used for training a model that allows detecting ovulation based on measurements from a subset of these five properties. Our resulted model can be very useful for women trying to conceive that have trouble in recognizing their ovulation period.

Rather than treating the problem as a classification problem, we treat it as a regression problem. That is, given a full menstrual cycle, we detect the day in which the ovulation occurred. We therefore use the difference between the ovulation date predicted by the model and the actual ovulation date as a measure for the performance of our model.

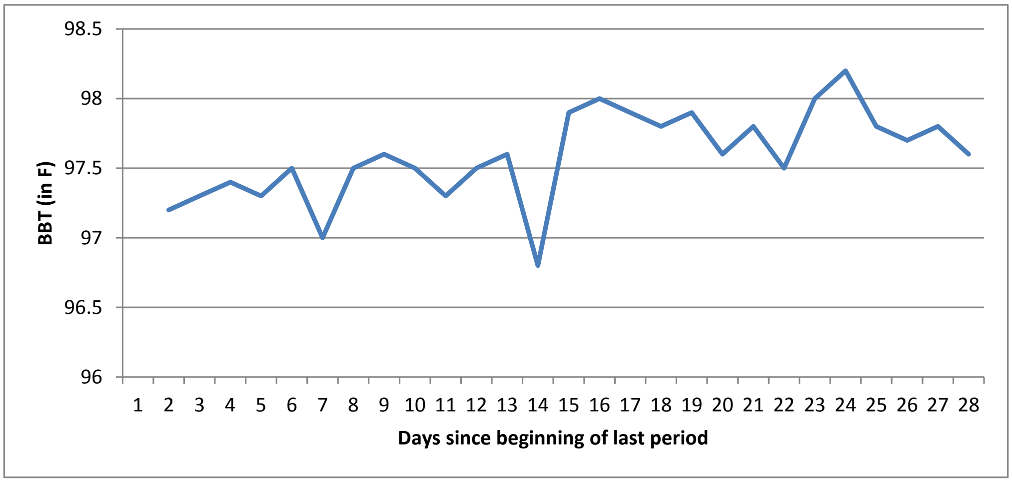
## II. Related Work

Methods for ovulation detection were introduced many decades ago, both as an attempt at helping couples trying to conceive and as a method for birth-control. An increase in basal body temperature (BBT) should be expected immediately after ovulation occurs. Figure 1 illustrates the cyclic pattern in BBT during the menstrual cycle. As can be seen in the figure, such temperatures suggest that ovulation occurred in day 14 of the cycle. However, such a clear ovulation indication of a rise in BBT immediately after ovulation is not that common. Bauman [8] has found that BBT alone is not a good enough predictor, as only in $20\%$ of the cycles Bauman recorded was a rise in BBT detected within a $\pm 1$ day range of the peak in LH (which Bauman used as ground truth).

Most of the current methods for ovulation detection are rule-based; e.g., [14], [27] they do not require a training corpus, and thus cannot be seen as machine learning approaches. One exception is the work by Chen et al. [12], which detects the beginning of menstrual cycles by applying a hidden Markov model with two hidden states, with the input being skin temperature. The measurements in their study were taken every $10$ minutes during the subject sleep period using a special device. Despite ovulation detection being the main goal of Chen et al.'s work, since it is very complex to achieve ground truth with respect to the day in the menstrual cycle on which ovulation has actually occurred, the authors instead use the beginning of the menstrual cycle as their prediction target. Ground truth labels at the beginning of the menstrual cycle were reported directly by the subjects. Their method reached $92\%$ accuracy, after filtering out some false-positive, false-negative, and undetected cycles. Another exception that uses a Bayesian approach for detecting ovulation based upon BBT is [11].

Many of the previous studies have focused on women with normal menstrual cycles [21]. However, infertility issues

Fig. 1. Illustration of Basal Body Temperature (BBT) in menstrual cycle. Notice the slight drop in BBT around the 14th day, followed by a slight rise in BBT on the 15th day. Also note, BBT remains high in most of the following days. These results suggest that ovulation has occurred on day 14.




are often coupled with irregular menstrual cycles [20], thus increasing the importance of detecting the exact ovulation date in women with irregular menstrual cycles.

