

Treatment management algorithm for natural frozen embryo transfer cycles using a real-time ovulation prediction machine learning model

Received: 7 January 2025

Accepted: 28 February 2026

Published online: 08 March 2026

Cite this article as: Moran E., Hourvitz A., Luz A. *et al.* Treatment management algorithm for natural frozen embryo transfer cycles using a real-time ovulation prediction machine learning model. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-42921-1>

Eden Moran, Ariel Hourvitz, Almog Luz, Nevo Itzhak, Rohi Hourvitz, Michal Youngster, Micha Baum, Ettie Maman & Zerifin Israel

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Treatment management algorithm for natural frozen embryo transfer cycles using a real-time ovulation prediction machine learning model

Eden Moran^{1*}, Ariel Hourvitz MD^{2,*}, Almog Luz¹, Nevo Itzhak¹, Rohi Hourvitz¹, Michal Youngster MD², Micha Baum MD^{3,4} and Etti Maman MD^{3,4}

*Equal Contribution

¹ FertilAI, Ramat Gan Israel

² Shamir Medical Center, In Vitro Fertilization Unit, Department of Obstetrics and Gynecology, Zerifin Israel, Faculty of Medical & Health Sciences, Tel Aviv University, Tel-Aviv, Israel

³ Sheba Medical Center In Vitro Fertilization Unit, Department of Obstetrics and Gynecology, Ramat Gan Israel, Faculty of Medical & Health Sciences, Tel Aviv University, Tel-Aviv, Israel

⁴ Herzliya Medical Center, In Vitro Fertilization Unit, Herzliya Israel

Corresponding author: Etti Maman MD, ettiemaman@gmail.com

Article type: Cohort study

Abstract

Our aim was to develop an AI-based NC-FET Treatment Management Algorithm (NTMA), decision-support system that predicts ovulation in real time to manage and optimize natural frozen embryo transfer cycle (NC-FET) scheduling.

The algorithm was developed using a “teacher-student” machine learning approach and was trained on a total of 3,975 labeled NC-FET, including 3,432 training cycles and 543 test cycles. A second test group included 166 documented ovulation cycles (documented follicular rupture and LH surge in two consecutive days of ultrasound scans).

The algorithm showed high ovulation detection accuracy in both tests’ groups particularly one day before and the day of ovulation (95.4% and 94.6% in the labeled test group and 95.5% and 95.3% in the documented test group, respectively). Most influential predictive features of the algorithm included LH levels, the estrogen/progesterone ratio, and leading follicle size during the monitored test days. The NTMA yielded 92.04% correct prediction ⁷ in identifying ovulation with an average of 3.1 tests per cycle.

We propose an AI-based algorithm for the complete management of a NC-FET. The algorithm shows high accuracy in predicting the time of ovulation and therefore can serve as a useful decision support tool for clinicians in their daily practice. Prospective studies are warranted to validate these results.

Keywords:

Artificial Intelligence, Machine Learning, Ovulation Determination, Natural Cycle, Frozen Embryo Transfer

ARTICLE IN PRESS

Introduction

The introduction of vitrification and advancements in cryopreservation techniques have led to higher embryo survival rates and, consequently, more available embryos [1]. As a result, the number of frozen embryo transfer (FET) cycles has increased in recent years [2]. Among many fertility physicians, one of the preferred methods is time-frozen embryo transfer based on the natural cycle (NC-FET) [3, 4]. NC-FET is associated with lower rates of early pregnancy loss [5, 6] and reduced obstetric complications, such as hypertensive disorders of pregnancy and large-for-gestational-age infants, compared to cycles involving artificial endometrial preparation [4, 7, 8]. Additionally, a recent systematic review and meta-analysis by Wu et al.[9] reported increased live birth rates (LBRs) in NC-FET cycles compared to artificial FET cycles.

NC-FET requires close monitoring of the natural menstrual cycle to identify the timing of ovulation. Currently, the method considered by many as the most accurate method in NC-FET is the use of a combination of blood tests (Estrogen, Progesterone, and LH) and ultrasound monitoring of the growth and rupture of the leading follicle during ovulation [10-12]. This method requires patients to undergo several follow-up tests throughout the month to accurately monitor ovulation and consequently requires extensive expertise and time on the part of the physician.

In recent years, artificial intelligence (AI) technologies have been introduced in the field of reproduction, offering a range of benefits that could revolutionize the way we approach fertility treatments [13]. Several recently published studies utilized machine learning technologies in the field of reproduction including predicting IVF outcomes [14], optimizing ovarian stimulation [15, 16], sperm selection [17], trigger day [18-20], IUI (Intrauterine Insemination) scheduling [21, 22], workload balancing [23] embryo selection [24] and

euploidy prediction [25]. In a recent study, we examined the accuracy of AI-based ovulation prediction model on the results of NC-FET cycles. We showed that the pregnancy rate in NC-FET cycles, where the transfer was performed according to the ovulation date predicted by the model (matched cycles), was 9% higher than in cycles without such a match between the model and the physician's ovulation time determination [26].

Based on the model developed in a previous study, we introduce in this study a novel decision-support AI software designed to predict ovulation in real-time and fully manage the treatment of a natural frozen embryo transfer cycle. The algorithm supports the treatment cycle from its early stages, schedules the necessary tests, and ultimately predicts ovulation, thereby determining the optimal date for embryo transfer.

ARTICLE IN PRESS

Results

In this section, we first present the performance of the real-time ovulation day prediction model used by the NC-FET Treatment Management Algorithm (NTMA), followed by the evaluation of the NTMA itself. As detailed in the Materials and Methods section, the real-time prediction model was trained on data labeled according to a previously published ovulation prediction model [26], using the “student-teacher” approach [27]. The real-time ovulation day prediction model was trained on instances comprising one or two test days, with features derived from these days.

