Estimated Date of Delivery with Electronic Medical Records by a Hybrid GBDT-GRU Model

Yina Wu
Hangzhou Normal University
Yichao Zhang
Hangzhou Normal University
Xu Zou
Hangzhou Hele Tech.Co
Zhenming Yuan
Hangzhou Normal University
Wensheng Hu
Hangzhou Women's Hospital
Sha Lu
Hangzhou Women's Hospital
Xiaoyan Sun
Hangzhou Normal University
Yingfei Wu (wyf@hznu.edu.cn)
Hangzhou Normal University

Research Article

Keywords: Estimated date of delivery, Gated Recurrent Unit, Gradient Boosting Decision Tree, Hybrid model

Posted Date: January 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-147356/v1

Research Article

Estimated date of delivery with electronic medical records by a hybrid GBDT-GRU model

Yina Wu[1+], Yichao Zhang [1+], Xu Zou[2], Zhenming Yuan[1], Wensheng Hu[3], Sha Lu[3], Xiaoyan Sun[1], Yingfei Wu[1*]

* Correspondence: wyf@hznu.edu.cn;

[+] Yina Wu and Yichao Zhang contributed equally to this work;

[1] Engineering Research Center of Mobile Health Management Ministry of Education, Hangzhou Normal University, Hangzhou, China;

[2] Hangzhou Hele Tech. Co, Hangzhou, China;

[3] Hangzhou Women's Hospital, Hangzhou, China

Abstract

Background: An accurate estimated date of delivery (EDD) helps pregnant women make adequate preparations before delivery and avoid the panic of parturition. EDD is normally derived from some formulates or estimated by doctors based on last menstruation period and ultrasound examinations. The main aim of this study was to develop a hybrid model to improve the accuracy of EDD and promote the health and safety of pregnant women and fetuses.

Methods: This study attempted to combine antenatal examinations and electronic medical records to develop a hybrid model based on Gradient Boosting Decision Tree and Gated Recurrent Unit (GBDT-GRU). Besides exploring the features that affect the EDD, GBDT-GRU model obtained the results by dynamic prediction of different stages. The mean square error (MSE) and coefficient of determination (R^2) were used to compare the performance among the different prediction methods. In addition, we evaluated predictive performances of different prediction models by comparing the proportion of pregnant women under the error of different days.

Results: The clinical data were collected with 33,222 pregnancy examination records from 5537 Chinese pregnant women who have given birth. Experimental results showed that the hybrid GBDT-GRU model outperformed other prediction methods with coefficient of determination

(R^2) of 0.84, mean square error (MSE) of 41.73. We also found that the GBDT-GRU model had a smaller deviation by comparing the bias between the actual delivery date and the EDD under different methods.

Conclusions: In comparison with other prediction methods, the GBDT-GRU model provided better performance results. The results of this study are helpful for the development of guidelines for clinical delivery treatments, as it can assist clinicians in making correct decisions during obstetric examinations.

Keywords: Estimated date of delivery, Gated Recurrent Unit, Gradient Boosting Decision Tree, Hybrid model

---

Background

Accurate estimated date of delivery (EDD) is helpful for pregnancy outcomes and clinical decisions making [1], including diagnosing preterm and full-term, formulating measures for fetal dysplasia, arranging the timing of prenatal examination, preparing nursing measures for parturition and improving the efficiency of delivery. A reliable EDD is very important to reduce the occurrence of premature or postmature babies and is critical for both short-term and long-term health outcomes in neonates. Inaccurate EDD may have adverse effects on the health and safety of pregnant women and fetuses.

The current clinical method of determining the EDD is based on the information about last menstrual period(LMP) and ultrasound [2, 3, 4]. Among them, the Naegele's rule based on LMP is the most common and wide method to calculate the EDD [5]. The Naegele's rule is calculated by adding seven days and nine months to the first day of the LMP. Alternatively, the EDD is 280 days after the first day of the LMP [6]. However, the limitations of LMP include deviations in recalling the last menstruation, irregular menstrual cycles, oral contraceptives and early pregnancy bleeding[7]. In several studies, calculating EDD by ultrasound of the first trimester of pregnancy is more accurate than the LMP[8]. However, all ultrasound measurements need to be performed by specially trained medical staff and the precision probably is influenced by the technical level of ultrasound examiners. Besides, some studies believed that the accuracy of EDD will gradually decrease with the increase of gestational weeks[9]. Therefore, the EDD should be determined once we obtain the data from LMP or the first accurate ultrasound examination[10, 11]. However, some studies showed that only about 5% of births are born exactly on EDD, regardless of the LMP methods or ultrasound[12].

In the study of medical prediction methods, machine learning models are widely used since its high accuracy and high efficiency. Liang Liang et al[13] used a linear regression model to find the blood metabolites that can predict gestational age and delivery date accurately. Thomas L.A et al.[14] used convolution neural network(CNN) to estimate fetal head circumference, so as to determine gestational age and delivery date. Russell Fung et al.[15] developed a machine learning approach based on ultrasound-derived and fetal biometric data to estimate gestational age and delivery date, but this article did not mention the type of ML. Schink T at al.[16] developed an algorithm to estimate the beginning of pregnancy in German claims data focusing on the potential of the expected delivery date. Torres M T et al.[17] designed a system to calculate the gestational age and delivery date. They used images from the feet, face and ear of 130 newborn babies and a combination of fully convolutional networks, CNN and support vector regressors (SVR). However, the above-mentioned prediction models ignored the effect of time series factors. Accurate EDD needs to evaluate the physical condition of pregnant women, and analyze the recent trend by judging the fetal development status of pregnant women at all stages. Since the data of antenatal examination is time series data, the EDD is closely related to the results of each examination.

The purpose of this study was to propose a hybrid prediction model that combined two models to obtain a more reliable prediction for EDD. The hybrid model was based on Gradient Boosting Decision Tree and Gated Recurrent Unit (GBDT-GRU). The rest of the paper is organized as follows: Section 2 introduces data preparation, feature selection, detail description of GBDT-GRU model and other technical methods. Section 3 provides the actual data and experimental results, and compares the proposed model with other prediction methods. Section 4 explores the application in EDD. Finally, section 5 summarizes the main conclusions and future prospects of this paper.

Methods

Framework for the estimated date of delivery

This study aimed to predict the EDD by using a hybrid model of GBDT and GRU. GBDT-GRU model made more effective and reasonable decisions by obtaining information from experience and mining hidden knowledge in data. The block diagram of the prediction process is shown in Fig. 1 and the detailed explanations of each step are as follows:

Step 1: Data preparation: We obtained a high-quality dataset by processing the original

maternal data, such as data cleaning and transformation. Then, according to the characteristics of physiological parameters during different periods of pregnancy, the data was divided into two datasets: the first trimester of pregnancy dataset, the second and third trimesters of pregnancy dataset.

Step 2: Feature selection: The GBDT model was used in the feature selection of the two datasets from step 1. Feature selection could solve the problem of information redundancy and reduce the dimension of data.

Step 3: GBDT-GRU model for prediction: First, we used GBDT model based on the first trimester of pregnancy dataset to predict the initial EDD, then regarded this EDD as a new feature. Second, the new feature was combined with the original features of the second and third trimester of pregnancy to obtain a new fused dataset. Finally, GRU model generated the final EDD on the bias of the new fused data.

Step 4: Comparison and evaluation: We verified the accuracy of this method by comparing different models. The qualities of the predictive models were validated by standard concepts of accuracy as well as clinical usefulness.