In this paper we use semi-supervised learning to further improve the performance of the learned model. Semi-supervised learning requires a small labeled dataset as well as a large unlabeled dataset. There are many different semi-supervised methods [31], [7]; we use a bootstrapping semi-supervised self-training method [32].

## III. The Dataset

The dataset for this work was obtained from the OvaGraph website [30]. This website allows its users to record various menstrual characteristics with a goal of assisting them in finding their ovulation days. The website enables its users to record the following properties for each day[1]:

1) Basal Body Temperature (BBT): the lowest body temperature attained during rest, and usually measured right after the user wakes up in the morning. These values are generally between 96 and 99 degrees Fahrenheit, with accuracy of up to 0.01 degree. Measurements are usually taken with a BBT dedicated oral thermometer (i.e., that is usually more accurate in the relevant range of temperatures).
2) Salivary Electrical Resistance (SER): These values are between 10 and 395. Measurements are taken with the OvaCue Fertility Monitor oral sensor.
3) Vaginal Electrical Resistance (VER): These values are between 1 and 399. Measurements are taken with the OvaCue Fertility Monitor vaginal sensor.
4) Ovulation Prediction Kit (OPK): These values are either "Positive" or "Negative". This measurement checks the level of LH in the subject's urine (high LH indicates ovulation).
5) Cervical mucus (5 different classes), breast tenderness (indicated if present), and ov pain (indicated if present): The values for these properties are reported by the users and require no additional accessories. An overview of these properties appears in recent work by Mulcaire-Jones et al. [23].

The website also allows its users to manually label their own assessment of their ovulation date (likely taking into account all recorded factors). One prediction is generated on the ovulation date based on BBT and a different one based on both SER and VER (while the exact algorithm used is not publicly available, it is likely based on [27] and [15], [14], respectively). The website users may choose whether to make their profile private or publicly available.

We crawled 118 publicly available profiles (women) from OvaGraph, which had a total of 283 cycles with measurements for each of the five properties (for a total of 6301 days with measurements). However, only 71 cycles out of these had an ovulation date labeled by the user (for a total of 2069 days with measurements). While these cycles were manually labeled, we believe that subjects with access to measurements of all these properties are likely to be able to detect their own ovulation day. Furthermore, subjects who went through the process of monitoring so many properties are likely to have a well developed understanding of ovulation and awareness of their own body. Such labeling may be the most accurate one might expect to obtain and may be as accurate as methods that detect ovulation based on ultrasound [24].

## IV. Deep Learning Based Methods

We developed two methods of deep learning, a convolutional neural network [19] and an LSTM neural network [17]. The convolutional neural network used two convolutional layers each with 12 filters, each of size $1 \times 6$. The convolution was across days and not across features. LSTM was selected as it performs well with sequences. Both neural networks use dropout for regularization and batch normalization. After

[1]The website enables the recording of additional properties that were not used in this work, such as medication and Home Pregnancy Test (HPT) results.

the convolutional or LSTM layers, both networks have fully connected layers with a final linear activation layer, which predicts the ovulation date.

These deep learning approaches cannot simply ignore missing data, but have to fill it in. Therefore, missing data-points were filled in by the latest known measurement for that specific property, or with the average, if no previous measurement was taken for that property. For example, if the BBT was 97.3 on day 3, but the user did not record her BBT on day 4, the BBT value for day 4 was set to 97.3 as well. If the user did not record her BBT from day 1, it was set to 97.67 which was the average BBT in the data-set. The CNN based method will be referred to as CNN Ovulation Detector (CNN-OD) and the LSTM-based method will be referred to as LSTM Ovulation Detector (LSTM-OD). The advantage of using deep learning methods is that they do not require manual feature generation. However, due to the lack of labeled data, we did not expect these methods to perform as well.