Model Performance

The real-time ovulation day prediction model’s performance on each test set is represented as a confusion matrix (actual vs predicted relative-to-ovulation days). Figure 1A shows the results on the Labeled Dataset, and Figure 1B presents the results on instances on the Documented Ovulation Dataset. In those figures, the model shows strong predictive power in two key ranges: far from ovulation (≤ -6) and close to ovulation (-1 or 0). For the far-from-ovulation predictions (≤ -6), we observed improved model performances because the test days are early enough to capture patterns where ovulation is not imminent, allowing for more accurate predictions that align with these distant days. On the other hand, the model also performs well around -1 and 0, as these ranges correspond to days immediately before or on the ovulation day itself. During these days, the biological signals, such as hormonal changes, are most prominent, providing the model with the information needed to make accurate predictions. As a result, the combination of early test days for distant predictions and clear hormonal patterns near ovulation leads to strong performance across these ranges.

Figure 1A: Model performance on the Labeled Dataset displayed as a confusion matrix

1A

Predicted	≤ -6	356 16.78%	47 2.22%	19 0.90%	5 0.24%	1 0.05%	0 0.0%	0 0.0%	0 0.0%	1 0.05%	429 82.98% 17.02%
	-5	19 0.90%	14 0.66%	12 0.57%	3 0.14%	0 0.0%	1 0.05%	0 0.0%	0 0.0%	0 0.0%	49 28.57% 71.43%
	-4	21 0.99%	52 2.45%	67 3.16%	35 1.65%	0 0.0%	1 0.05%	0 0.0%	0 0.0%	1 0.05%	177 37.85% 62.15%
	-3	3 0.14%	15 0.71%	67 3.16%	137 6.46%	37 1.74%	0 0.0%	0 0.0%	0 0.0%	6 0.28%	265 51.70% 48.30%
	-2	0 0.0%	0 0.0%	2 0.09%	33 1.56%	194 9.15%	14 0.66%	1 0.05%	0 0.0%	0 0.0%	244 79.51% 20.49%
	-1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	11 0.52%	501 23.62%	12 0.57%	0 0.0%	0 0.0%	524 95.61% 4.39%
	0	1 0.05%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	8 0.38%	334 15.75%	21 0.99%	2 0.09%	366 91.26% 8.74%
	+1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 0.28%	45 2.12%	4 0.19%	55 81.82% 18.18%
	≥ +2	4 0.19%	0 0.0%	2 0.09%	1 0.05%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 0.24%	12 41.67% 58.33%
	Column Summation	404 88.12% 11.88%	128 10.94% 89.06%	169 39.64% 60.36%	214 64.02% 35.98%	243 79.84% 20.16%	525 95.43% 4.57%	353 94.62% 5.38%	66 68.18% 31.82%	19 26.32% 73.68%	2121 77.93% 22.07%
	≤ -6	-5	-4	-3	-2	-1	0	+1	≥ +2	Row Summation	
	Actual										

Figure 1B: Model performance on the Documented Ovulation Dataset

B

Predicted	≤ -6	145 16.11%	10 1.11%	3 0.33%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	158 91.77% 8.23%
	-5	1 0.11%	18 2.00%	5 0.56%	1 0.11%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	25 72.00% 28.00%
	-4	3 0.33%	13 1.44%	42 4.67%	6 0.67%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	64 65.62% 34.38%
	-3	2 0.22%	1 0.11%	10 1.11%	104 11.56%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	117 88.89% 11.11%
	-2	0 0.0%	1 0.11%	0 0.0%	13 1.44%	31 3.44%	5 0.56%	1 0.11%	0 0.0%	0 0.0%	51 60.78% 39.22%
	-1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.11%	170 18.89%	3 0.33%	0 0.0%	0 0.0%	174 97.70% 2.30%
	0	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.33%	285 31.67%	5 0.56%	0 0.0%	293 97.27% 2.73%
	+1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	10 1.11%	2 0.22%	2 0.22%	14 14.29% 85.71%
	$\geq +2$	2 0.22%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.22%	4 50.00% 50.00%
	Column Summation	153 94.77% 5.23%	43 41.86% 58.14%	60 70.00% 30.00%	124 83.87% 16.13%	32 96.88% 3.12%	173 95.51% 4.49%	299 95.32% 4.68%	7 28.57% 71.43%	4 50.00% 50.00%	900 99.78% 0.22%
	≤ -6	-5	-4	-3	-2	-1	0	+1	$\geq +2$	Row Summation	
	Actual										

Figure 1A and 1B: Model performance. Columns represent instances' actual classes, relative to the instance's latest test day. Rows represent the model-predicted classes. The number and percentage of total occurrences for actual vs. prediction pairs are presented in each cell. Percentages were normalized cumulative sum of 100%. The bottom row and right column highlight the accuracy in green and the errors in red, respectively. The number of occurrences is positioned at the top of each cell. The bottom-right cell signifies the overall model accuracy. In Figures 1A and 1B, the model shows strong predictive power in two key ranges: far from ovulation (≤ -6) and close to ovulation (-1 or 0).

