Journal of Translational Medicine

RESEARCH

Open Access

Deep learning-based prediction of individualized Real-time FSH doses in GnRH agonist long protocols

Na Kong[1,2,3†], Yu Xia[5†], Zhilong Wang[2,3†], Hui Zhang[2,3], Liyan Duan[3], Yingchun Zhu[2,3], Chenyang Huang[2,3], Guijun Yan[3,7,8], Jie Mei[2,3*†], Wujun Li[4,5,6*†] and Haixiang Sun[1,2,7,8*†]

Abstract

Background Individualizing follicle-stimulating hormone (FSH) dosing during controlled ovarian stimulation (COS) is critical for optimizing outcomes in assisted reproduction but remains difficult due to patient heterogeneity. Most existing models are limited to static predictions of initial doses and do not support real-time adjustments throughout stimulation.

Methods We developed a deep learning model that integrates cross-temporal and cross-feature encoding (CTFE) to predict personalized daily FSH doses in patients undergoing COS using the GnRH agonist long protocol. A total of 13,788 IVF/ICSI cycles conducted between January 2018 and December 2020 were retrospectively analyzed. Women with baseline antral follicle counts between 7 and 30 were included. Data were randomly divided into training (n = 6761), validation (n = 2898), and test (n = 4135) sets. The model encodes both static (e.g., age, BMI, basic hormone levels) and dynamic (e.g., follicle development, hormone trends during COS) variables across stimulation days. Final dose predictions were generated using a K-nearest neighbor algorithm applied to low-dimensional latent representations derived from the deep encoder layers.

Results The CTFE model achieved a dose classification accuracy of 0.737 (± 0.004) and a weighted F1-score of 0.732 (± 0.005) on the test set. On key stimulation days 1 and 5, the CTFE model significantly outperformed traditional LASSO regression models (F1-score: 0.832 vs 0.699 on day 1; 0.817 vs 0.523 on day 5; $p < 0.001$). Prediction performance was maintained beyond day 13 using a sliding window mechanism, despite reduced data availability in longer stimulation cycles.

Conclusions This is the first study to apply a cross-temporal and cross-feature deep learning framework for daily, individualized FSH dose prediction across the full duration of COS. The model demonstrated superior performance over conventional approaches and offers a promising tool for standardizing COS management. Although currently

†Na Kong, Yu Xia and Zhilong Wang are co-first authors and contributed equally to this work.

†Jie Mei, Wujun Li and Haixiang Sun are contributed equally to this work.

*Correspondence:
Jie Mei
meijie560@163.com
Wujun Li
liwujun@nju.edu.cn
Haixiang Sun
stevensunz@163.com
Full list of author information is available at the end of the article

limited by its retrospective, single-center design, the model may support future clinical decision-making and improve COS outcomes. Prospective, multicenter validation studies are warranted to confirm its utility and generalizability.

Keywords Deep learning, Controlled ovarian stimulation, FSH dosing, Personalized medicine, Artificial intelligence

Background

Global epidemiological data indicate that infertility impacts millions of individuals and couples across demographic and geographic boundaries [1, 2]. Since the pioneering implementation of in vitro fertilization (IVF) in 1977, assisted reproductive techniques (ART) have become the predominant therapeutic intervention for addressing human infertility [3, 4]. Despite abundant technological and methodological advances in the field, considerable challenges persist in optimizing treatment protocols to maximize success rates while simultaneously minimizing procedural complications and ensuring comprehensive patient safety [5].

A pivotal component of in vitro fertilization (IVF) is controlled ovarian hyperstimulation (COH), which is crucial for acquiring multiple oocytes. Among various protocols, the GnRH-agonist long protocol has been associated with higher oocyte quality and cumulative clinical pregnancy rates, according to Cochrane reviews [6]. This protocol employs a GnRH agonist to prevent premature luteinizing hormone (LH) surges and ensure controlled ovarian stimulation, followed by follicle-stimulating hormone (FSH) administration to produce multiple follicles.

Precise FSH dosing is fundamental to successful COH. Both overdosing and underdosing carry significant risks: excessive FSH can trigger ovarian hyperstimulation syndrome (OHSS) and compromise oocyte quality, while insufficient doses may lead to poor follicular development and cycle cancellation [7, 8]. To optimize outcomes, reproductive endocrinologists conduct comprehensive monitoring that integrates multiple parameters, including daily follicular development and serum levels of estrogen, progesterone, and luteinizing hormone. This information guides precise adjustments to gonadotropin dosing and timing, with the goal of maximizing mature oocyte retrieval and treatment success rates. However, current FSH dosing practices remain largely subjective, varying significantly among clinicians, facilities, and countries. This lack of standardization poses a major challenge to treatment optimization and can result in suboptimal outcomes, increased costs, and avoidable complications.

Several researchers have attempted to develop more objective approaches to FSH dosing [9, 10]. La Marca et al. pioneered a predictive model incorporating age, serum FSH, and AMH levels to determine initial FSH doses [11]. Fanton et al. employed a KNN regression model to determine the initial FSH dosage, aiming to enhance treatment standardization [12]. However, their model was limited by its reliance on only three indicators and its inability to guide daily dose adjustments. Meanwhile, Letterie et al. utilized various machine learning methods to assist decision-making regarding the continuation of ovulation stimulation medications, the algorithm is robust but not as accurate for the dosing decision [13]. More recently, Xu et al. developed prediction models for both initial and day-6 FSH dosing for GnRH antagonist COS cycles [14], but this approach still falls short of addressing the need for continuous, real-time dose adjustment throughout the stimulation period.