## V. Conditional Random Fields Approach

Another method presented in this paper is based on Conditional Random Fields (CRF) [18]. This method was selected since the data can clearly be observed as a sequence and the features that we use may be redundant. CRF was shown to work well also when its features may capture redundant properties [18]. CRF works well with missing data, a very important feature for our domain, as obviously not all properties had values for every day - and some days did not have measurements for any properties at all.

For each of the five properties, if it was present on a given day, we used the following features for the CRF model: the raw measurement value, the delta from each of the past 5 days and each of the 5 future days, the cumulative sum, and the cumulative sum of the future days. Additional important features that were added include the day within the cycle, the number of days remaining until the end of the cycle, and the number of days in the cycle. In cases in which the data was missing, no feature was emitted. The labeling of each day is either '0', which implies that ovulation did not occur yet in the current cycle, or '1', which indicates that ovulation has occurred in the current cycle. The day in the labeled sequence in each cycle on which there is a shift from '0' to '1' (i.e. the first '1' in the sequence) indicates the ovulation day. We refer to this method as the CRF Ovulation Detector (CRF-OD).

### A. Partially Observed Data

The vast majority of the data in the data-set does not include all properties. In fact, the more common problem is detecting ovulation with only partially observed data; that is, only some properties. In order to train models that only consider a subset of the properties, we only consider features that come from specific properties (and disregard the rest). However, as labeled data, we only use cycles that were labeled by the users and have measurements for *all* 5 properties. Despite the fact that we want to detect ovulation based on a subset of these properties, we cannot rely on hand-labeled ovulation dates to be accurate, if they are based on fewer properties. Since OPK measurement relies on the usage of disposable units, the first model we consider is CRF-OD without OPK. Since both SER and VER require the use of expensive sensors, we also consider CRF-OD without OPK, SER, and VER.

### B. Semi-Supervised Methods

We consider a bootstrapping semi-supervised self-training method [32]. We use the generated model to label our unlabeled data and then use the newly labeled data as if its labels were ground-truth. It has been widely argued as to whether such methods of semi-supervised learning are useful. However, while we will show that we gain some (non-statistically significant) improvement from using self-training, our main goal was to increase the detection of the partially observed data methods. We believe that since the newly labeled data receives its labels from the model that observes all the data, these new labels are quite accurate and are, therefore, useful for training the partially observed data models, which observe only part of the data (i.e., only consider some properties).

## VI. Experiments

### A. Methods Used

As baselines, we used the following 3 state-of-the-art methods. These methods rely either on measurements of a single property, or two properties at most.

1) **BBT** This method is the detected ovulation date based on BBT only, as presented by the website and is most likely based on the CUSUM algorithm which appears in [27].
2) **Color** This method is the detected ovulation date based on both SER and VER, as presented by the website and is most likely based on the algorithm which appears in [15]. These predictions are encoded by color, with "pink" and "purple" indicating that ovulation was detected. In case of multiple predictions, we select the first one to be this method's prediction.
3) **OPK** This method predicts that the first "positive" on the OPK measurement is the ovulation date.

We tested the performance of the following proposed methods in detecting ovulation. The labels used for ground truth were the ovulation dates suggested by the users themselves.

1) **CNN-OD** the CNN based detector as described above.
2) **LSTM-OD** the LSTM based detector as described above. Both deep learning approaches used 10 random trials (500 epochs each), with a split of 51 cycles for training data and 20 for test data.
3) **CRF-OD** the CRF method as described above, which takes all features into account. Leave-one-out cross validation was used.

In addition, we tested the performance of the following partially observed data methods:

1) **CRF-OD w/o OPK** the CRF method as described above, which takes all features *but OPK* into account.

2) **CRF-OD w/o OPK, SER, VER** the CRF method as described above, which takes all features *but OPK, SER and VER* into account.

Leave-one-out cross validation was used for all these methods.

Finally, we tested the following semi-supervised based methods:

1) **SSCRF-OD** the Semi Supervised CRF method as described above, which takes all features into account.
2) **SSCRF-OD w/o OPK** the Semi Supervised CRF method as described above, which takes all features *but OPK* into account.
3) **SSCRF-OD w/o OPK, SER, VER** the Semi Supervised CRF method as described above, which takes all features *but OPK, SER and VER* into account.