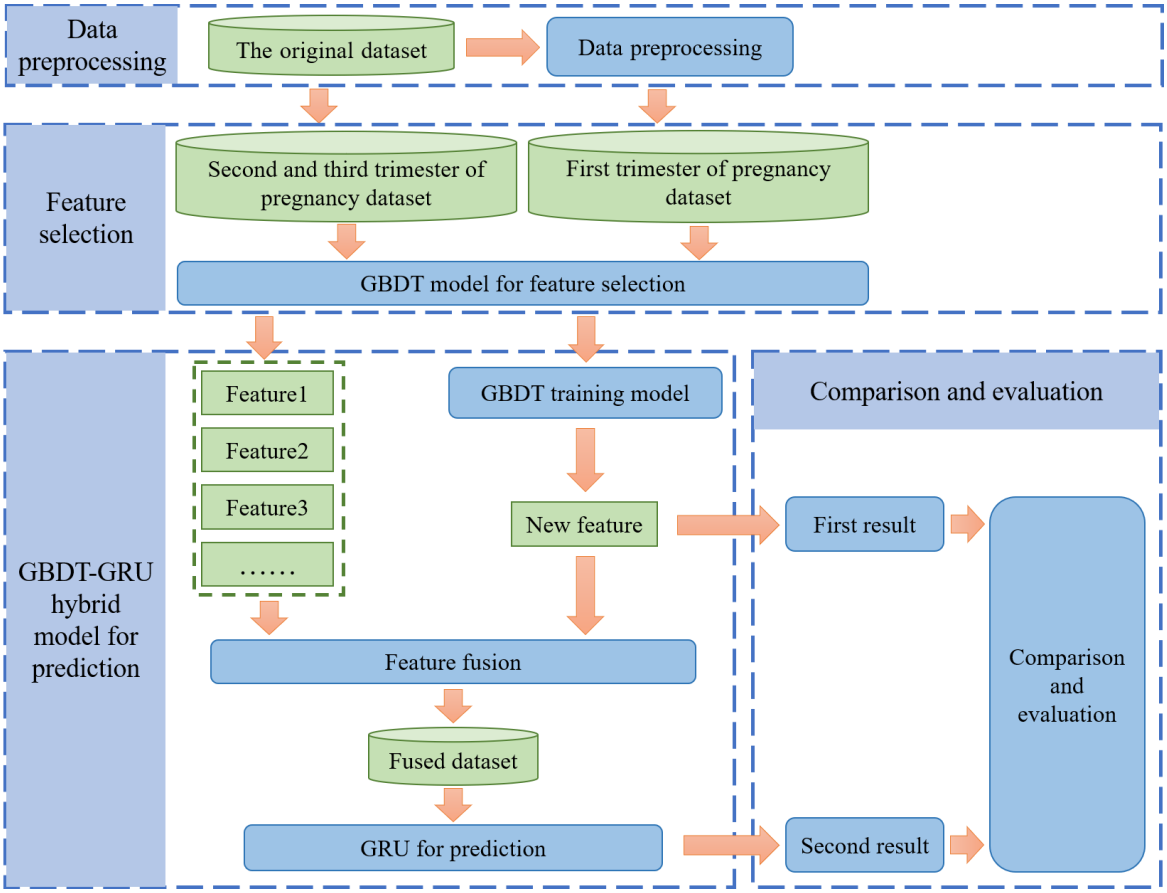


Fig. 1 GBDT-GRU framework for the estimated date of delivery

Datasets preparation

In this study, the data were collected from the electronic medical records (EMR) of pregnant women in a maternity hospital in eastern part of China. We extracted the physical examination and ultrasound records of the pregnant women who natural and full-term delivery between 2017 and 2020，with approval from Hangzhou Women's Hospital (Hangzhou Maternity and Child Health Care Hospital) (written permission with approval NO. 2019-02-2). All medical procedures were performed in accordance with relevant guidelines and regulations. The informed consent requirement for this study was waived by the board because the researcher only accessed the database for analysis purposes, and all patient data were de-identified. Considering the necessity of predict EDD in advance, only the physical examination records before 35 weeks were used in this study. The count of physical examinations of pregnant women in different gestational weeks is shown in Fig 2. According to Fig 2, there are too few pregnancy examinations between 13 and 22 weeks. In addition to the frequency of pregnancy examination, pregnant women have different examination items at different stages of pregnancy. Some ultrasound indicators only appear in the first trimester of pregnancy and will disappear with the increase of pregnancy weeks, such as the gestational sac (e. g, Features of the pregnancy examination Table 1). Therefore, we divided the dataset into two subsets according to the time of pregnancy examination: the first trimester of pregnancy dataset (pregnant week: 4 to less than 14 weeks); the second and third trimester of pregnancy dataset (pregnant week: 23 to less than or equal 35 weeks).

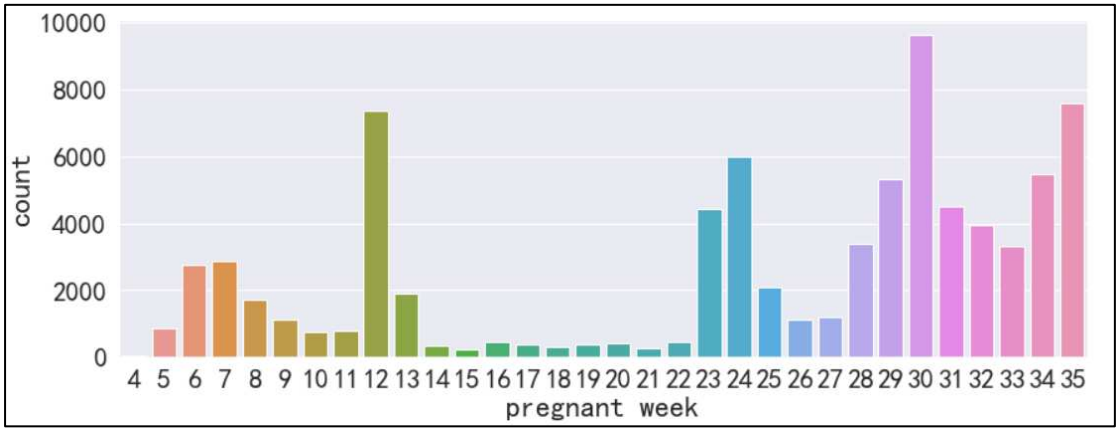


Fig. 2 Count of the pregnant women in different pregnant weeks

Table 1 Features of the pregnancy examination

Types	Feature	Notation	Pregnant week
Static	Height	Height of pregnant woman(cm)	/
	Age	Age of pregnant woman	/

	P-W	Pre-pregnant weight of pregnant woman(kg)	/
	Gravidity	Gravidity	/
	Parity	Parity	/
	F-SBP	Systolic blood pressure of the first pregnancy examination	/
	F-DBP	Diastolic blood pressure of the first pregnancy examination	/
	LMP	Last menstrual period	/
	MD	Menstrual days	/
	MC	Menstrual cycle	/
	MA	Menarche age	/
	MV	Menstrual volume	/
	DY	Dysmenorrhea(0,1)	/
	DH	Disease history	/
Time series	SBP	Systolic blood pressure(mmHg)	4-35
	DBP	Diastolic blood pressure(mmHg)	4-35
	FUH	Fundal height (cm)	11-35
	AC	Abdomen circumference(cm)	11-35
	FHR	Fetal heart rate(times/min)	11-35
	HRF	High risk factors	4-35
	BMI	Body mass index(kg/m2)	4-35
	GSL	Gestational sac length(cm)	4-16
	GSW	Gestational sac width(cm)	4-16
	GSH	Gestational sac height(cm)	4-16
	FP	Fetal position	11-35
	PMG	Placental Mature Grading	12-35
	AFI	Amniotic fluid index(cm)	13-35
	S/D	Systolic to diastolic(ratio)	21-35
	NT	Nuchal translucency(cm)	10-14
	CRL	Crown-rump length(cm)	7-13
	BPD	Biparietal diameter(cm)	12-35
	FAC	Fetal abdomen circumference(cm)	12-35
	FL	Femur length(cm)	12-35
	HC	Head circumference(cm)	12-35
	HGB	Hemoglobin(g/L)	23-35