Feature Importance

The top ten influential features, used by the model to determine ovulation day, are presented in Figure 2. The terms “first” and “second” test days refer to their chronological order within each instance. In cases where only a single test day is used, it is considered both the first and second day. Overall, features associated with the second test day showed greater influence than those linked to the first test day. The most influential features included LH, the ratio between estrogen (E2) and progesterone (P4), P4, and E2 on the instance’s second test day. The other features used by the model were: E2 value on the instance’s first test day, leading follicle size on the instance’s second test day, LH value on the instance’s first test day, average LH difference per day between the instance’s first and second test days, endometrial thickness value on the instance’s second test day, and E2 value on the instance’s first test day. The LH value on the second test day had higher importance compared to the LH value on the first test day. This is consistent with the physiology of the peri-ovulatory LH rise. Because the second test day is typically closer to ovulation, the model naturally assigns it greater weight. In contrast, the LH value on the first test day contributes less, as it is usually obtained earlier in the follicular phase, when LH levels are lower, more stable, and therefore less informative.

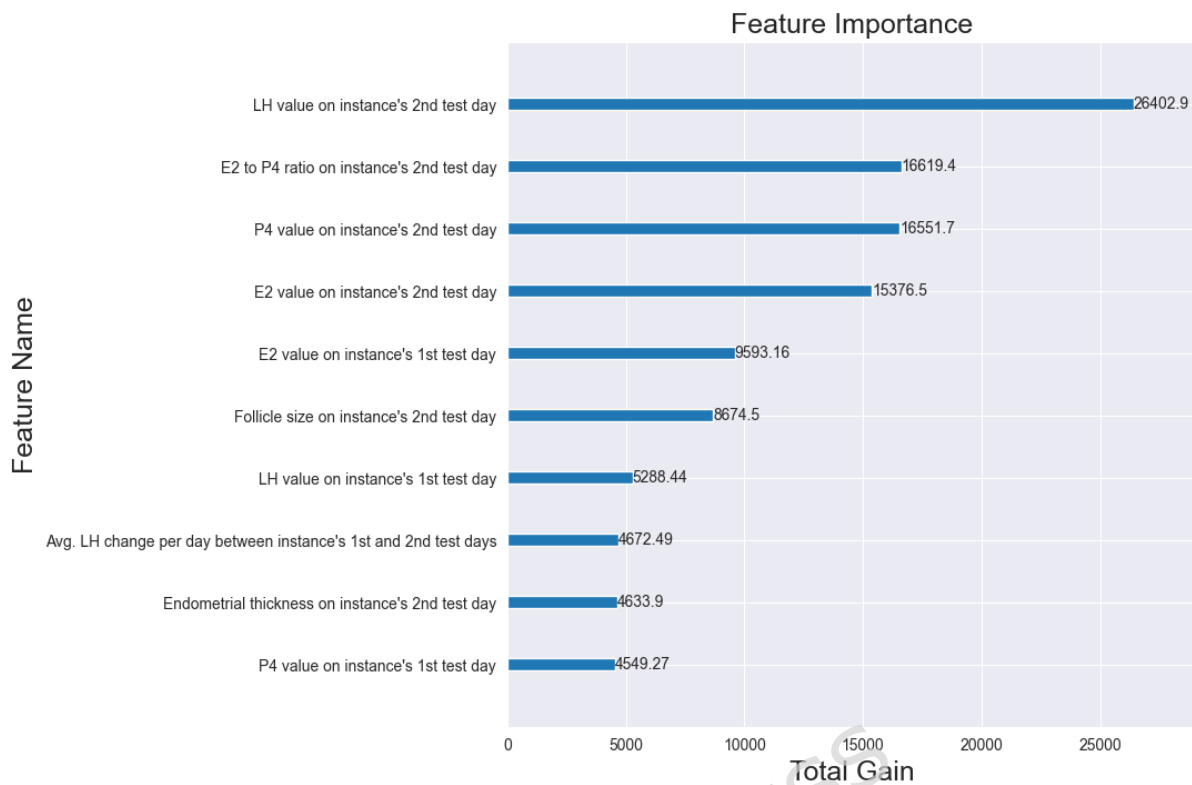


Figure 2: Real-time ovulation day prediction model's most influential features. For each feature, the relative contribution is presented as the total gain. Summary of the real-time ovulation day prediction model's ten most important features by total gain of each feature.

NTMA Performance

Table 1 shows the results for NTMA performance. The algorithm's performance was estimated using a statistical simulation-based approach that incorporates cycle-day ovulation probabilities, model prediction accuracy, and treatment-management logic (see Supplementary Part 3: NTMA's Performance Evaluation). The following metrics were used to evaluate performance.:

- **Correct Prediction:** The algorithm correctly determined the ovulation day
- **Incorrect Prediction:** The algorithm, based on the prediction of the real-time ovulation prediction model, was not able to correctly identify the ovulation day

- **No Prediction:** The algorithm could not accurately predict ovulation as the test was performed two or more days following ovulation
- **Number of tests performed**

Overall, the model demonstrates a 92.04% correct prediction rate, with a 1.16% no prediction rate (NTMA's Objective 1), and incorrect prediction rate of 6.79%. An average number of tests of 3.10 were performed. The NTMA approach reduces the number of blood tests and ultrasounds required. Table 1 highlights that the average number of tests using NTMA is 3.10, compared to an average of 3.54 [3.32-3.67] blood tests (excluding baseline tests) in the Labeled Dataset and average of 3.48 [3.35-3.61] number of ultrasounds in the Labeled Dataset (excluding baseline tests). Comparison of monitoring burden (number of blood tests and ultrasounds) is reported only for the Labeled Dataset. Specifically, the Documented Ovulation Dataset was created from cycles with two consecutive days of testing on the day before and the day of ovulation and was used solely for model validation; therefore, it does not represent routine clinical monitoring practices.