The advent of Electronic Health Records (EHR) has created new opportunities for developing more sophisticated approaches to treatment optimization [15]. EHR data, with its complex temporal nature, high dimensionality, and heterogeneity, is particularly well-suited for analysis using deep learning methods [16]. Among various approaches, long short-term memory networks (LSTM) [17] and temporal delay neural networks (TDNN), especially the recently introduced D-TDNN [18], have shown promise in efficiently processing medical time series data. However, while deep learning has been widely applied to disease progression prediction, the modeling of treatment processes—particularly medication interactions and dosing—has received less attention.

To address these challenges, we propose a novel Cross-Temporal and Cross-Feature Joint Encoding (CTFE) model, built upon the D-TDNN architecture. This model uniquely integrates both temporal monitoring data and time-invariant patient characteristics to generate personalized FSH dosing recommendations. By incorporating comprehensive patient data and real-time monitoring information, CTFE aims to optimize treatment efficacy while minimizing complications and costs, ultimately providing more precise and effective care for individuals seeking fertility treatment.

Methods

Ethical approval and clinical data collection

This study was approved by the Ethics Committee of Nanjing Drum Tower Hospital (approval number: 2021–384-01). We conducted a retrospective analysis of

45,257 treatment cycles from the Reproductive Medicine Center of Nanjing Drum Tower Hospital between January 2018 and December 2020. Of these, 18,832 cycles using GnRHa long protocol (IVF/ICSI) were selected. The final study cohort comprised 13,788 patients with antral follicle counts between 7 and 30. This cohort was randomly divided into training (n = 6761), validation (n = 2898), and test (n = 4135) sets (Fig. 1). To avoid data leakage, only one cycle per patient was retained, and the dataset was split at the patient level. Hence, each patient's data appears in only one of the training, validation, or test sets. A successful treatment cycle was defined as achieving 6–20 follicles ≥ 14 mm in diameter on hCG trigger day. All patient information were de-identified prior to analysis for leak-proof processing in accordance with institutional privacy protocols and ethical guidelines.

Data processing

The initial dataset comprised 274 variables encompassing essential clinical parameters for determining FSH dosing in fertility treatment. These included basic patient characteristics (height, weight, age), reproductive history (primary or secondary infertility, previous IVF outcomes, duration of infertility), ovarian reserve markers (AMH, baseline FSH, LH, estradiol, and progesterone levels on cycle days 2–3), antral follicle counts (AFC), and daily hormone measurements (estradiol, progesterone, LH, FSH) throughout the treatment cycle.

During data preprocessing, variables with excessive missing data (> 60%), such as laparoscopic records, were excluded. Categorical variables were converted using one-hot encoding. For static patient characteristics, missing continuous variables were filled using mean imputation, while missing categorical values were set as 'None' prior to one-hot encoding. For dynamic monitoring data, missing values were filled using the value of the previous observation (forward fill) to maintain temporal continuity. In addition, all continuous variables were scaled using the min–max scaling method to ensure that the feature values fall within a consistent range.

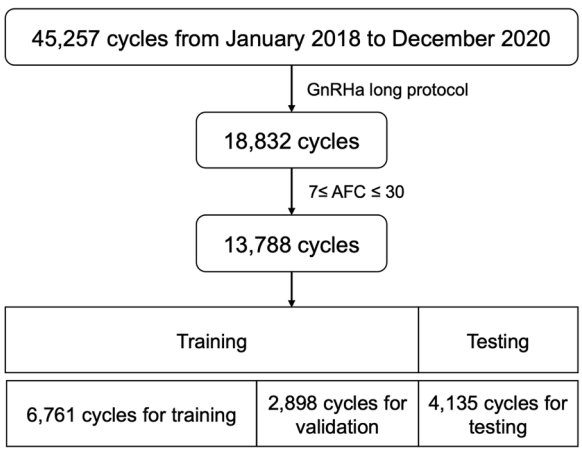

Fig. 1 Flow chart of the study selection process

The daily follicle monitoring data encompassed 13 key characteristics, including hormonal measurements (FSH, E2, LH, P), endometrial thickness, and bilateral ovarian follicle measurements. These monitoring parameters were ultimately transformed into 52 daily follicle detection attributes for analysis.

CTFE embedding system

Overview of the proposed model

We propose a novel time series data processing model for FSH daily dose prediction, comprising three main components: a time series data encoder, a time-invariant data joint encoder, and a fusion gate. The model employs D-TDNN (Deep Time Delay Neural Network) for time series data processing. Unlike traditional RNN-based approaches, our model analyzes feature relationships across both time series and features, capturing interactions between different monitoring parameters and the influence of initial states on dosage selection. We implement a feature weight extraction method to derive statistical measures (mean, variance, kurtosis, and skewness) from time series data for subsequent network layer processing. Patient cycle recommendations are generated using a similarity-based cycle retrieval strategy (Fig. 2).