These methods used leave-one-out-cross-validation in the following manner. For each of these four methods, in each round, we trained the CRF-OD model on all the training data, except one cycle which was reserved for testing. Note that the CRF-OD model uses measurements from *all* five properties. Using this CRF-OD model, we labeled all the unlabeled data. Once all the unlabeled data was labeled, it was combined with all the training data (excluding the cycle left for testing) and each of the four semi supervised methods used only the features from the relevant properties to train a new model which was then used to label the left out cycle. This process was repeated 71 times (once for every cycle left out). This process ensured that no data was used both as training and as testing.

### *B. Results*

Table I presents the results for the three baseline methods mentioned above. Performance is measured using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), between predicted ovulation day and reported ovulation day. Let $p_i$ be the model's prediction for ovulation day for cycle $i$, $r_i$ be the reported ovulation day for cycle $i$, and let $n$ denote the number of cycles. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (p_i - r_i)^2} \tag{1}$$

and MAE is defined as:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |p_i - r_i|. \tag{2}$$

These results suggest that the OPK measurement, even by itself, is an adequate indicator for ovulation. However, the OPK is one of the least frequent measurements in the full data-set as it is the only measurement that requires a one-time use article.

Table II presents the performance of the proposed machine learning based methods using measurements from *all* properties. As can be seen in the table, CRF-OD performed much better than all other approaches (including the baseline methods) and its detected ovulation days were close to those labeled by users. AAD of 1.00 implies that on average the

TABLE I
PERFORMANCE OF EACH OF THE THREE BASELINE (CURRENT STATE-OF-THE-ART) METHODS FOR OVULATION DETECTION. THE TABLE PRESENTS BOTH ROOT MEAN SQUARED ERROR (RMSE) AND MEAN ABSOLUTE ERROR (MAE) FOR EACH METHOD. LOWER RESULTS INDICATE MORE ACCURATE DETECTION.

| Method | RMSE | MAE |
|---|---|---|
| BBT | 3.301 | 1.944 |
| Color | 5.311 | 4.127 |
| OPK | 2.86 | 1.592 |

TABLE II
PERFORMANCE OF THE PROPOSED MACHINE LEARNING METHODS, BASED ON MEASUREMENTS FROM ALL PROPERTIES.

| Method | RMSE | MAE |
|---|---|---|
| CNN-OD | 4.213 | 3.479 |
| LSTM-OD | 3.507 | 2.307 |
| CRF-OD | **1.895** | **1.00** |

CRF-OD method was off only by 1 day from the actual labeled data. Note, the labels themselves are not perfect, and since ovulation was labeled by different users, we cannot expect a method which will be correct at all times.

As suspected, both deep learning methods (CNN-OD and LSTM-OD) did not perform as well, since we did not have enough training data. Furthermore, despite OPK being one of the values in the input vector, for these methods neither of them performed even as well as the baseline OPK method.

We now turn to test the performance of the methods which do not observe all the data (Table III). As can be seen in the table, while the performance decays as we remove more and more properties, these methods still perform decently well, with the CRF-OD without OPK method outperforming all baseline methods, and the CRF-OD without OPK, SER, VER outperforming the baseline methods which do not use OPK measurements.

The most interesting results come from the semi-supervised methods. Table IV presents the performance of each of the semi-supervised methods. These results are further illustrated in Figure 2 in which they are compared to the performance of the same methods using only the labeled data. The error rate when observing all the properties was improved from a value of 1.00 (MAE) in CRF-OD to 0.901 in SSCRF-OD, and from 1.295 in CRF-OD without OPK to only 1.155 in SSCRF-OD without OPK. While the performance on each of these

TABLE III
PERFORMANCE OF CRF, BASED PARTIALLY ON OBSERVED DATA METHODS.

| Method | RMSE | MAE |
|---|---|---|
| CRF-OD w/o OPK | 2.289 | 1.295 |
| CRF-OD w/o OPK,SER,VER | 2.737 | 1.718 |

Fig. 2. Mean Absolute Error (MAE) of CRF based methods using only labeled data compared to the same methods which also use unlabeled data (lower is better).