Table 1: Summary results of the NTMA.

Correct Prediction Rate (%)	92.04 [91.75-92.34]
Incorrect Prediction Rate (%)	6.79% [6.53–7.05]
No Prediction Rate (%)	1.16 [0.98-1.33]
Average Number of Tests	3.10 [3.09-3.11]

[]95% Confidence Interval

ARTICLE IN PRESS

Discussion

Embryo transfer in the natural cycle is currently considered the preferred method for frozen embryo transfer [3, 4]. In this novel study, we present an AI-based NC-FET treatment management algorithm that uses an ovulation determination model to schedule embryo transfer time. With an average of 3.1 follow-up test days per cycle, the model successfully determined the ovulation date in 92.04% of cases. This new algorithm can fully manage frozen cycles and serve as a decision support system for NC-FET.

In the previous article [26], we demonstrated that, theoretically, pregnancy rates could be improved when the ovulation date was matched with the model's recommendation. In this article, however, we present an algorithm trained on the results of the previous model, capable of achieving the same level of prediction accuracy while completely managing treatment. The real time ovulation prediction algorithm was developed using the teacher-student method [27]. The ovulation labeling model (teacher) was trained on hundreds of cycles in which ovulation was determined by fertility experts. Afterward, the real-time (student) algorithm was trained on several thousand cycles where the ovulation date was labeled by the ovulation labeling model. This approach allows models to be trained on a large number of treatment cycles without the need for manual labeling. However, because the real time algorithm was trained on labels generated by the teacher model rather than directly on expert annotations for all cycles, any residual inaccuracies of the teacher may also be reflected in the student. This reliance on model generated labels should be considered when interpreting the results. Developing a treatment management algorithm that attempts to optimize the treatment by selecting the optimal test days using retrospective data proved challenging. The test days in the data were selected by the

attending physicians and rarely matched the algorithm's recommendation; therefore, the algorithm performance could not be tested directly using the available treatment cycles and the statistical method described in the materials and methods was used. This method has the advantage of being capable of analyzing the performance of the treatment management algorithm just by using statistical information about the probability of a patient to ovulate on each cycle day, the different error probabilities of the ovulation determination model and the logic of the treatment management algorithm. The method allowed us to quickly repeat through multiple versions of the treatment management algorithm logic and select the optimal one.

The accuracy of ovulation determination by the model was tested using two datasets: the ovulation labeling model-Labeled Dataset and the Documented Ovulation Dataset. The advantages of the model-Labeled Dataset include its larger size and random selection. In contrast, the Documented Ovulation Dataset is smaller, however ovulation there was confirmed using two "gold standard" criteria: evidence of follicular rupture and an LH surge [3, 4, 28, 29]. It is important to note that there is no standard definition for an LH surge; to address this, we define an LH surge as $LH > 17 \text{ IU/L}$ [30] with a documented decrease in LH level the following day. The fact that accuracy of ovulation prediction model in the Labeled Dataset was similar to that in a documented ovulation test set (up to 95.4 and up to 95.5%, respectively) strongly suggests that the algorithm is highly accurate in determining the true ovulation time. Nevertheless, additional prospective studies are needed to validate the model's impact on NC-FET timing and clinical outcomes.

The development of this algorithm raises some clinical questions. One might question the necessity of a natural ovulation detection model when ovulation can be timed through the

administration of an hCG trigger and a scheduled ovulation and embryos transfer (modified NC-FET). While the question of whether frozen embryo transfer should occur in a fully natural cycle versus a modified natural cycle lies outside the primary focus of this article, it remains without a definitive answer. Some studies indicate no significant difference in success rates between modified cycles and natural cycles, which generally necessitate more frequent clinic visits and monitoring [3, 31, 32]. Conversely, other studies suggest an advantage for the fully natural cycle approach [33-35]. Variability in reported success rates may be attributed to factors such as the precise timing of hCG administration, which, if misaligned, can impair endometrial receptivity or lead to suboptimal corpus luteum function. Future approaches may benefit from a hybrid model that uses the strengths of both strategies—preserving the natural cycle where feasible, while also providing the option to control ovulation timing through scheduled hCG administration when needed. The most effective pathway to achieve this balance likely lies in the development of AI-driven algorithms. Such models could integrate serial hormonal measurements, ultrasound findings and scheduling constraints to propose individualized treatment strategies. In particular, they could recommend hCG triggering when the predicted transfer day based on spontaneous ovulation is not feasible due to patient or clinic constraints.

The algorithm presented in this article is designed to identify ovulation, and thereafter the physician can determine embryo transfer time according to embryo age, making it suitable for all types of frozen embryos. It is important to emphasize that the proposed NTMA is specifically designed for frozen embryo transfer cycles. This includes that if the ovulation date identified by the algorithm has passed by one day, frozen embryos can still be transferred.

As expected, LH, estrogen and progesterone were the major factors used by the model for ovulation prediction. Interestingly, one of the influential features affecting the algorithm performance was the Estrogen/Progesterone ratio on the prediction day. The finding of the Estrogen/Progesterone ratio as an influential factor is in agreement with previous studies [33, 34] that consider the decrease in the ratio of Estrogen/Progesterone metabolites in urine to correspond to the day of luteal transition in ovulatory women. It is important to note that the Estrogen/Progesterone ratio is not in routine daily clinical use, probably because of its relative complexity. However, the ability of AI algorithms to quickly process data, increases the possibility of using this feature and many others, such as the ratios and derivative between different hormones in consecutive tests. As can be seen in Figure 1, the model uses these additional features to increase the accuracy of ovulation determination and optimal FET timing.