Joint coding for cross-time series and cross-feature

Following data preprocessing, a patient's data can be represented as two distinct components: Ti, which encompasses the fundamental physiological information on day i of the current treatment cycle, like age, weight, GnRHa, etc. And Ts, which represents the sequential daily monitoring information as a time series (ts1, ts2, …, tsk), where k indicates the number of monitoring variables, like FSH, LH, E2, P, follicular size, etc.

After preprocessing the data, it is possible to describe the fundamental physiological information for a certain patient on day **I** of the current treatment cycle as **Ti** and the daily detection information as Ts = ts1, ts2, …, tsk.

The related inputs for the Ti and Ts generated by preprocessing the fundamental physiological information and daily detection information are $\boldsymbol{I_{Ti}} \in R^{B\times D_1}$, $\boldsymbol{I_{Ts}} \in R^{B\times L\times D_2}$, respectively.

Repeat for I_{Ti} in the dimension of the time series to obtain $\boldsymbol{I'_{Ti}} \in R^{B\times L\times D_1}$, Then link I_{Ti}' and I_{Ts} to get

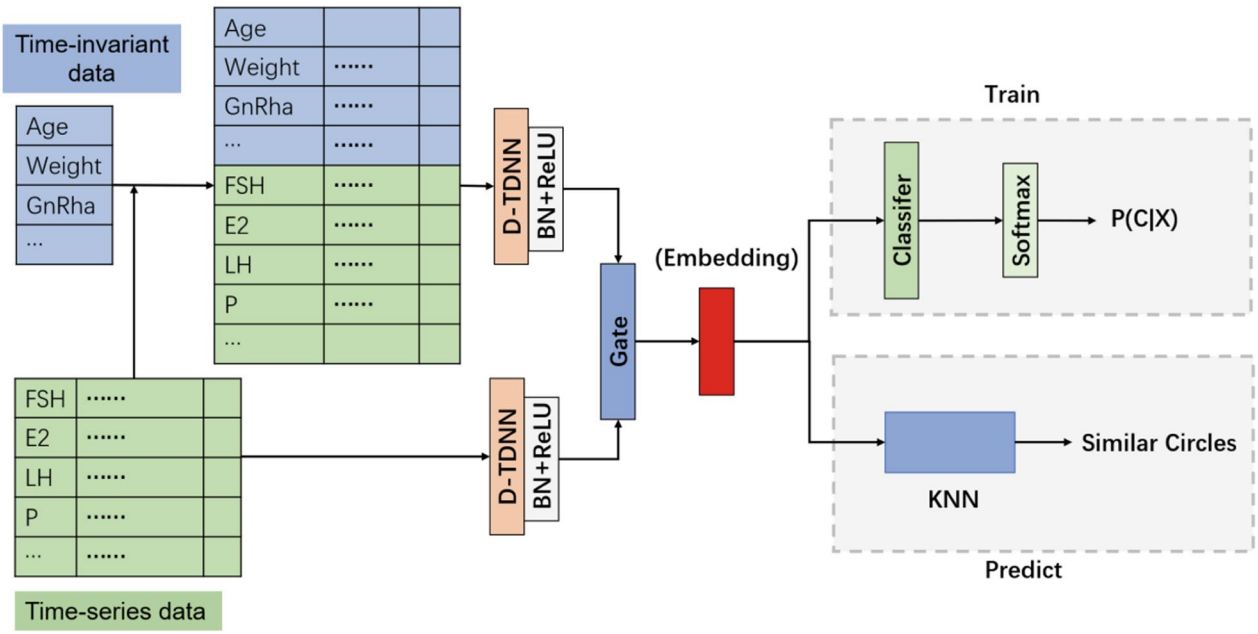

Fig. 2 Model structure overview

$\boldsymbol{I_T} \in R^{B\times L\times(D_1+D_2)}$.Using D-TDNN as an encoder, we get:

$$E'_{\mathrm{T}} = \mathrm{DTDNN\_1}(I_T) \tag{1}$$

$$E'_{\mathrm{TS}} = \mathrm{DTDNN\_2}(\mathrm{I_{Ts}}) \tag{2}$$

The DTDNN_1 and DTDNN_2 are two D-TDNN networks used in the model. DTDNN_1 is used to jointly extract cross-temporal and cross-feature relationships between basic physiological information and daily monitoring information, while DTDNN_2 is used to extract temporal features in the data. In DTDNN_2, one-dimensional convolution is used to obtain temporal feature correlations in the data, and multiple channels are used to capture corresponding feature correlations. E'_{T} and E'_{TS} are the results of $\mathrm{I_T}$ and $\mathrm{I_{Ts}}$ passing through the D-TDNN networks, respectively.

D-TDNN extracts temporal feature weights. We input mean, variance, kurtosis, and skewness data to the next layer network during pooling. This approach enriches temporal feature representations, improving feature connections in temporal changes. Fully linked layers project the feature representation.

$$\mathrm{E'_{Ts}} = ReLU(\mathrm{Linear(E_{Ts})}) \tag{3}$$

$$\mathrm{E'_{T}} = ReLU(\mathrm{Linear(E_{T})}) \tag{4}$$

Linear_1 and Linear_2 are linear layers, and ReLU is the activation function used.