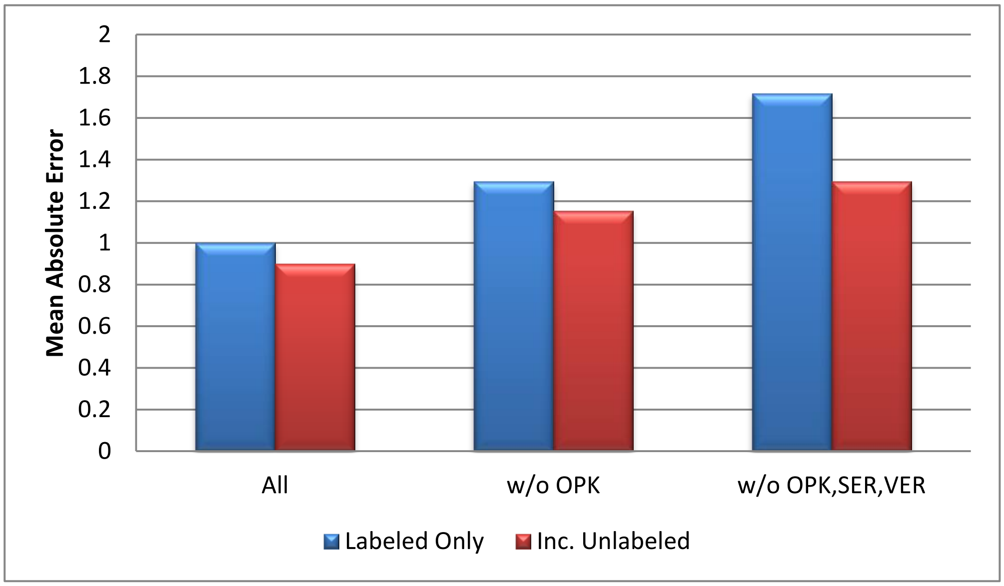



TABLE IV
PERFORMANCE OF THE PROPOSED SEMI SUPERVISED METHODS.

| Method | RMSE | MAE |
|---|---|---|
| SSCRF-OD | 1.838 | 0.901 |
| SSCRF-OD w/o OPK | 2.116 | 1.155 |
| SSCRF-OD w/o OPK,SER,VER | 2.207 | 1.296 |

methods has improved, the results reach statistical significance only for SSCRF-OD w/o OPK,SER,VER ($p < 0.05$) in which the error rate fell from $1.718$ down to only $1.296$.

## VII. DISCUSSION

We believe that the performance obtained by the SSCRF-OD is nearly as accurate as user given labels for ovulation detection *when measurements for all properties* are present. However, taking advantage of unlabeled data with fewer properties represents a more complex problem. Such unlabeled cycles are what the data-set is actually composed of. The full data-set includes $2286$ profiles (women) with a total of $8466$ cycles (with measurements over a total of $183,266$ days). In future work, we intend to use the full data-set in order to develop a method that will attain an increased performance level when using only a subset of the properties.

One limitation of using CRF is that it is a discriminative model. Therefore, confident scores are not particularly valuable. Especially when using only a subset of the properties, it would be very useful if the system could tell the user how confident it is in its prediction. In the domain of ovulation detection, this confidence value could have great impact both on the users trying to conceive and those who may be using ovulation detection as an additional method for birth control.

All results presented in this paper were obtained with very little training data. The nature of the data, which may be split into different views, may suggest using co-training [9]. However, using different techniques of co-training did not outperform SSCRF-OD. Such techniques may become useful for detecting ovulation based on a subset of the properties. Other semi-supervised learning approaches may be harnessed for our problem, as there are so few cycles with measurements for all properties which can be used as unlabeled data. Therefore, approaches such as the ladder framework [25] on the unlabeled data may not work as well either.