While this retrospective study highlights the NTMA's clinical utility, prospective validation is needed to assess its impact on pregnancy outcomes. The study does not include pregnancy rates, which is essential to fully assess the algorithm's clinical utility and effectiveness in improving FET success. However, the algorithm was trained on labeled data, and the labeling model was retrospectively tested on NC-FET cycles, showing a 9% increase in pregnancy rates in cycles where ovulation matched the model's prediction[26]. Another drawback of this study is that AI algorithms are susceptible to bias, which can arise from data selection or labeling processes. Although the algorithm was tested using two datasets to minimize this risk, some bias may still persist.

In conclusion, to our knowledge, this is the first highly accurate and useful AI-based algorithm designed for the full management of an IVF protocol. The performance achieved by the NTMA model for ovulation detection and management of NC-FET cycles are

reassuring. The high accuracy is attributed to the capability of the algorithm to analyze effectively and accurately multiple and complex features. We believe that this algorithm and others that will follow can serve clinicians in their daily practice, improving clinical outcomes while increasing their clinical activity. We anticipate the growing availability of this and other AI algorithms will greatly impact the accessibility, cost and success of fertility treatments.

ARTICLE IN PRESS

Materials and Methods

The study was approved by the Institutional Review Board of Herzliya Medical Center (HMC-0008-21). Due to the retrospective nature of the study, informed consent was waived by the Institutional Review Board of Herzliya Medical Center. We confirm that the study was performed in accordance with relevant guidelines and regulations.

The "Student-Teacher" Approach

The Ovulation Labeling Model ("Teacher"): The teacher model that was previously described [26], was trained and evaluated on NC-FET cycles between 2018-2023. Briefly, the model was first trained on 399 cycles, where ovulation timing was determined by a majority decision from three reproductive endocrinology and infertility independent specialists (REI). The model was evaluated on 101 cycles where ovulation was determined by the same majority method and 101 cycles with documented ovulation. The model achieved high accuracy in these test groups, 93.85% and 92.89% matching rates with the REI and Documented Ovulation Datasets, respectively. We used this "Teacher" prediction on multiple instances from a cycle to determine the ovulation day, relying on aggregated information from the cycle's data, in a retrospective way, for a final decision (see Supplementary Material Part 1).

The Real-Time Ovulation Day Prediction Model ("Student"): The real-time ovulation day prediction model stands at the core of the treatment management algorithm and is used to predict the probability of a patient ovulating. The real-time ovulation day prediction model was trained using a "student-teacher" approach [27]. The teacher algorithm was used to label the precise ovulation day in NC-FET cycles, which were then used to train the real-time ovulation day prediction model ("student") accurately. Achieving an accurate real-time model required extensive training on a large dataset, which would not have been feasible

with only a few hundred manually labeled cycles by fertility experts, as was done with the teacher model. The primary distinction between the ovulation labeling model and the real-time ovulation day prediction model lies in how they are applied. In contrast to the teacher model, the real-time ovulation day prediction model is optimized for real-time prediction, considering only the latest instance to provide immediate, actionable insights without the context of the entire cycle. This approach enabled the student model to accurately predict ovulation several days in advance, unlike the ovulation labeling model (teacher model).

Natural FET Treatment Management Algorithm (NTMA): In this study, we propose a treatment management algorithm for spontaneous natural FET cycles that fully manages the cycle. After the Real-Time Ovulation Day Prediction Model predicts the day relative to ovulation, a set of rules (the NTMA) is applied explained further below. At first, the NTMA recommends when the patient should come for the initial test day. Then, after each test day, the algorithm considers the patient's likelihood and timing of ovulation, provided by the machine learning model, and recommends if and when another test day should be performed, or if an accurate prediction of ovulation can be made. Further details regarding the logic of the NTMA are provided in the Methods section.

Datasets

Labeled Dataset

To train and evaluate the real-time ovulation day prediction model ("Student"), we used the ovulation day determination model ("Teacher") to generate the dataset. The initial dataset is from Herzliya Medical Center, Israel, and consisted of 5,162 natural FET cycles performed between August 2018 and May 2024. Filtering out cycles with uncertain class prediction (i.e., the model could not accurately predict when ovulation will occur given the existing test

results) of the teacher algorithm [26] reduced the number to 4,361 cycles. Removing cycles with expert prediction, utilized in the teaching process of the teacher algorithm, resulted in the remaining 3,975 cycles.

After the teacher model completed ovulation determination on the cycles, we used the resulting dataset to train and evaluate the real-time ovulation day prediction model (“Student”). The data obtained for each cycle consisted of baseline characteristics, including the patient's age at the time of embryo freezing and embryo transfer, body mass index (BMI), and cycle characteristics, including the leading follicle’s size, endometrial thickness, blood test results for estradiol, progesterone, and LH levels, and day of embryo transfer. Table 2A shows patients’ characteristics of the real-time ovulation day prediction model training and test cycles. The table presents baseline patient characteristics only. Time-dependent variables, such as hormone levels and embryo transfer day, were not included because they vary across test days and are analyzed dynamically within the model.