Cross-feature information embedding

After obtaining temporal monitoring and cross-temporal cross-feature representations, we implement a novel fusion strategy. While temporal monitoring information is fundamental for medication dose prediction, it alone may not capture the complete clinical picture. The basic physiological information introduces important biases that influence temporal patterns. We address this through our AddGate mechanism, which integrates these complementary information sources.

$$H = AddGate(E_{Ts}; E_T) \tag{5}$$

Specifically:

$$E_{Tc} = ReLU(W[E_{Ts}; E_T] + b) \tag{6}$$

$$H = E_{Ts} + E_{Tc} \tag{7}$$

Note: W and b are learnable matrix parameters, E_{Tc} is the bias representation after weighted fusion, and ReLU is the activation function used.

Model training and prediction

We use the SoftMax function to classify the obtained Embedding:

$$Y = Softmax(Linear(H)) \tag{8}$$

And use the cross-entropy function as the loss function for training.

In the prediction phase, we search for the most comparable patient COS cycles using a Cosine-Similarity-KNN based method. The hidden layer representation set for the KNN retrieval sample is produced using the pre-network as the encoder:$M = \{H_1, H_2, \ldots, H_N\}$; the predicted sample hidden layer representation H_y. For the case where the parameter K of KNN is set to k, the similarity period retrieved is $M_s = \{H_{S1}, H_{s2}, \ldots, H_{sk}\}$, and then the predicted dose category C as well as the estimated success rate P can be obtained.

Our approach utilizes the Cross-Temporal and Cross-Feature Encoding (CTFE) framework to jointly encode both static patient attributes and dynamic monitoring data into a compact, low-dimensional latent representation. The CTFE model outputs fixed-size embeddings by integrating temporal stimulation-day variables with baseline clinical features, thereby effectively mitigating the challenges typically associated with high-dimensional input spaces. These embeddings are specifically optimized for cosine similarity calculations, allowing for efficient and meaningful proximity measurements in the latent space.

Rather than employing a traditional end-to-end deep neural regressor—such as a sequence of fully connected layers—for direct dose prediction, we adopt a cosine similarity–based K-nearest neighbor (KNN) retrieval strategy. This design enables the model to identify and reference prior IVF/ICSI cycles that exhibit similar encoded representations, thus generating dose recommendations grounded in clinically relevant precedents. In addition to yielding competitive prediction accuracy—particularly on critical stimulation days such as Day 1 and Day 5—this case-based retrieval approach enhances interpretability and clinical transparency. By presenting historical cases that are most similar to the current patient profile, the model can support clinicians in making personalized and evidence-informed dosing decisions.

We categorized the daily doses in the experiment because, in the operating process, the distribution of dosage data is discrete and different clinicians would treat patients at various doses differently. The FSH doses are grouped from high to low: 0 1 2 3 4 represent stop, low dose (< 80), medium and low dose (80 ~ 160), medium and high dose (160 ~ 240), high dose (> 240). The predicted result is the group in which the target dose is located, and the result is used for evaluation.

System implementation

We developed and implemented a real-time clinical decision support system for FSH dosage prediction. The system architecture comprises three main components: data acquisition, preprocessing, and prediction generation.

For data input, we designed a clinical interface that enables healthcare providers to input two categories of information: (1) patient baseline characteristics; and (2) daily monitoring parameters. The system processes these inputs through a data preprocessing pipeline implemented in Pandas framework, which performs standardization and validation procedures to ensure data quality and format consistency.

The preprocessed data is then analyzed by our trained model, which generates three key outputs: the KNN-predicted dosage recommendations, selection rates for each proposed dose, and their corresponding anticipated success rates. This provides clinicians with comprehensive information to support treatment decisions.

The model training and validation were performed on a high-performance computing system running Ubuntu 16.04 LTS, equipped with an NVIDIA GeForce RTX 2080 TI graphics processing unit. This infrastructure ensures efficient processing of complex calculations required for real-time clinical applications.

Results

Model performance comparison and analysis

The joint pair E_{TS} and E_T (AddGate) model framework exhibited the highest performance metrics, specifically in terms of Accuracy and Weighted-F1 scores. Through comprehensive analysis of the single-day model (day 9) across multiple model architectures, we demonstrated that the ETS and ET (AddGate) framework consistently achieved superior results compared to alternative approaches (DTDNN, LSTM, CTFE).

Implementation of our CTFE model, specifically designed to analyze relationships between cross-time-series and cross-feature data, yielded a significant 2.8% accuracy improvement over the baseline model (Table 1). The performance analysis revealed several key findings across different model configurations:

When utilizing E_T only, the DTDNN achieved an accuracy of 0.7018 (± 0.0082), while the LSTM model reached 0.6978 (± 0.0100) on the test set. After introducing the independent **Ets**, simple FC layer can improve the results substantially, with DTDNN accuracy increasing to 0.7230 (± 0.0053) and LSTM accuracy reaching 0.7206 (± 0.0026).