This paper assumes that subjects with access to measurements of many properties are likely to be able to detect their own ovulation day. However, since it is well known that humans suffer from many psychological biases (e.g. anchoring, framing) [29], [1], [10], [6], [26], the ovulation day detected may possibly be inaccurate. Nevertheless, the methods used in this paper can be used to better understand human behavior and predict what the subject believes her ovulation date is. This information can then be used by an autonomous agent interacting with a human, when trying to convince the human that the ovulation date is different than what she believes, or when explaining health symptoms. Modeling human behavior to support an agent interacting with a human, is common practice in the field of human-agent interaction [3], [5], [2], [4].

## VIII. FUTURE WORK

While the current work focused on ovulation detection, future research should target predicting ovulation in advance. This task is more challenging, but far more useful in practice, as couples trying to conceive (or avoid pregnancy) must detect ovulation *before* it happens. We would like to note that While our model cannot currently predict ovulation ahead of time, it can help women understand when the ovulation has occurred, which may be very useful for prognosis of infertility and may also assist in prediction of ovulation in future cycles.

We intend to further improve our methods for detecting ovulation based on fewer properties and use these methods to label the whole data-set. Increase in data-set size should allow the use of deep learning methods such as LSTM to predict ovulation before it occurs. Using this data, we hope to be able to predict ovulation in advance also based on a subset of properties.

## IX. Conclusions

Despite the importance of ovulation detection in human female and the considerable research it has attracted, only a few machine learning approaches exist to solve this problem. In this paper, we proposed a machine learning-based method for ovulation detection drawing on measurements of different properties. To the best of our knowledge, this is the first work to take into account more than one or two properties, as our method considers *five* different properties. We show that our CRF-OD yields very good performance measures in detecting ovulation and outperforms current state-of-the-art methods (which do not take into account all properties). We demonstrate that a semi-supervised approach can further increase the method's performance, especially when not all five property measurements are available. Namely, we show that by labeling unlabeled data, our CRF-based approach, which relies on only two of these properties, significantly outperforms its counterpart, which relies only on labeled data (while observing the same two properties), while also outperforming any current state-of-the-art method. Our resulted machine learning model can be very useful for women trying to conceive that have trouble in recognizing their ovulation period, especially when some measurements are missing.

## X. Acknowledgments

The CRF is based on the Java implementation of CRF by Sunita Sarawagi [28]. The deep learning methods use the Keras library [13].

# References


[1] D. Ariely, G. Loewenstein and D. Prelec, "coherent arbitrariness": Stable demand curves without stable preferences, *The Quarterly Journal of Economics* **118**(1) (2003) 73–106.

[2] A. Azaria, Y. Gal, S. Kraus and C. V. Goldman, Strategic advice provision in repeated human-agent interactions, *Autonomous Agents and Multi-Agent Systems* **30**(1) (2016) 4–29.

[3] A. Azaria, Y. Gal, S. Kraus and C. Goldman, Strategic advice provision in repeated human-agent interactions, *Autonomous Agents and Multi-Agent Systems* (2015) 1–26.

[4] A. Azaria, S. Kraus, C. V. Goldman and O. Tsimhoni, Advice provision for energy saving in automobile climate control systems, in *Twenty-Sixth IAAI Conference* (2014)

[5] A. Azaria, Z. Rabinovich, C. V. Goldman and S. Kraus, Strategic information disclosure to people with multiple alternatives, *ACM Transactions on Intelligent Systems and Technology (TIST)* **5**(4) (2015) p. 64.

[6] A. Azaria, A. Richardson and S. Kraus, An agent for the prospect presentation problem, in *AAMAS* (2014) pp. 989–996.

[7] A. Azaria, A. Richardson, S. Kraus and V. Subrahmanian, Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data, *IEEE Transactions on Computational Social Systems* **1**(2) (2014) 135–155.

[8] J. E. Bauman, Basal body temperature: unreliable method of ovulation detection., *Fertility and sterility* **36**(6) (1981) 729–733.

[9] A. Blum and T. Mitchell, Combining labeled and unlabeled data with co-training, in *Proceedings of the eleventh annual conference on Computational learning theory* (1998) pp. 92–100.