Table 2A: Baseline characteristics of patients included in the Labeled Dataset

	<i>Training sets*</i>	<i>Test set</i>	<i>P-value</i>
Number of cycles	3,432	543	
Mean patient age [y]** (Mean±SD)	36.44 ± 5.80	36.33 ± 5.93	0.58
BMI [kg/m ²] (Mean±SD)	24.27 ± 5.66	24.68 ± 5.51	0.05
Smokers (%)	7.66	7.37	0.65
Ovulation day as marked by as Ovulation Labeling Model (Mean±SD)	16.00 ± 5.08	16.34 ± 4.95	0.05
Endometrial width [mm]*** (Mean±SD)	9.55 ± 1.99	9.59 ± 2.14	0.95

Presented as mean ± standard deviation (SD)

*Including the train and validation sets that were used to learn the model

**At the time of embryo transfer

***As measured on the last test day

Documented Ovulation Dataset

The Documented Ovulation Dataset included 166 cycles from 2019-2024 and was used only to evaluate the real-time ovulation day prediction model performance (a test set only).

Inclusion criteria were: two consecutive days of blood and ultrasound tests, documented follicular rupture of a single primary follicle measuring ≥ 17 mm, and documented LH surge, defined as $LH \geq 17$ IU on the initial day [30], followed by a decrease on the next day. The LH cut-off of 17 IU/L was used solely for defining ovulation in the Documented Ovulation Dataset and does not affect model training or feature-importance estimation. The student model uses continuous LH values rather than any threshold.

Table 2B shows patients' characteristics of the training cycles of the Labeled Dataset compared with the Documented Ovulation Dataset.

Table 2B: Baseline characteristics of patients included in the Labeled Dataset compared with the Documented Ovulation Dataset

	<i>Training sets*</i>	<i>Documented ovulation set</i>	<i>P-value</i>
Number of cycles	3,432	166	
Mean patient age [y]** (Mean \pm SD)	36.44 \pm 5.80	36.60 \pm 6.23	0.63
BMI [kg/m ²] (Mean \pm SD)	24.27 \pm 5.66	24.19 \pm 4.23	0.28
Smokers (%)	7.66	4.82	0.10
Ovulation day (Mean \pm SD)	16.00 \pm 5.08	17.96 \pm 6.32	0.84

Endometrial width [mm]*** (Mean±SD)	9.55 ± 1.99	9.57 ±1.94	0.82
--	-------------	------------	------

Presented as mean ± standard deviation (SD)

*Including the train and validation sets that were used to learn the model

**At the time of embryo transfer

***As measured on the last test day

Instances Creation

To train, and evaluate a real-time ovulation prediction model, multiple instances were created from each cycle in the train, validation, and test sets, where each instance included a different combination of test days from that cycle. The instances included all single test days and any combination of two test days that are at most four days apart. Training the model with additional test days or test days further apart did not improve its performance. The target class for each instance is determined by calculating the difference between the instance's last test day and the ovulation day. The nine ovulation classes are defined as six or more days before (≤ -6), five days before (-5), four days before (-4), three days before (-3), two days before (-2), one day before (-1), the same day as (0), a day after (+1), or two or more days after ($\geq +2$) the ovulation day. For example, Table 3 shows an example of a cycle with ovulation on the 13th day, while the blood tests were performed on the 3rd, 10th, 12th, and 14th days. As can be seen from the example in Table 3, seven possible instances can be generated.

Table 3. An example of NC-FET

Status			T							T		T	O	T	
Day	1	2	3	4	5	6	7	8	9	10		12	13	14	15

Possible combinations of days to generate the instances

#Instance	1st Test Day	2nd Test Day	Ovulation Class
1	-	3	≤ -6
2	-	10	-3
3	10	12	-1
4	10	14	1
5	-	12	-1
6	12	14	1
7	-	14	1

Table 3: An example of NC-FET over the cycle days and that status at each day, where T with a blue-colored background represents a test day and O with a purple-colored background represents the ovulation day. The seven possible combinations to create the instances for the presented NC-FET are presented.

For the Labeled Dataset, these instances were randomly split by the patient unique ID, into the train (60%), validation and calibration (25%), and test (15%) sets, ensuring that each patient appeared in only one set—thereby preventing any data leakage, while maintaining the same class distribution between the sets. Table 4 summarizes the dataset groups.

Table 4: Summary of the number of cycles and instances per dataset

Dataset	Cycles				Instances			
	Total	Train	Validation	Test	Total	Train	Validation	Test
Labeled Dataset	3,975	2,465	967	543	21,369	14,360	4,888	2,121
Documented Ovulation	166			166	900			900
All datasets	4,141			709	22,269			3,021

The model performance results for the Labeled Dataset included 2,121 instances: a single test day conducted between the 6th and 10th cycle days, encompassing 481 instances from 391 distinct cycles, and two test days conducted on the first test day following the 6th cycle day, encompassing 1,640 instances from 509 distinct cycles.

Model Design

Based on instances that were extracted, described in the data preparation section, a machine-learning model was trained. The model was trained on the train set to be able to classify, given an instance, the prediction of the ovulation class ($\leq -6, -5, -4, -3, -2, -1, 0, +1, \geq +2$). The features included information related to each instance consisting of baseline characteristics, alongside cycle-specific measurements derived from blood and ultrasound tests. These included both raw clinical values and calculated features reflecting hormonal and ultrasound dynamics across test days. A comprehensive description of all features is provided in Supplementary Table S1, and the extent of missing data for each feature is summarized in Supplementary Table S2. Because the model was designed to flexibly support

cases both with and without ultrasound data, an appropriate handling of missing values was essential.

ARTICLE IN PRESS

XGBoost classifier, a machine learning model, was selected due to the combination of its high performance and its powerful handling of missing values. Multi-class calibration using one versus rest methodology was performed as a part of the models' development (see Supplementary Material Part 2). The grid search cross-validation method was applied and evaluated on the validation and calibration set to optimize the model hyperparameters, such as learning rate and maximal depth. After the hyperparameters optimization, the final model was trained on the training and validation sets. During the prediction stage, the ovulation class with the greatest likelihood, per instance, was selected as the model's final prediction.