The incorporation of the AddGate mechanism further enhanced model performance, resulting in accuracy measurements of 0.7301 (± 0.0081) for DTDNN and 0.7242 (± 0.0046) for LSTM. These improvements underscore the effectiveness of AddGate in capturing

Table 1 Performance comparison

Model	Accuracy	Weighted_F1
E_T		
LSTM	0.6978(± 0.0100)	0.6921(± 0.0107)
DTDNN	0.7018(± 0.0082)	0.6952(± 0.0082)
E_{TS}&E_T		
LSTM(FC Layer)	0.7206(± 0.0026)	0.7155(± 0.0030)
DTDNN(FC Layer)	0.7230(± 0.0053)	0.7189(± 0.0067)
DTDNN(MAG)	0.7261(± 0.0064)	0.7218(± 0.0064)
DTDNN(AddGate)	0.7301(± 0.0081)	0.7254(± 0.0082)
CTFE(AddGate)	0.7367(± 0.0043)	0.7320(± 0.0048)

CTFE shows the best result

FC Layer uses a simple fully connected layer to add E_{TS} and E_T.MAG uses multimodal fusion to align data

complex relationships between cross-time-series and cross-feature data.

Our proposed CTFE model ultimately achieved the highest performance metrics, with an accuracy of 0.7367 (± 0.0043) and a Weighted-F1 score of 0.7320 (± 0.0048). These results demonstrate the robust capability of our CTFE model in handling both accuracy requirements and comprehensive class-specific metrics, establishing its effectiveness in clinical applications.

To validate the effectiveness of our proposed approach, we conducted a comparative analysis against traditional regression methods. It should be noted that our prediction target is the categorical grouping of FSH doses, where doses are classified as: stop, low dose (< 80), medium–low dose (80–160), medium–high dose (160–240), and high dose (> 240). Thus, the evaluation metrics are defined in terms of classification accuracy and weighted F1 score. Lasso regression was selected as a benchmark model due to its established role in medical dose prediction. The comparison focused on two critical time points in the treatment protocol: Day 1 (initial dose determination) and Day 5 (first dose adjustment).

For both our proposed method and Lasso regression, we utilized identical training and testing datasets to ensure a fair comparison. The evaluation metrics primarily focused on prediction accuracy, defined as the proportion of correct dosage predictions within the acceptable clinical range. All experiments were conducted using consistent data preprocessing procedures and validation protocols to maintain methodological rigor.

The results showed that our method significantly outperformed the lasso regression-based method. In terms of results, Lasso regression achieved an accuracy of 0.699 and 0.523 on Day 1 and Day 5, respectively, while our method achieved accuracies of 0.832 and 0.817, respectively (Table 2).

Table 2 Performance comparison between Lasso regression and our model (CTFE). CTFE shows better result

	Day1 accuracy	Day5 accuracy
Lasso regression	0.699	0.523
CTFE	0.832	0.817

Experimental results and effect of sliding window approach

In the dataset, as the number of treatment days increases, the amount of data decreases, making it difficult to train the model for the subsequent days. Therefore, in the model training after day 13, a sliding window of 10 data segments is used as the input data for the model, and the data for the later days are combined for training. The results are shown in the table below, and in the figure, using the sliding window (indicated by the"*sw*"suffix) leads to a significant improvement in the results for the subsequent days, alleviating the problem of decreasing training results over time.

Implementation of the sliding window approach demonstrated significant improvements in prediction accuracy for later treatment stages. While performance remained identical for days 1–12, marked improvements emerged from day 13 onward. The sliding window method achieved accuracy rates of 0.732 (day 16), 0.720 (day 17), and 0.781 (day 18), representing improvements of 4.6%, 5.6%, and 11.0% respectively over the standard approach. The most substantial improvement was observed on day 18, where the 11.0% accuracy gain highlighted the effectiveness of our approach in addressing data scarcity challenges during later treatment phases (Table 3, Fig. 3).

Qualitative evaluation

The output of the model is given to the doctor as a reference during the real prediction procedure so that more data may be generated. The model generated three key metrics for each dosage recommendation: predicted dosage, selection rate (frequency of dosage choice in similar cases), and success rate (proportion of successful outcomes with the recommended dosage).

To assess the practical utility of our model, we conducted qualitative evaluations using one representative case. For patient 1****, the model accurately predicted the progression of dosage adjustments from days 7–10, recommending low-to-medium doses initially, transitioning to low dose, and finally suggesting treatment termination. These predictions aligned with the actual treatment course (shows in Fig. 4).

Table 3 Daily accuracy results for the prediction

Day	Daily count	CFTE	CFTE-sw
1	3037	0.832	0.832
2	3031	0.981	0.981
3	3012	0.974	0.974
4	2979	0.974	0.974
5	2939	0.817	0.817
6	2882	0.933	0.933
7	2809	0.798	0.798
8	2695	0.858	0.858
9	2397	0.737	0.737
10	1999	0.798	0.798
11	1529	0.723	0.723
12	1126	0.782	0.782
(*sw* start)13	745	0.705	0.672
14	471	0.758	0.759
15	291	0.663	0.677
16	194	0.686	0.732
17	125	0.664	0.720
18	73	0.671	0.781
19	43	0.721	0.744

The sliding window (sw) starts to work from day 13 and the performance is improved in most days

The model's ability to provide selection and success rates for each recommendation enhanced its clinical interpretability. This feature allows physicians to evaluate recommendation reliability based on historical outcomes in similar cases, potentially improving clinical decision-making.