[10] C. F. Camerer, *Behavioral Game Theory. Experiments in Strategic Interaction* (Princeton University Press, 2003), ch. 2.

[11] R. L. Carter and B. J. Blight, A bayesian change-point problem with an application to the prediction and detection of ovulation in women, *Biometrics* (1981) 743–751.

[12] W. Chen, M. Kitazawa and T. Togawa, HMM-based estimation of menstrual cycle from skin temperature during sleep, in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* (2008) pp. 1635–1638.

[13] F. Chollet, Keras https://github.com/fchollet/keras, (2015).

[14] A. Fazleabas, M. Segraves and F. Khan-Dawood, Evaluation of salivary and vaginal electrical resistance for determination of the time of ovulation., *International journal of fertility* **35**(2) (1989) 106–111.

[15] G. Freundl, M. Bremme, P. Frank-Herrmann, S. Baur, E. Godehardt and U. Sottong, The cue fertility monitor compared to ultrasound and lh peak measurements for fertile time ovulation detection, *Advances in contraception* **12**(2) (1996) 111–121.

[16] T. W. Hilgers, F. G. E. ABRAHAM and D. CAVANAGH, Natural family planning i. the peak symptom and estimated time of ovulation., *Obstetrics & Gynecology* **52**(5) (1978) 575–582.

[17] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural computation* **9**(8) (1997) 1735–1780.

[18] J. Lafferty, A. McCallum and F. C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in *ICML* (2001) p. 282289.

[19] S. Lawrence, C. L. Giles, A. C. Tsoi and A. D. Back, Face recognition: A convolutional neural-network approach, *Neural Networks, IEEE Transactions on* **8**(1) (1997) 98–113.

[20] B. Lessey, L. Damjanovich, C. Coutifaris, A. Castelbaum, S. Albelda and C. Buck, Integrin adhesion molecules in the human endometrium. correlation with the normal and abnormal menstrual cycle., *Journal of Clinical Investigation* **90**(1) (1992) p. 188.

[21] P. B. Miller and M. R. Soules, The usefulness of a urinary lh kit for ovulation prediction during menstrual cycles of normal women., *Obstetrics & Gynecology* **87**(1) (1996) 13–17.

[22] K. Moghissi, Prediction and detection of ovulation., *Fertility and sterility* **34**(2) (1980) 89–98.

[23] G. Mulcaire-Jones, R. J. Fehring, M. Bradshaw, K. Brower, G. Lubega and P. Lubega, Couple beads: An integrated method of natural family planning, *The Linacre Quarterly* **83**(1) (2016) 69–82.

[24] J. Queenan, G. O'Brien, L. Bains, J. Simpson, W. Collins and S. Campbell, Ultrasound scanning of ovaries to detect ovulation in women., *Fertility and sterility* **34**(2) (1980) 99–105.

[25] A. Rasmus, M. Berglund, M. Honkala, H. Valpola and T. Raiko, Semi-supervised learning with ladder networks, in *Advances in Neural Information Processing Systems* (2015) pp. 3532–3540.

[26] A. Rosenfeld, A. Azaria, S. Kraus, C. V. Goldman and O. Tsimhoni, Adaptive advice in automobile climate control systems, in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (2015) pp. 543–551.

[27] J. Royston and R. Abrams, An objective method for detecting the shift in basal body temperature in women, *Biometrics* (1980) 217–224.

[28] S. Sarawagi, The crf project: a java implementation (2004).

[29] A. Tversky and D. Kahneman, The framing of decisions and the psychology of choice, *Science* **211**(4481) (1981) 453–458.

[30] T. Weschler, Ovagraph: Free ovulation calendar and fertility bbt charting http://ovagraph.com, (2009).

[31] Z. Zhang, Y. Zhang, G. Liu, J. Tang, S. Yan and M. Wang, Joint label prediction based semi-supervised adaptive concept factorization for robust data representation, *IEEE Transactions on Knowledge and Data Engineering* (2019).

[32] X. Zhu, Semi-supervised learning literature survey (2005).