Natural FET Treatment Management Algorithm (NTMA)

The objective of the Natural FET Treatment Management Algorithm (NTMA) is to autonomously manage NC-FET cycles. Embryo transfer day was calculated as the day of ovulation + plus the embryo's developmental stage [3, 4].

This algorithm includes recommending the ideal timing for test days until an accurate prediction of ovulation can be made and a reliable suggestion for the optimal transfer day can be provided.

The NTMA's objectives are as follows:

- Objective 1: Transfer on the correct day according to the patient's ovulation
- Objective 2: Reduce the number of tests required per cycle

NTMA's logic

The algorithm's logic considers the probability of natural ovulation to occur on each day following the last test day using the real-time ovulation prediction model. We present here a simplified logic that represents the general principles of the NTMA:

- Initial test on the 8th day of the menstrual cycle
- Following each test day:
 - If the model predicts ovulation in 6 or more days, schedule the next test in 4 days.
 - If the model predicts that the ovulation occurred yesterday, today, or will occur tomorrow, schedule a transfer, according to the predicted ovulation day
 - If the model determines that ovulation occurred two or more days ago (No prediction), suggest cycle cancellation, since precise ovulation day cannot be determined
 - If neither condition is met, schedule another test one day before the predicted ovulation

NTMA's evaluation***Statistical evaluating approach:***

Measuring the performance of the NTMA without a prospective study was challenging. The algorithm's accuracy relied on scheduling blood tests on specific days, while our dataset consisted of blood test data on days chosen by attending physicians. To address this, we use a novel statistical approach to calculate the algorithm's performance metrics. This approach was originally proposed [21] and factored in the likelihood of a patient ovulating on each cycle day, the prediction accuracy of the real-time ovulation day prediction model on natural cycles, and the logic of the treatment management algorithm. This was done for varying ovulation cycle days, according to the distributions observed in the data. A simplified pseudo-code for this calculation is presented in Supplementary Material Part 3. The results are presented with 95% confidence intervals, calculated using the Wilson Score method and the Monte Carlo technique. Each scenario was sampled 30 times to determine the mean, standard deviation, and confidence intervals for all evaluation metrics.

Additional Information:

Funding Statement: This study was supported by FertilAI LTD

Conflict of interest: A.H, R.H, A.L, M.B and E.M are shareholders and board members of FertilAI. N.I and E.M are employees of FertilAI. M.Y reports no conflict of interest.

Authors' contribution

Eden Moran - Conception and design, acquisition of data, analysis and interpretation of data, approval of the version to be published, agree to be accountable for all aspects of the work

Ariel Hourvitz - Conception and design, analysis and interpretation of data, revising the article, approval of the version to be published, agree to be accountable for all aspects of the work

Almog Luz -Conception and design, acquisition of data, analysis and interpretation of data, drafting the article and revising, approval of the version to be published, agree to be accountable for all aspects of the work.

Nevo Itzhak - Conception and design, acquisition of data, analysis and interpretation of data, drafting the article and revising, approval of the version to be published, agree to be accountable for all aspects of the work

Rohi Hourvitz - Conception and design, analysis and interpretation of data, approval of the version to be published, agree to be accountable for all aspects of the work

Michal Youngster - Conception and design, analysis and interpretation of data, revising the article, approval of the version to be published, agree to be accountable for all aspects of the work

Micha Baum - Conception and design, analysis and interpretation of data, revising the article, approval of the version to be published, agree to be accountable for all aspects of the work

Ettie Maman - Conception and design, analysis and interpretation of data, drafting the article and revising, approval of the version to be published, agree to be accountable for all aspects of the work

Data Availability Statements:

Data is provided within the manuscript or supplementary information files.

The data underlying this article will be shared on reasonable request to the corresponding author.