Discussion

The advent of Electronic Health Records (EHRs) has revolutionized medical data analysis, particularly in time series applications. EHR data's complex characteristics—including high dimensionality, multimodality, and heterogeneity—make it particularly suitable for deep learning approaches [19]. This digital transformation has accelerated the development of computational methods for analyzing patient histories, identifying cohorts, predicting risks, and exploring practical applications. In our study, rigorous data preprocessing – including mean imputation for static continuous features, forward filling for dynamic measurements, and min–max scaling to standardize all continuous variables – was applied to ensure data integrity for deep learning analysis.

Controlled ovarian stimulation (COS) is a crucial component of ART, directly determining embryo quality and pregnancy outcomes. However, COS faces numerous clinical challenges, including significant individual patient variability, varying ovarian responses, and asynchronous follicular development, all of which can impact treatment efficacy and safety. In recent years, artificial intelligence (AI) applications in COS have garnered increasing attention. By integrating and analyzing multidimensional data, including patients' clinical characteristics, ovarian ultrasound imaging features, and endocrine hormone levels, AI can construct predictive models to support clinicians in developing individualized COS protocols, thereby enhancing treatment safety and effectiveness [20].

Previous studies have demonstrated promising applications of artificial intelligence in controlled ovarian stimulation, yet significant limitations persist in existing

Fig. 3 Results shows that sliding window mechanism helps maintain good performance despite decreasing accuracy due to insufficient data

A

OID 1****

Days 1-6 omitted

Day 7:

Actual: low to medium dose

Predicted: low to medium dose

Alternatives:

- low to medium dose (Selectivity: 1.00, Success rate: 0.86)

Day 8:

Actual: low to medium dose

Predicted: low to medium dose

Alternatives:

- termination (Selectivity: 0.36, Success rate: 0.80)
- low to medium dose (Selectivity: 0.93, Success rate: 0.75)

Day 9:

Actual: low dose

Predicted: low dose

Alternatives:

- termination (Selectivity: 0.29, Success rate: 0.75)
- low dose (Selectivity: 0.50, Success rate: 0.71)
- low to medium dose (Selectivity: 0.21, Success rate: 1.00)

Day 10:

Actual: termination

Predicted: termination

Alternatives:

- termination (Selectivity: 0.71, Success rate: 0.90)
- low dose (Selectivity: 0.29, Success rate: 0.50)

B

	OID:1****				Similar Cycle OID:1****				Dissimilar Cycle OID:3****			
Day	GonalF	E2	LH	size	GonalF	E2	LH	size	GonalF	E2	LH	size
1	225	37.87	1.52	7	225	24.58	1.82	6.5	187.5	34.5	2.16	7
2	225	-	-		225	-	-		187.5	-	-	
3	225	-	-		225	-	-		187.5	-	-	
4	225	-	-		225	-	-		187.5	-	-	
5	225	539.78	0.88	11.5	225	568.36	0.96	13	112.5	1189	2.55	12.5
6	150	-	-		225	-	-		112.5	-	-	
7	150	1447.2	1.35	17.5	150	1213.1	1.88	14.5	112.5	3103	2.91	15..5
8	150	-	-		150	-	-		37.5	4613	2.12	18.5
9	75	4311.4	1.54	21.5	75	3372.8	3.88	18.5	0	5855	1.75	20.5
10	-	4856	1.36	23	-	4800	3.24	20	-	-	-	

Fig. 4 Partial prediction outputs for patient 1****. On day 10, the patient's cycle was terminated, and the model accurately predicted the termination of the cycle

approaches. The CONSORT calculator, while pioneering in its machine learning-based individualization of r-FSH dosing, was constrained by its small sample size and singular focus on initial dosing [21]. Similarly, Fanton's explainable machine learning model, despite incorporating comprehensive patient characteristics and employing advanced Random Forest algorithms, was limited by its retrospective nature and inability to provide

real-time adjustments [12]. The Stim Assist platform, though validated through multicenter prospective studies, primarily focused on starting dose optimization without addressing dynamic treatment adjustments [22]. In the realm of trigger timing prediction, while various AI systems have shown impressive accuracy [13, 23, 24], they typically function as standalone solutions, disconnected from the broader context of COS management.

To address these limitations, we propose a Cross-Temporal and Cross-Feature Encoding (CTFE) model that uniquely combines real-time FSH dose adjustment capabilities with trigger timing prediction. This comprehensive approach represents a significant advancement over existing models by providing dynamic, personalized treatment optimization throughout the entire COS cycle.

Our proposed represents a significant advancement in medication dosing prediction during ovulation stimulation. The model's innovative approach successfully captures the dynamic relationship between patients'baseline physiological characteristics and sequential treatment information, resulting in superior drug dosage prediction performance compared to traditional regression-based and temporal-related methods. A key strength of CTFE is its ability to provide interpretable outputs, including dosage selection rates and success probabilities, enhancing its clinical utility.