Due to the variability and irregularity of pregnancy examination dates, some samples will be lost. In this work, we deleted the samples that lacked key features. For example, a sample only has basic features such as height and weight, but it lacks all important features such as gestational sac size, FAC, HC and so on. Moreover, samples with more than 50% missing values were excluded from further analysis. And the missing values of time series data were filled by linear interpolation according to the time of two adjacent pregnancy examinations. The gap between the values of variables, resulting from the different dimensions and dimensional units of variables, could affect the performance of the model. Therefore, it was necessary to normalize the data to avoid the influence of the larger range of values on other features and improve the convergence speed of the model. The min-max normalization is used to scale the values of the result to range [0,1], which is represented in Equation (1) as:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x is the current eigenvalue, x_{min} and x_{max} are the minimum and maximum values of the current feature, and x^* is the standardized eigenvalue. The prediction results generated by the model need to be further denormalized as shown in equation (2), where y is the true value and $y_{predict}$ is the predicted value.

$$y = y_{predict} * (x_{max} - x_{min}) + x_{min} \tag{2}$$

Feature selection

GBDT[18] is a boosting algorithm based on the classification and regression tree(CART). In this study, we used GBDT model for feature importance analysis and selection in the two datasets. The feature selection of GBDT is based on calculating the gain of the split nodes of the decision tree and using cumulative summation to evaluate the appropriateness of features[19]. The importance of a feature is measured by calculating the average importance of a feature in a single tree. GBDT uses formula (3) as a measure of influence, $\hat{I}_j$ is the relative influence.

$$\hat{I}_j^{\,2} = \frac{1}{M}\sum_{m=1}^{M} \hat{I}_j^{\,2}(T_m) \tag{3}$$

Where $\{T_m\}_1^M$ means a collection of decision tree, M represents the number of trees. The importance of feature j in a tree is calculated according to the formula(4):

$$\hat{I}_j^{\,2}(T) = \sum_{t=1}^{J-1} \hat{\imath}_t^{\,2} 1(v_t = j) \tag{4}$$

Where $\hat{\imath}_t^{\,2}$ represents the squared loss, v_t means a feature associated with j nodes, and $J-1$ is the number of non-leaf node.

In the feature selection process, we generated the feature weights group $W = \{w_1, w_2, \dots, w_n\}$ from prenatal examination datasets and selection results of GBDT model, where w_i describes the weight of each feature. Feature selection was performed on the two subsets based on the contribution degree of each feature. In this paper, we added the features one by one according to the weights from high to low, and selected the features used in this experiment by comparing the error and running time.

Hybrid GBDT-GRU model

Since our aim was to predict the remaining days of pregnancy, the uncertainty of the future of pregnancy and the strict requirement of accuracy was really challenging. We designed a hybrid GBDT-GRU model and the structure was shown in Fig. 3. GBDT model is a kind of boosting algorithm, which belongs to the category of ensemble learning[20]. Among the machine learning methods used in practice, GBDT runs faster when training large amounts of data and have stronger robustness when processing outlier value. In this study, we used the GBDT model for the first prediction with the dataset of first trimester of pregnancy, then took the predicted results as the initial EDD. As a new feature, the initial EDD was fused with the second and third trimester of pregnancy dataset to obtain a fused dataset.

GRU[21] is a variant of Recurrent Neural Network (RNN)[22], which is proposed to solve the problems of gradient in long-term memory and back propagation[23]. With the design of update gate and reset gate, GRU model is capable of handling the time series data as well. The input layer of GRU is the time series from fused dataset, which can be noted as $X = \{x_1, x_2, \dots, x_t\}$, where x_i represents the record of the i th physical examination of pregnancy women. The hidden state h_{t-1} contains the information of the previous node. Where z_t and r_t denote the update gate and reset gate, respectively. Wr and Wz are the weight matrices from hidden states at previous time step to the update gate and reset gate, respectively. σ is a sigmoid function. The formula is expressed as follows:

$$r_t = \sigma(Wr \cdot [h_{t-1}, x_t]) \tag{5}$$

$$z_t = \sigma(Wz \cdot [h_{t-1}, x_t]) \tag{6}$$

The reset data obtained by the reset gate of the hidden layer data at the final moment is

combined with the current input x_t, and $tanh$ is the activation function. The activation state of the hidden layer at the current moment $\tilde{h}_t$ can be defined as:

$$\tilde{h}_t = tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \tag{7}$$

Then the same gate z_t is used to select and forget memory, and the hidden state h_t of time t can be calculated by:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{8}$$

Finally, we used the result of the last moment of the output layer as the final EDD. Meanwhile, the final EDD could predict more accurately than initial EDD.

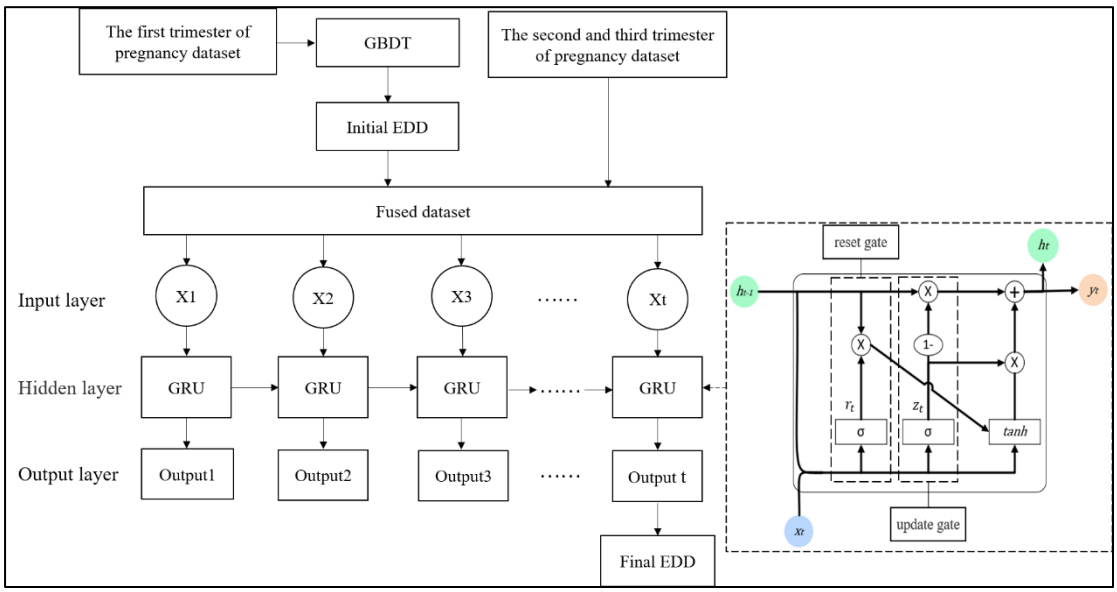


Fig. 3 Structure of GBDT-GRU

The parameters of these prediction models were determined by grid search. The grid search method combined all possible parameters, then trained each group of parameters to find the best combination of parameters. After cross-validation, the hyperparameter combination with the highest average score was taken as the best choice, and the model object was returned. Table 2 shows the parameter settings of GBDT and GRU models.

Table 2 Parameters settings of GBDT and GRU models

Model	Parameters	Values
GDBT	Learning rate	0.01
	N_estimators	500
	Min_samples_leaf	4
	Min_samples_split	3
	Max_depth	3
GRU	Batch size	100

Loss function	MSE
Layers	1
Optimizer	Adam
Hidden_size	37
Input size	18
Learning rate	0.002
Epochs	200

Evaluation methodology

The prediction errors were considered as an essential factor to evaluate the proposed model. In this study, the coefficient of determination (R^2), Mean Absolute Errors(MAE) and Mean Square Error(MSE) were used as the evaluation indices of the models. The parameters of these prediction models were determined by grid search and the models were validated with 5-fold cross-validation. The 5-fold cross-validation splits the training dataset into two sections, where 80% of dataset is used for training and the remaining 20% is used for testing. The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_i^n \left(\hat{y}^{(i)} - y^{(i)}\right)^2}{\sum_i^n (\bar{y} - y^{(i)})^2} \quad (9)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m} \left| y^{(i)} - \hat{y}^{(i)} \right| \quad (10)$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2 \quad (11)$$

Where $y^{(i)}$ and $\hat{y}^{(i)}$ are the real and predicted values, respectively, and $\bar{y}$ is the average value of real values.