Code availability

The code used to develop and evaluate the NTMA is publicly available at

<https://github.com/fertilai/nc-fet-simulator/>

ARTICLE IN PRESS

References

1. Venetis, C.A., *Pro: Fresh versus frozen embryo transfer. Is frozen embryo transfer the future?* Hum Reprod, 2022. **37**(7): p. 1379-1387.
2. Roelens, C. and C. Blockeel, *Impact of different endometrial preparation protocols before frozen embryo transfer on pregnancy outcomes: a review.* Fertil Steril, 2022. **118**(5): p. 820-827.
3. Mackens, S., et al., *Frozen embryo transfer: a review on the optimal endometrial preparation and timing.* Hum Reprod, 2017. **32**(11): p. 2234-2242.
4. Roelens, C., et al., *Artificially prepared vitrified-warmed embryo transfer cycles are associated with an increased risk of pre-eclampsia.* Reprod Biomed Online, 2022. **44**(5): p. 915-922.
5. Vinsonneau, L., et al., *Impact of endometrial preparation on early pregnancy loss and live birth rate after frozen embryo transfer: a large multicenter cohort study (14 421 frozen cycles).* Human Reproduction Open, 2022. **2022**(2): p. hoac007.
6. von Versen-Hoynck, F. and G. Griesinger, *Should any use of artificial cycle regimen for frozen-thawed embryo transfer in women capable of ovulation be abandoned: yes, but what's next for FET cycle practice and research?* Hum Reprod, 2022. **37**(8): p. 1697-1703.
7. Maheshwari, A., et al., *Is frozen embryo transfer better for mothers and babies? Can cumulative meta-analysis provide a definitive answer?* Hum Reprod Update, 2018. **24**(1): p. 35-58.
8. Moreno-Sepulveda, J., J.J. Espinos, and M.A. Checa, *Lower risk of adverse perinatal outcomes in natural versus artificial frozen-thawed embryo transfer cycles: a systematic review and meta-analysis.* Reprod Biomed Online, 2021. **42**(6): p. 1131-1145.
9. Wu, H., et al., *Endometrial preparation for frozen-thawed embryo transfer cycles: a systematic review and network meta-analysis.* J Assist Reprod Genet, 2021. **38**(8): p. 1913-1926.
10. Mumusoglu, S., et al., *Preparation of the Endometrium for Frozen Embryo Transfer: A Systematic Review.* Front Endocrinol (Lausanne), 2021. **12**: p. 688237.
11. Erden, M., et al., *The LH surge and ovulation re-visited: a systematic review and meta-analysis and implications for true natural cycle frozen thawed embryo transfer.* Hum Reprod Update, 2022. **28**(5): p. 717-732.
12. Maman, E., et al., *Prediction of ovulation: new insight into an old challenge.* Sci Rep, 2023. **13**(1): p. 20003.
13. Letterie, G., *Artificial intelligence and assisted reproductive technologies: 2023. Ready for prime time? Or not.* Fertil Steril, 2023. **120**(1): p. 32-37.
14. Goyal, A., M. Kuchana, and K.P.R. Ayyagari, *Machine learning predicts live-birth occurrence before in-vitro fertilization treatment.* Sci Rep, 2020. **10**(1): p. 20925.
15. Letterie, G. and A. Mac Donald, *Artificial intelligence in in vitro fertilization: a computer decision support system for day-to-day management of ovarian stimulation during in vitro fertilization.* Fertil Steril, 2020. **114**(5): p. 1026-1031.
16. Letterie, G., A. MacDonald, and Z. Shi, *An artificial intelligence platform to optimize workflow during ovarian stimulation and IVF: process improvement and outcome-based predictions.* Reprod Biomed Online, 2022. **44**(2): p. 254-260.
17. McCallum, C., et al., *Deep learning-based selection of human sperm with high DNA integrity.* Commun Biol, 2019. **2**: p. 250.
18. Reuveny, S., et al., *An artificial intelligence-based approach for selecting the optimal day for triggering in antagonist protocol cycles.* Reprod Biomed Online, 2024. **48**(1): p. 103423.
19. Hariton, E., et al., *A machine learning algorithm can optimize the day of trigger to improve in vitro fertilization outcomes.* Fertil Steril, 2021. **116**(5): p. 1227-1235.

20. Fanton, M., et al., *An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation*. Fertil Steril, 2022. **118**(1): p. 101-108.
21. Youngster, M., et al., *Artificial intelligence in the service of intrauterine insemination and timed intercourse in spontaneous cycles*. Fertil Steril, 2023. **120**(5): p. 1004-1012.
22. Youngster, M., et al., *Intrauterine insemination timing models-LH can only take you so far*. J Assist Reprod Genet, 2024. **41**(7): p. 1843-1850.
23. Youngster, M., et al., *Optimizing workload balance using artificial intelligence*. Fertil Steril, 2024. **122**(1): p. 178-180.
24. Zaninovic, N. and Z. Rosenwaks, *Artificial intelligence in human in vitro fertilization and embryology*. Fertil Steril, 2020. **114**(5): p. 914-920.
25. Diakiw, S.M., et al., *An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos*. Reprod Biomed Online, 2022. **45**(6): p. 1105-1117.
26. Luz, A., et al., *Improved clinical pregnancy rates in natural frozen-thawed embryo transfer cycles with machine learning ovulation prediction: insights from a retrospective cohort study*. Sci Rep, 2024. **14**(1): p. 29451.
27. Hinton, G., *Distilling the Knowledge in a Neural Network*. . arXiv 2015. **preprint arXiv:1503.02531**.
28. Wetzels, L.C. and H.J. Hoogland, *Relation between ultrasonographic evidence of ovulation and hormonal parameters: luteinizing hormone surge and initial progesterone rise*. Fertil Steril, 1982. **37**(3): p. 336-41.
29. Marinho, A.O., et al., *Real time pelvic ultrasonography during the periovulatory period of patients attending an artificial insemination clinic*. Fertil Steril, 1982. **37**(5): p. 633-8.
30. Irani, M., et al., *Optimal parameters for determining the LH surge in natural cycle frozen-thawed embryo transfers*. J Ovarian Res, 2017. **10**(1): p. 70.
31. Groenewoud, E.R., et al., *What is the optimal means of preparing the endometrium in frozen-thawed embryo transfer cycles? A systematic review and meta-analysis*. Hum Reprod Update, 2017. **23**(2): p. 255-261.
32. Weissman, A., et al., *Spontaneous ovulation versus HCG triggering for timing natural-cycle frozen-thawed embryo transfer: a randomized study*. Reprod Biomed Online, 2011. **23**(4): p. 484-9.
33. Fatemi, H.M., et al., *Cryopreserved-thawed human embryo transfer: spontaneous natural cycle is superior to human chorionic gonadotropin-induced natural cycle*. Fertil Steril, 2010. **94**(6): p. 2054-8.
34. Montagut, M., et al., *Frozen-thawed embryo transfers in natural cycles with spontaneous or induced ovulation: the search for the best protocol continues*. Hum Reprod, 2016. **31**(12): p. 2803-2810.
35. Gao, D.D., et al., *Is Human Chorionic Gonadotropin Trigger Beneficial for Natural Cycle Frozen-Thawed Embryo Transfer?* Front Med (Lausanne), 2021. **8**: p. 691428.