The application of deep neural networks (DNNs) to drug dosage prediction [25], particularly in IVF treatment, addresses a critical clinical need. IVF's complexity and cost necessitate precise FSH dosing, which must account for multiple patient-specific factors including age, body weight, AMH, FSH, LH, and ovarian reserve [13]. By reformulating IVF as a medication dose prediction problem, CTFE represents the first comprehensive DNN-based approach to predicting time-series FSH daily dosing.

Technical innovations in our model include the use of D-TDNN architecture for time-series data encoding, enhanced by our novel cross-temporal and cross-feature joint encoding approach. This combination effectively addresses a common limitation of existing deep learning models: the challenge of integrating temporal and time-invariant data. The superior performance of CTFE (accuracy: 0.7367 ± 0.0043, Weighted_F1: 0.7320 ± 0.0048) compared to simpler models using only temporal data demonstrates the value of this integrated approach. Although this study reports overall accuracy and weighted F1 score to evaluate model performance, we acknowledge the importance of class-wise precision and recall for assessing prediction fidelity across specific FSH dose groups. These metrics are particularly relevant in clinical scenarios where distinguishing between adjacent or less frequent dose categories may inform treatment safety and personalization. Due to current limitations in our experiment pipeline, detailed per-class metrics were not available for this version but will be included in subsequent analyses.

A notable methodological advancement of our study is the development of an innovative solution to address the common challenge of diminishing patient numbers during treatment progression. Our sliding window approach represents a significant technical contribution with three distinct advantages. First, it enables effective aggregation of multi-day treatment data, ensuring sufficient sample size for model training. Second, it transforms the traditional modeling paradigm by emphasizing treatment process patterns rather than fixed starting points. Third, it substantially expands the retrievable dataset, thereby enhancing overall model robustness.

The effectiveness of this approach is particularly evident in the later stages of treatment, where we observed significant improvements in prediction accuracy: 4.6% for day 16, 5.6% for day 17, and 11.0% for day 18. These improvements are especially meaningful given the traditional challenges in maintaining model performance during advanced treatment stages.

Despite these promising results, several limitations warrant consideration. The model's performance may be influenced by inherent biases arising from variations in treatment protocols and individual physician preferences. Additionally, the temporal scope (2018–2020) and size of our dataset may constrain the model's generalizability. Furthermore, while our results are encouraging, randomized controlled trials are necessary to definitively establish the model's clinical utility.

Future research directions should prioritize several key areas. First, expanding the dataset both temporally and geographically will enable further optimization and generalizability of the model, with prospective multicenter validations planned to confirm clinical applicability. Second, implementing mechanisms for continuous model updates would ensure adaptability to evolving clinical practices. Third, conducting rigorous randomized clinical trials will be essential to validate the system's real-world effectiveness. Finally, exploring the integration of alternative interpretability methods—such as attention mechanisms and SHAP values—and the application of this methodology to other sequential treatment scenarios in reproductive medicine and beyond will further enhance the impact of this approach.

In conclusion, CTFE represents a significant advancement in personalized FSH dosing for long-protocol ovulation induction. Its ability to provide data-driven, interpretable dosing recommendations

has the potential to enhance treatment efficacy, optimize resource utilization, and improve patient outcomes. While further validation is needed, this work establishes a promising framework for AI-assisted precision medicine in reproductive healthcare.

Conclusions

This study presents a novel deep learning model using cross-temporal and cross-feature joint encoding (CTFE) for optimizing daily FSH dosing during controlled ovarian stimulation. The model demonstrated robust performance with 73.7% accuracy in predicting personalized FSH doses, significantly outperforming conventional methods. By successfully integrating both static patient characteristics and dynamic monitoring data, our model addresses a critical gap in current COS management approaches.

While the single-center retrospective design presents certain limitations, the model's strong performance suggests its potential as a valuable clinical decision support tool. Future multicenter prospective studies are warranted to validate these findings and assess their impact on clinical outcomes. This work represents a significant step toward standardizing COS protocols through artificial intelligence, potentially advancing personalized medicine in reproductive healthcare and improving treatment outcomes. Notably, with a 19–56% improvement over LASSO on critical treatment days, the CTFE model demonstrates superior adaptability to dynamic patient responses.

Acknowledgements
Not applicable.

Author contributions
Jie Mei, Wujun Li, Haixiang Sun conceptualized the study, and all the authors contributed to refining the study concept and methods. Na Kong, Yu Xia, Zhilong Wang prepared and performed the data collection and analysis; Na Kong, Yu Xia, Zhilong Wang wrote the original draft of the manuscript. Hui Zhang, Liyan Duan, Yingchun Zhu, Chenyang Huang, Guijun Yan assessed and verified the underlying data and revised the final version of the manuscript. Jie Mei, Wujun Li, Haixiang Sun supervised the study and acquired funding. All authors participated in the interpretation of the data, reviewed and edited the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (NO.82030040), Scientific research project of Jiangsu Province Association of Maternal and Child Health (No. FYX202203), Reproductions Research Program of Young and Middle-aged Physicians. The authors state that no commercial or financial ties that might be considered as a possible conflict of interest existed during the research.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of Nanjing Drum Tower Hospital (Approval number: 2021–384-01).

Consent for publication
All authors have approved the manuscript to be submitted.