In order to further assess the effectiveness of prediction based on the GBDT-GRU model, the bias in predicting EDD of each method was used as another critical index of prediction reliability. The D_{bias} is defined as formula (12), where D_{real} is the actual date of delivery and $D_{predict}$ is the EDD.

$$D_{bias} = |D_{real} - D_{predict}| \quad (12)$$

By counting the proportion of people under different D_{bias}, we could get the performance and availability of different methods in practical applications. We calculated the accuracy

under specific requirements $Accuracy_{bias}$ by Eq (13).

$$Accuracy_{bias} = \frac{n_{D_{bias}}}{N} * 100\% \quad (13)$$

Where N is the total number of pregnant women, $n_{D_{bias}}$ means the number of pregnant women whose prediction bias are less than D_{bias}.

Results

Description of the experimental data

The dataset used in this study comes from a hospital in the eastern part of China, which includes a large amount of data such as the maternal ultrasound records, prenatal examination reports and so on. After data preprocessing, the pregnancy dataset was obtained includes 33,222 pregnancy examination records and ultrasound records of 5537 pregnant women. Table 2 describes the features of the dataset.

Feature importance and feature selection

We used GBDT to selected the features that have vital influences on EDD. The selected features were used as the input of the prediction model, which reduced the dimension of the input and solved the problem of information redundancy. Feature importance reflected the contribution of each variable in EDD. The results for feature selection of different datasets results are shown in Fig. 4 a, c. The GSL was the most important variable to affect EDD in the first pregnancy dataset, followed by GSW, GSH, P-W, Age, MC and MD(Fig. 4 a). At the same time, FAC was the feature with the highest weight value in the second and third trimester of pregnancy dataset, followed by FL, HC, BPD, AFI, UH, P-W, HGB, DBP, BLG, BMI, Age and SBP(Fig. 4 c).

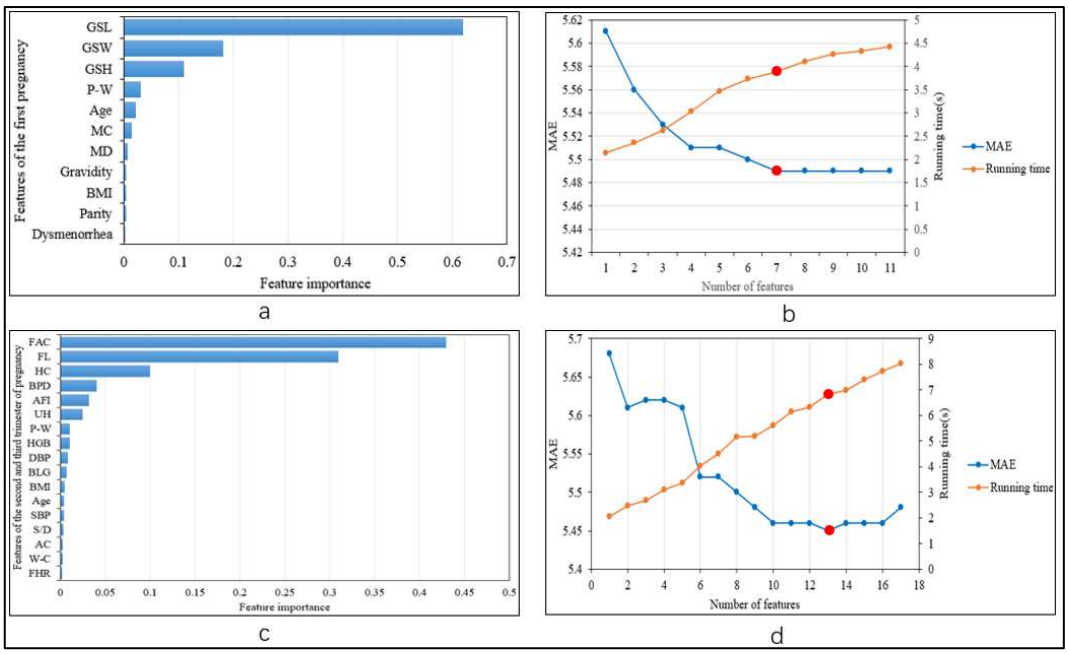


Fig. 4 Analysis result for feature selection of different datasets. **a** shows the feature importance of the first pregnancy. **b** shows the MAE and running time with different number of features. **c** represents the feature importance of the second and third trimester of pregnancy. **d** represents the MAE and running time with different number of features in the second and third trimester of pregnancy data.

We added the features one by one according to the weights from high to low. The corresponding MAE values and running time after training the different number of features with GBDT are shown in Fig 4 b, d. We chose the feature group with the shortest running time in the case of the lowest MAE. Finally, seven features were retained in the first pregnancy dataset and 13 features were reserved in the second and third trimester of pregnancy data. Table 3 shows the summary statistics of these parameters.

Table 3 Summary statistics of parameters

First pregnancy	**Value (Mean ± SD)**	**The second and third trimester of pregnancy**	**Value (Mean ± SD)**
Pregnant Days	54.2±10.8	Pregnant Days	218.5±14.9
GSL	3.6±1.6	FAC	26.8±2.4
GSW	2.7±1.3	FL	5.8±0.5
GSH	2.1±1.2	HC	28.5±1.8
P-W	53.9±7.2	BPD	7.9±0.6
Age	29.3±3.4	AFI	11.7±2.6
MC	30.6±4.4	FUH	28.8±2.5

MD	5.9±1.1	P- W	53.5±7.2
/	/	HGB	116±10.1
/	/	DBP	67.3±8.1
/	/	BLG	4.3±0.4
/	/	BMI	24.4±2.6
/	/	Age	29.3±3.4
/	/	SBP	113.4±10.7

Evaluation and comparison of different models

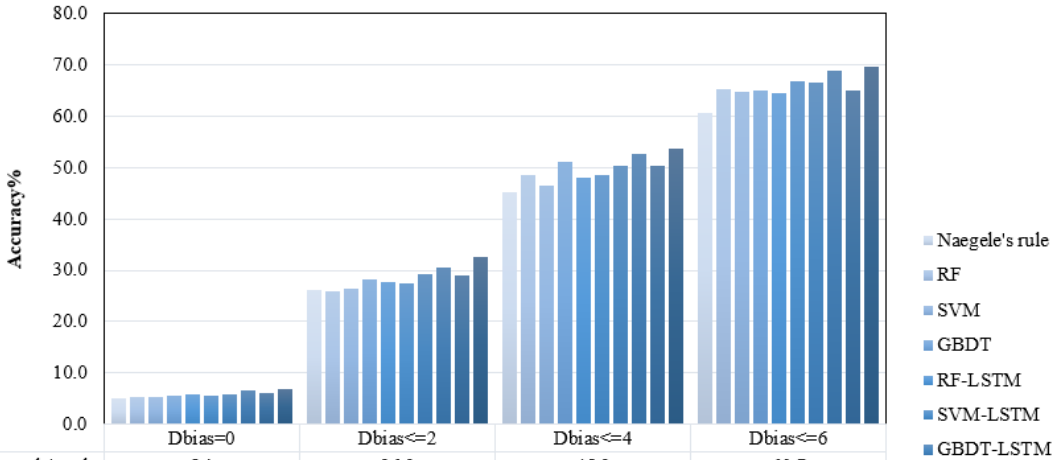
In order to effectively evaluate the experimental results of GBDT-GRU model, we compared the prediction results of the Naegele's rule and some machine learning models. The machine learning models we used for comparison include Random Forest(RF), Support Vector Regression (SVR) and Long short-term memory (LSTM). RF is a powerful algorithm for classification and regression, the prediction is made by majority vote or averaging the results of the ensemble[24, 25]. SVR[26] is an extension of the concept of Support Vector Machine(SVM), which is used for regression purpose. LSTM is a kind of RNN, which can be used to predict events with long intervals and time[27, 28]. Based on the above-mentioned features in Table 3, we constructed these machine learning models to predict the EDD.