Competing interests
All authors declare no competing interest.

Author details
[1]Nanjing Drum Tower Hospital, Drum Tower Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing 210008, Jiangsu, China. [2]Center for Reproductive Medicine and Obstetrics and Gynecology, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China. [3]Center for Molecular Reproductive Medicine, Nanjing University, Nanjing 2100 08, China. [4]National Institute of Healthcare Data Science at Nanjing University, Nanjing 210008, Jiangsu, China. [5]National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China. [6]Center for Medical Big Data, Nanjing Drum Tower Hospital, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing 210008, Jiangsu, China. [7]State Key Laboratory of Reproductive Medicine and China International Joint Research Center On Environment and Human Health, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [8]Department of Reproductive Medicine Center, Nanjing Drum Tower Hospital, State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing 210023, China.

Received: 18 March 2025 Accepted: 4 May 2025
Published online: 15 May 2025

References

1. Cox CM, et al. Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis. Hum Reprod Open. 2022;2022:51.
2. Bhattacharjee NV, et al. Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403:2057–99.
3. Wang C, et al. Prediction of pregnancy-related complications in women undergoing assisted reproduction, using machine learning methods. Fertil Steril. 2024;122:95–105.
4. Herbert M, Choudhary M, Zander-Fox D. Assisted reproductive technologies at the nexus of fertility treatment and disease prevention. Science. 2023;380:164–7.
5. TEGGO OS, Bosch E, Broer S, Griesinger G, Grynberg M, Humaidan P, Kolibianakis E, Kunicki M, La Marca A, Lainas G, Le Clef N, et al. ESHRE guideline: ovarian stimulation for IVF/ICSI†. Hum Reprod Open. 2020;2020:hoaa009.
6. Siristatidis CS, Gibreel A, Basios G, Maheshwari A, Bhattacharya S. Gonadotrophin-releasing hormone agonist protocols for pituitary suppression in assisted reproduction. Cochrane Database Syst Rev. 2015;2015:CD006919.
7. Ngwenya O, et al. Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI). Cochrane Database Syst Rev. 2024;1:CD012693.
8. Polyzos NP, Sunkara SK. Sub-optimal responders following controlled ovarian stimulation: an overlooked group? Hum Reprod Oxf Engl. 2015;30:2005–8.
9. Schouten N, et al. Development and validation of a gonadotropin dose selection model for optimized ovarian stimulation in IVF/ICSI: an individual participant data meta-analysis. Hum Reprod Update. 2024. https://doi.org/10.1093/humupd/dmae032.

10. Fatemi H, et al. Dose adjustment of follicle-stimulating hormone (FSH) during ovarian stimulation as part of medically-assisted reproduction in clinical studies: a systematic review covering 10 years (2007–2017). Reprod Biol Endocrinol. 2021;19:68.
11. La Marca A, et al. Development of a nomogram based on markers of ovarian reserve for the individualisation of the follicle-stimulating hormone starting dose in in vitro fertilisation cycles. BJOG Int J Obstet Gynaecol. 2012;119:1171–9.
12. Fanton M, et al. An interpretable machine learning model for individualized gonadotrophin starting dose selection during ovarian stimulation. Reprod Biomed Online. 2022;45:1152–9.
13. Letterie G, Mac Donald A. 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;114:1026–31.
14. Xu H, et al. POvaStim: an online tool for directing individualized FSH doses in ovarian stimulation. Innov Camb Mass. 2023;4: 100401.
15. Heumos L, et al. An open-source framework for end-to-end analysis of electronic health record data. Nat Med. 2024;30:3369–80.
16. Huang T, et al. Artificial intelligence for medicine: progress, challenges, and perspectives. Innov Med. 2023;1: 100030.
17. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80.
18. Yu YQ, Li WJ. Densely connected time delay neural network for speaker verification. Interspeech. 2020. https://doi.org/10.21437/Interspeech.2020-1275.
19. Chen S, Yu J, Chamouni S, Wang Y, Li Y. Integrating machine learning and artificial intelligence in life-course epidemiology: pathways to innovative public health solutions. BMC Med. 2024;22:354.
20. Letterie G. Artificial intelligence and assisted reproductive technologies: 2023. Ready for prime time? Or not. Fertil Steril. 2023;120:32–7.
21. Bloomberg M, Brocklebank L, Hamer M, Steptoe A. Joint associations of physical activity and sleep duration with cognitive ageing: longitudinal analysis of an english cohort study. Lancet Healthy Longev. 2023;4:e345–53.
22. Canon C, et al. Optimizing oocyte yield utilizing a machine learning model for dose and trigger decisions, a multi-center, prospective study. Sci Rep. 2024;14:18721.
23. Letterie G, MacDonald A, Shi Z. An artificial intelligence platform to optimize workflow during ovarian stimulation and IVF: process improvement and outcome-based predictions. Reprod Biomed Online. 2022;44:254–60.
24. Fanton M, et al. An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation. Fertil Steril. 2022;118:101–8.
25. Mtshali S, Jacobs BA. Machine learning-based prediction of pharmacokinetic parameters for individualized drug dosage optimization. Int J Inf Technol. 2024. https://doi.org/10.1007/s41870-024-02314-x.

Publisher's Note