Table 4 Performance of different methods compared in two datasets

Datasets	**Method**	**MSE**	**R^2**	**Training time(seconds)**
	Naegele's rule	60.74	/	0
First trimester of pregnancy dataset	RF	48.34±0.2	0.61±0.01	6.3
	SVR	48.66±0.2	0.60±0.01	620
	GBDT	46.73±0.2	0.63±0.01	3.9
	LSTM	47.35±0.2	0.65±0.01	436
	GRU	46.89±0.2	0.65±0.01	205
Fused dataset	SVM-LSTM	48.30±0.2	0.80±0.01	1025
	RF-LSTM	44.12±0.2	0.81±0.01	560
	GBDT-LSTM	46.13±0.2	0.81±0.01	510
	SVM-GRU	46.60±0.2	0.82±0.01	970
	RF-GRU	43.49±0.2	0.83±0.01	250
	GBDT-GRU	**41.73±0.2**	**0.84±0.01**	245

The average values of the results after 5-fold cross-validation are shown in Table 4. We provided a performance comparison of the prediction models in different datasets. First, GBDT, RF and SVR were used to predict the initial EDD from the first trimester of pregnancy dataset. Second, the final EDD was gained with the time series model based on the fused dataset. Finally, MSE, R^2 and training time were used to compare the prediction results of different models. Table 4 shows that the GBDT-GRU prediction model outperforms Naegele's rule, all the single models and other hybrid models, achieves average MSE of 41.73 and R^2 of 0.84. Moreover, comparing with the hybrid LSTM models, the hybrid GRU models have a shorter training time.

According to the difference between EDD and the actual date of delivery, we recorded and compared the accuracy rate of each model under four categories: D_{bias} smaller or equal to zero, two, four and six. The accuracy of different methods under different D_{bias} is shown in Fig 5.



	Dbias=0	Dbias<=2	Dbias<=4	Dbias<=6
Naegele's rule	5.1	26.3	45.3	60.7
RF	5.2	26.0	48.6	65.4
SVM	5.4	26.4	46.6	64.7
GBDT	5.7	28.2	51.1	65.0
RF-LSTM	5.9	27.7	48.1	64.5
SVM-LSTM	5.7	27.4	48.5	66.8
GBDT-LSTM	5.8	29.2	50.4	66.7
RF-GRU	6.5	30.5	52.6	68.9
SVM-GRU	6.2	29.1	50.3	65.1
GBDT-GRU	6.9	32.6	53.6	69.8

Fig. 5 The accuracy of different methods under different D_{bias}

As shown in Fig 5, our GBDT-GRU models achieved better prediction results than other methods for different D_{bias}. The accuracy of EDD by GBDT-GRU model was 6.9%, 32.6%, 53.6% and 69.8%, when D_{bias}<= 0, 2, 4 and 6 days. Significantly, with the increase of D_{bias}, the accuracy advantage of GBDT-GRU model is more obvious.

Discussion

In this study, we used a hybrid model of GBDT and GRU to generate features from EMR and to predicted the EDD of pregnant women. The accuracy of the GBDT-GRU model was superior to other prediction methods. In addition, we selected the features that have great influence on the EDD to make the model have better performance of prediction.

The experimental results showed that the performance of hybrid models (GBDT-GRU, GBDT-LSTM, RF-GRU and RF-LSTM) were better than all single models. Hybrid models achieved overall MSE is smaller than 44.12 and R^2 is larger than 0.81. This shows that hybrid models have better generalization ability compared to other models for EDD, which may better serve and support the medical staff in decision making. Furthermore, the GRU presented better performance than LSTM when dealing with the time series data, which was benefited by the simpler gates structure of GRU. The GBDT-GRU exhibited the best performance among all models. As far as we know, this study was the first attempt to apply a hybrid model to the data of different stages of pregnancy, which could adjust the EDD according to the characteristics of each period of pregnancy. Therefore, it was obvious that our models were well suitable for the EDD of healthcare service.

As shown in Fig.5, the proposed model not only optimizes the model running time but also improves the prediction accuracy. When D_{bias} is less than six days, the accuracy of GBDT-GRU model is 9.1% higher than the Naegele's rule. In addition, the results of this study were helpful for the EDD and had development of guidelines for clinical delivery treatments.

The clinical research about EDD was still focused on ultrasound and LMP, such as head circumference[29], cervical length, some improved formula methods[30] and so on[31, 32]. These studies provided a reference for feature selection of machine learning. In addition, datasets of EMR provided great potential for EDD in pregnancy. We found that several new features were closely related to childbirth, which could enhance the accuracy of the EDD. The results of our study indicated that days of pregnancy, gestational sac size have a great influence on EDD in the first trimester of pregnancy. And for the second and third trimester of pregnancy, the influence of days of pregnancy, FAC, AFI and BPD were relative important features. Moreover, the importance of features given by GBDT model provides a reference for doctors to pay more attention to the key physiological indicators of pregnant women.

Our study also had several limitations that need to be improved. First, this study only used physical examination data and ultrasound data for prediction. We did not consider the influence of laboratory parameters on EDD. Second, the primary limitation of our study was

a possible selection bias due to the center study with small sample size, and its accuracy and practicality should be verified in prospective studies with larger samples.

Conclusions

In this paper, a hybrid model of a GBDT model and GRU model was proposed to predict EDD. For a more accurate EDD, we established a hybrid model of the parameters related to pregnant women and fetal physical examination. The results show that GBDT-GRU achieves a satisfactory outcome in the experiment and the accuracy of the EDD can be improved by adjusting the number of features. Therefore, our hybrid model is an effective method to support clinical decision making and artificial intelligence methods have great application potential in obstetrical practice. Future studies should also solve the problem of predicting the EDD within the scope of preterm delivery.

Abbreviations

EDD: estimated date of delivery; LMP: last menstrual period; ML: Machine learning; DL: Deep learning; CNN: Convolution neural network; GBDT: Gradient Boosting Decision Tree; GRU: Gated Recurrent Unit; EMR: Electronic medical records; P-W: Pre-pregnant weight of pregnant woman; F-SBP: Systolic blood pressure of the first pregnancy examination; F-DBP: Diastolic blood pressure of the first pregnancy examination; LMP: Last menstrual period; MD: Menstrual days; MC: Menstrual cycle; MA: Menarche age; MV: Menstrual volume; DY: Dysmenorrhea; DH: Disease history; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; FUH: Fundal height; AC: Abdomen circumference; FHR: Fetal heart rate; HRF: High risk factors; BMI: Body mass index; GSL: Gestational sac length; GSW: Gestational sac width; GSH: Gestational sac height; FP: Fetal position; PMG: Placental Mature Grading; AFI: Amniotic fluid index; S/D: Systolic to diastolic; NT: Nuchal translucency; CRL: Crown-rump length; BPD: Biparietal diameter; FAC: Fetal abdomen circumference; FL: Femur length; HC: Head circumference; HGB: Hemoglobin; BLG: Blood glucose; CART: Classification and regression tree; RNN: Recurrent neural network; R^2: Coefficient of determination; MAE: Mean absolute errors; MSE: Mean square error; SVR: Support vector regression; RF: Random forest; LSTM: Long short-term memory; SVM: Support vector machine

Declarations

Ethics approval and consent to participate

This study is observational and presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside the research context. All the study procedures were approved by the ethics committee of Hangzhou Women's Hospital (Hangzhou Maternity and Child Health Care Hospital) (written permission with approval NO. 2019-02-2). All medical procedures were carried out in accordance with relevant guidelines and regulations. The informed consent requirement for this study was waived by the board because the researcher only accessed the database for analysis purposes, and all patient data were de-identified.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from Hangzhou Women's Hospital, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Hangzhou Women's Hospital.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by Primary Research and Development Plan of Zhejiang Province in China(Grant No 2020C03107) and Natural Science Foundation of Zhejiang Province in China (Grant NO GF20F020009).

Authors' contributions

YNW, YCZ contributed equally to this work. YNW and YFW were responsible for the study design. YCZ, SL and WSH extracted the data. YNW completed the relevant experiments. YFW, XZ, ZMY, WSH, SL and XYS provided feedback on analyses and interpretation of results. YCZ, YNW, YFW wrote this paper. All authors read and approved the final manuscript.

Acknowledgements

We also like to thank all pregnant women and healthcare professionals who participated in the different stage of the development of the prediction model.

References

[1] Nguyen T H, Larsen T, Engholm G, et al. Evaluation of ultrasound-estimated date of delivery in 17,450 spontaneous singleton births: do we need to modify Naegele's rule?. 1999, 14(1):23-8.

[2] W.B. Barr, C.C. Pecci. Last menstrual period versus ultrasound for pregnancy dating. 2004, 87(1):38-39.

[3] Shuang L , Xiao-Tian L I . Advances in Gestational Age Estimating Based on Last Menstrual Period and Ultrasound[J]. Journal of International Obstetrics and Gynecology, 2014,(4):448-450

[4] José Morales Roselló, David Hervás Marín, Alfredo Perales Marín. Can we predict delivery date with obstetric ultrasound?[J]. The Journal of Maternal-Fetal & Neonatal Medicine, 2013.

[5] Lawson G W . Naegele's rule and the length of pregnancy – A review[J]. Australian and New Zealand Journal of Obstetrics and Gynaecology.

[6] Committee on Obstetric Practice, the American Institute of Ultrasound in Medicine, and the Society for Maternal-Fetal Medicine. Committee Opinion No 700: Methods for Estimating the Due Date[J]. Obstetrics & Gynecology, 2017, 129(5):e150.

[7] Hunter L A . Issues in Pregnancy Dating: Revisiting the Evidence[J]. Journal of Midwifery & Womens Health, 2009, 54(3):184-190.

[8] Dietz P, Englanda, LJ., Callaghana, WM., Pearlb, M., Wierb, ML., Kharrazic, M.: A comparison of LMP-based and ultrasound-based estimates of gestational age using linked California livebirth and prenatal screening records. Paediatric and Perinatal Epidemiology 2007, 21:62– 71.

[9] Pratima Jaiswal, William Masih, Sanjay Jaiswal, et al. Assessment of fetal gestational age by ultrasonic measurement of bi-parietal diameter in the southern part of Rajasthan. Medical Journal of Dr. D.Y. Patil University, 2015, 8(1):27-30.

[10] Max Mongelli, Mark Wilcox, Jason Gardosi. Estimating the date of confinement: Ultrasonographic biometry versus certain menstrual dates. 1996, 174(1):278-281.

[11] Ka Ying Bonnie Ng, Philip J. Steer. Prediction of an estimated delivery date should take into account both the length of a previous pregnancy and the interpregnancy interval. 2016, 201:101-107.

[12] Amina Z. Khambalia, Christine L. Roberts, Martin Nguyen, et al. Predicting date of birth and examining the best time to date a pregnancy. 2013, 123(2):105-109.

[13] Liang Liang, Marie-Louise Hee Rasmussen, Brian Piening, et al. Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. 2020, 181(7):1680-1692.e15.

[14] Thomas L.A. van den Heuvel, Hezkiel Petros, Stefano Santini, et al. Automated Fetal Head Detection and Circumference Estimation from Free-Hand Ultrasound Sweeps Using Deep Learning in Resource-Limited Countries. 2019, 45(3):773-785.

[15] Russell Fung, Jose Villar, Ali Dashti, et al. Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. 2020, 2(7):e368-e375.

[16] Schink T , Wentzell N , Dathe K , et al. Estimating the Beginning of Pregnancy in German Claims Data: Development of an Algorithm With a Focus on the Expected Delivery Date[J]. Frontiers in Public Health, 2020, 8:350.

[17] Torres M T , Valstar M , Henry C , et al. Postnatal gestational age estimation of newborns using Small Sample Deep Learning[J]. Image and Vision Computing, 2019, s 83–84(MAR.-APR.):87-99.

[18] Friedman J H . Greedy Function Approximation: A Gradient Boosting Machine[J]. Annals of Statistics, 2001, 29(5):1189-1232.

[19] Juan Huan, Hui Li, Mingbao Li, et al. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China. 2020, 175(C)

[20] Zhou Z H. Ensemble methods: foundations and algorithms[M]. CRC Press, 2012:67-95.

[21] Chung J , Gulcehre C , Cho K H , et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. Eprint Arxiv, 2014.

[22] Williams R , Zipser D . A Learning Algorithm for Continually Running Fully Recurrent

Neural Networks[J]. Neural Computation, 2014, 1(2):270-280.

[23] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, et al. Recurrent Neural Networks for Multivariate Time Series with Missing Values. 2018, 8(1):581-592.

[24] Breiman L , Cutler R A . Random Forests Machine Learning[J]. Journal of Clinical Microbiology, 2001, 2:199-228.

[25] Svetnik V . Random forest: a classification and regression tool for compound classification and QSAR modeling.[J]. Journal of Chemical Information & Computer ences, 2003, 43.

[26] Chih-Chung Chang, Chih-Jen Lin. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011.

[27] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[28] Graves A . Supervised Sequence Labelling with Recurrent Neural Networks[J]. Studies in Computational Intelligence, 2012, 385.

[29] Kessler J, Johnsen SL, Ebbing C, Karlsen HO, Rasmussen S, Kiserud T. Estimated date of delivery based on second trimester fetal head circumference: A population-based validation of 21451 deliveries. Acta Obstet Gynecol Scand. 2019 Jan;98(1):101-105. doi: 10.1111/aogs.13454. Epub 2018 Oct 25. PMID: 30168856

[30] Bo ping YANG, Wei ZHOU, Li LI. Clinical analysis of the delivery date reckoned by traditional and modified formula for calculating the expected date of confinement. 2017, 42(3):230-233.

[31] Mittendorf R, Williams M A, Berkey C S, et al. Predictors of human gestational length. 1993, 168(2):480-4.

[32] Luke B . Date of Delivery in Multifetal Pregnancies[J]. Jama the Journal of the American Medical Association, 1996, 276(6):452.

Figures

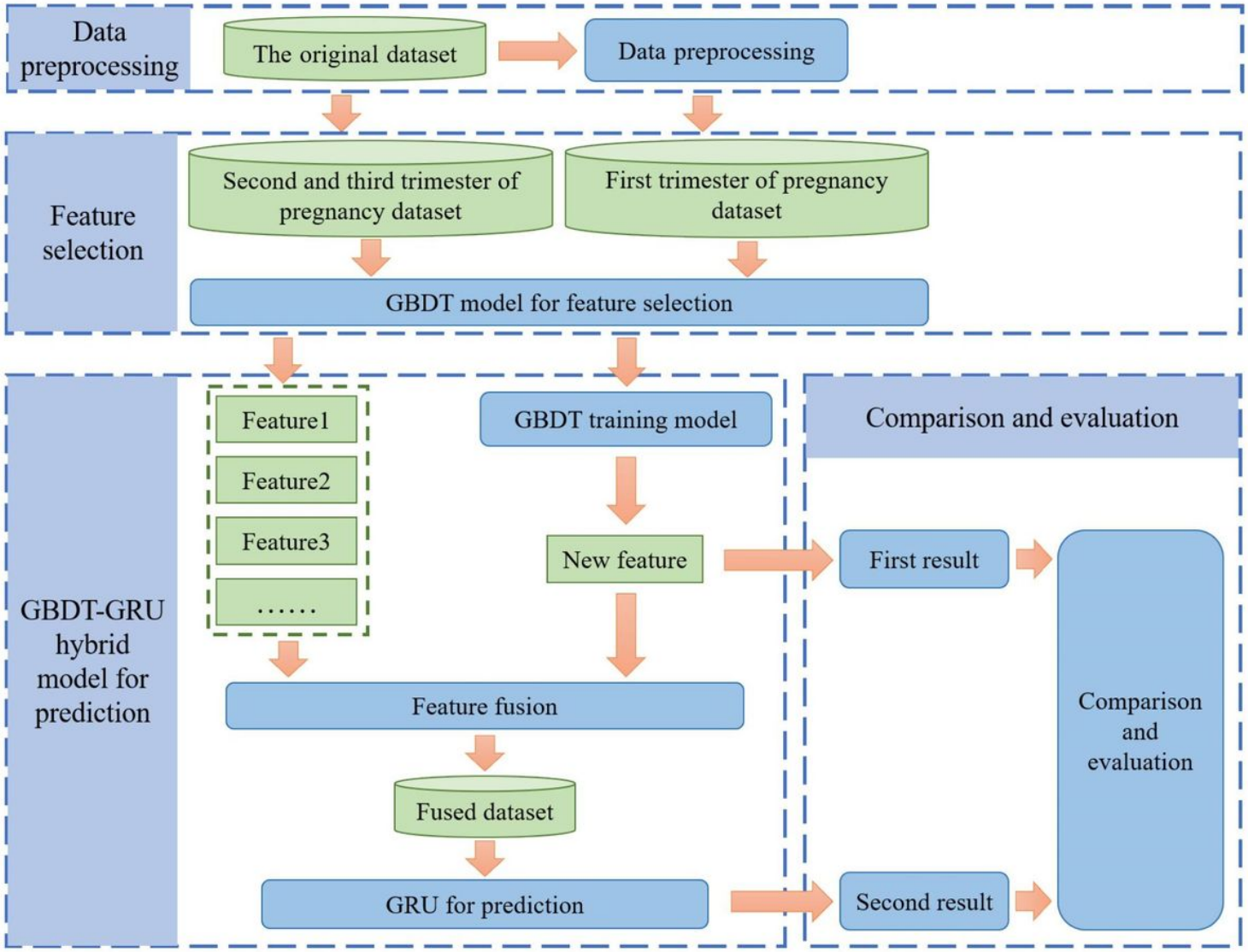


Figure 1

GBDT-GRU framework for the estimated date of delivery

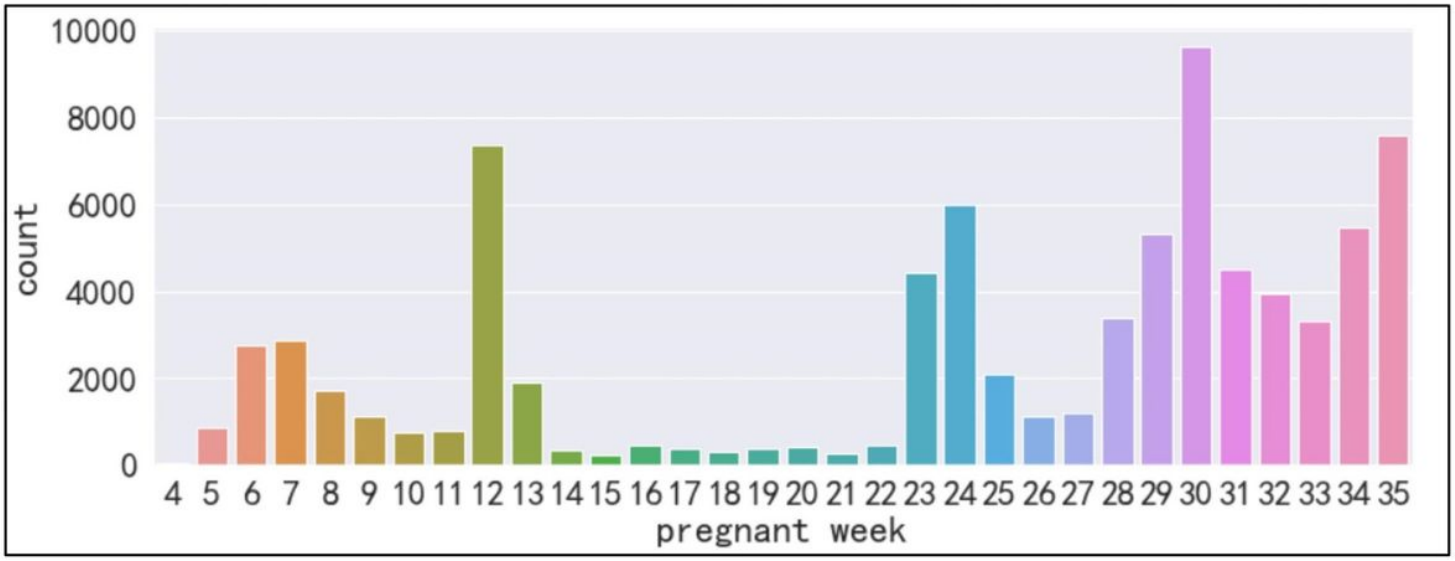


Figure 2

Count of the pregnant women in different pregnant weeks

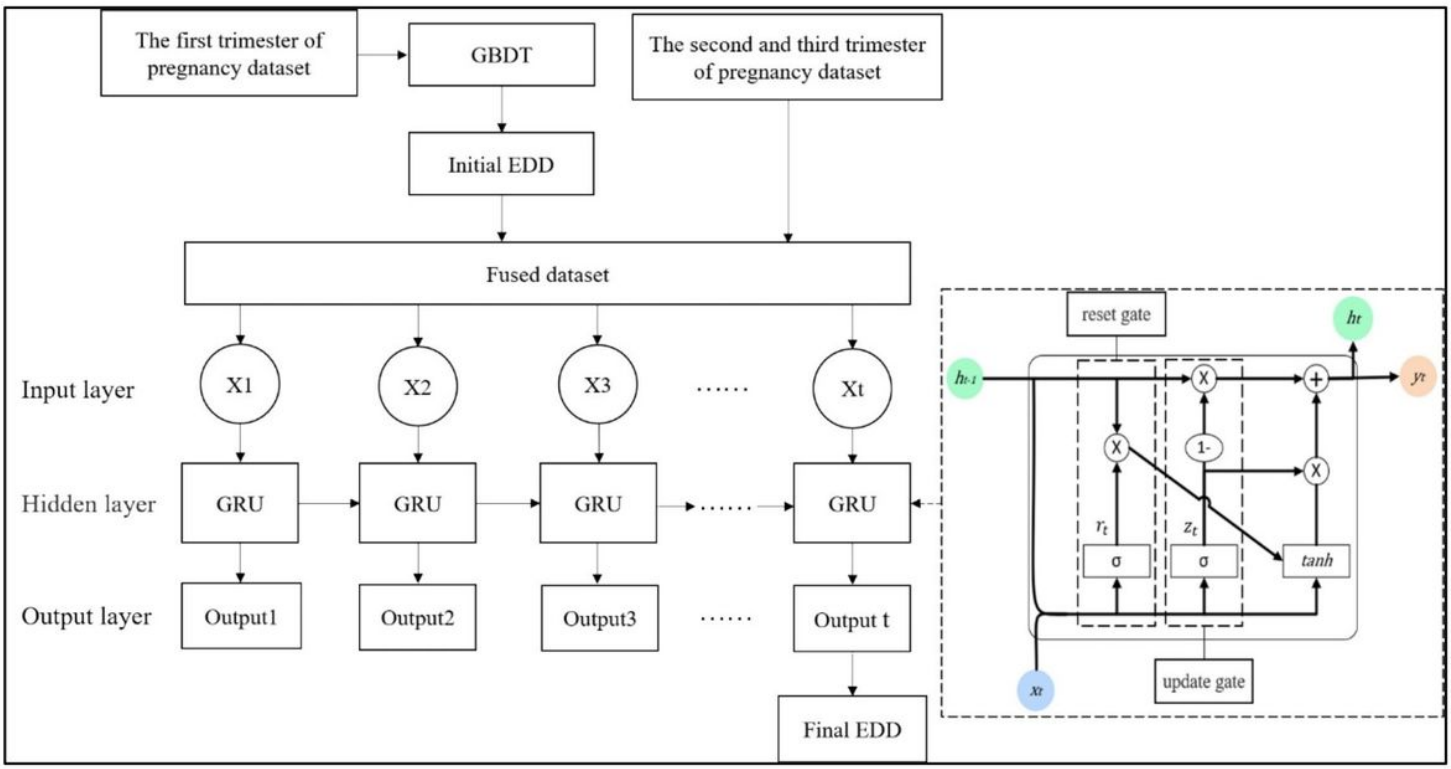


Figure 3

Structure of GBDT-GRU

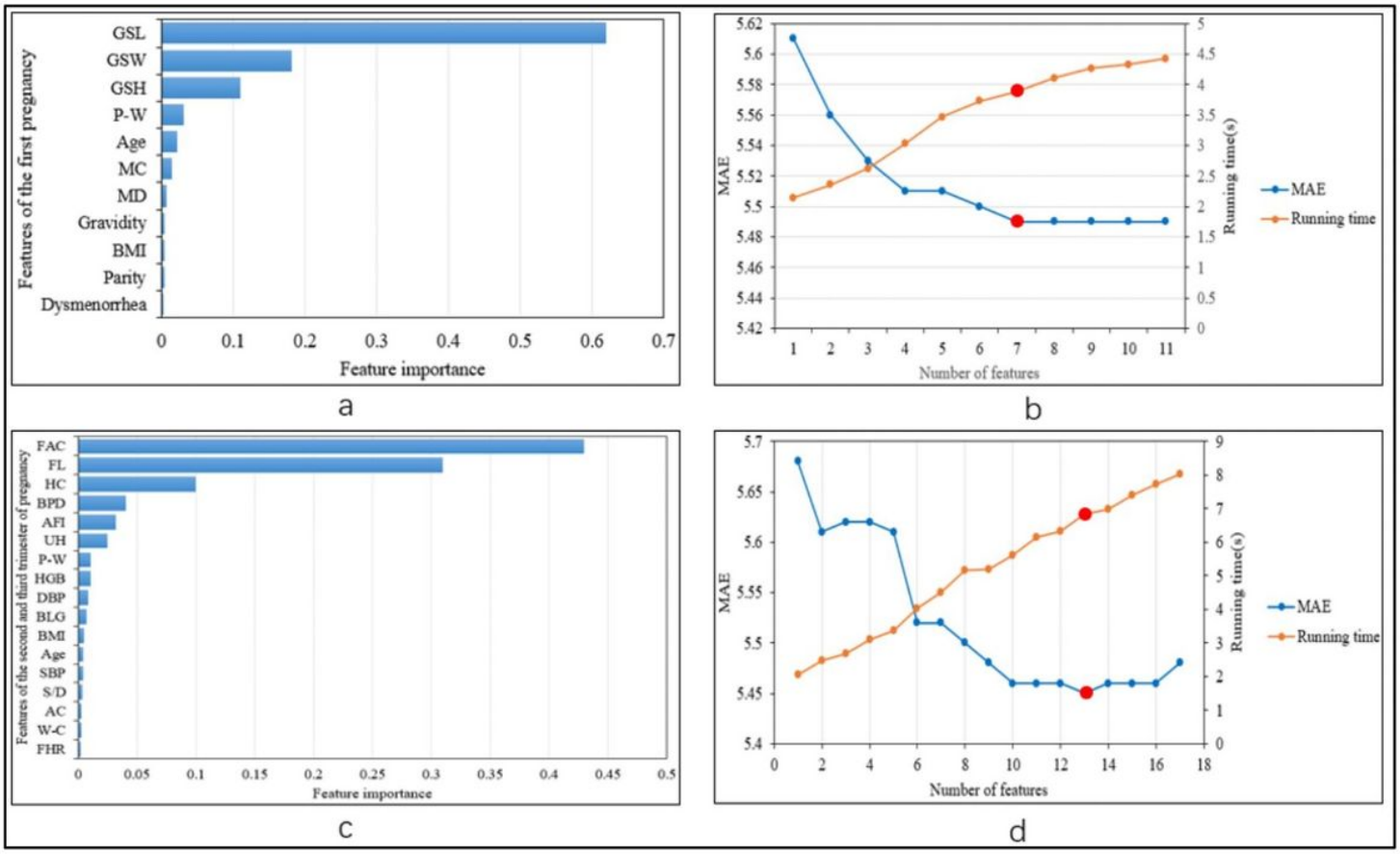
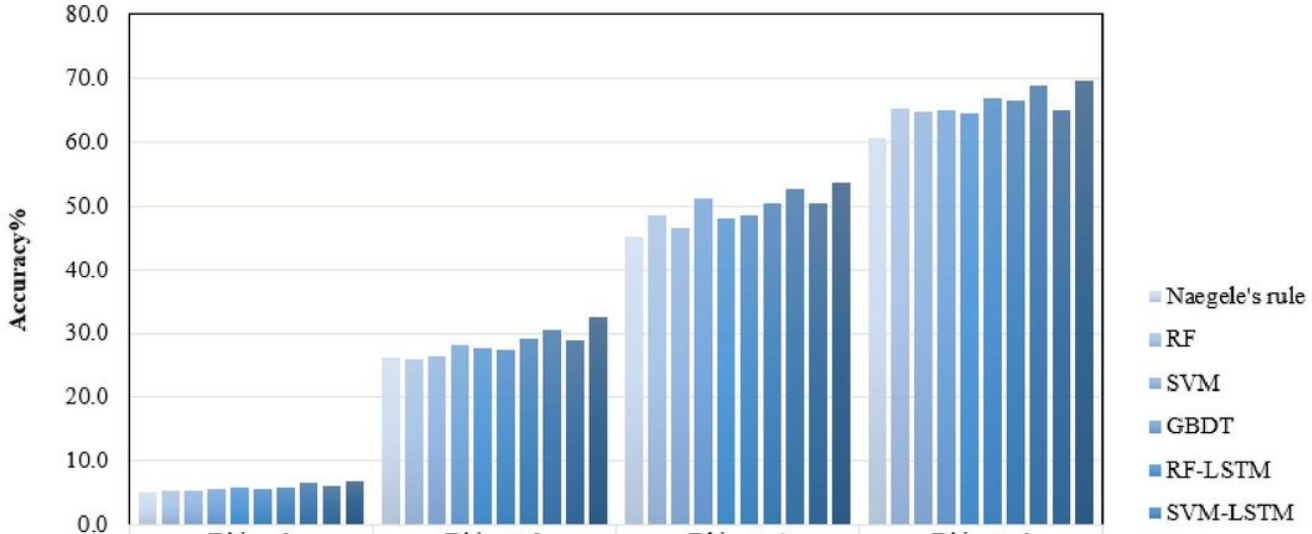


Figure 4

Analysis result for feature selection of different datasets. a shows the feature



	Dbias=0	Dbias<=2	Dbias<=4	Dbias<=6
Naegele's rule	5.1	26.3	45.3	60.7
RF	5.2	26.0	48.6	65.4
SVM	5.4	26.4	46.6	64.7
GBDT	5.7	28.2	51.1	65.0
RF-LSTM	5.9	27.7	48.1	64.5
SVM-LSTM	5.7	27.4	48.5	66.8
GBDT-LSTM	5.8	29.2	50.4	66.7
RF-GRU	6.5	30.5	52.6	68.9
SVM-GRU	6.2	29.1	50.3	65.1
GBDT-GRU	6.9	32.6	53.6	69.8

Figure 5

The accuracy of different methods under different □□□□□