

Development of a Machine Learning Model to Predict Epidural-Related Maternal Fever During Labor Analgesia: A Multi-Algorithm Comparative Study with Prospective Implementation Framework

Guoxiu Zhang¹, Yihui Yang², Rugang An², Zhengquan Tan²

¹Department of Critical Care Medicine, Zunyi First People's Hospital (Third Affiliated Hospital of Zunyi Medical College), Zunyi, 563000, People's Republic of China; ²Department of Anesthesiology, Zunyi First People's Hospital (Third Affiliated Hospital of Zunyi Medical College), Zunyi, 563000, People's Republic of China

Correspondence: Zhengquan Tan, Department of Anesthesiology, Zunyi First People's Hospital (Third Affiliated Hospital of Zunyi Medical College), No. 98 Fenghuang Road, Huichuan District, Zunyi, Guizhou, 563000, People's Republic of China, Tel +86 15985082482, Email zhengquantanzq01@126.com

Objective: To develop and validate a machine learning-based predictive model for assessing the risk of epidural-related maternal fever (ERMF)- a common complication during labor analgesia.

Methods: A prospective cohort study was conducted among 500 parturients with term singleton pregnancies who received epidural labor analgesia between September 2022 and August 2023. Key variables collected include maternal demographic characteristics, anesthesia-related indicators for complications, and obstetric features. Following application of exclusion criteria, 422 parturients were included and allocated into a modeling cohort (n = 337) and a validation cohort (n = 85) using stratified random sampling at an 8:2 ratio. Eleven machine learning algorithms were utilized to construct predictive models. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), accuracy, precision, recall, and F1-score.

Results: The incidence of ERMF was 28.1% (119/422). Among the evaluated algorithms, Categorical Boosting (CatBoost) demonstrated the highest performance, with an AUC of 0.94 (95% CI: 0.86–0.98), accuracy of 90.59%, precision of 0.88, and average precision (AP) of 0.86 in the validation cohort. Analysis using SHapley Additive exPlanations (SHAP)-an interpretable artificial intelligence method- identified prolonged duration of rupture of membranes, higher maternal body mass index, and nulliparity as the top predictors of ERMF risk. An interactive web-based interface was developed to facilitate real-time clinical risk evaluation.

Conclusion: A machine learning model with high discriminative ability was constructed to predict the risk of ERMF. The CatBoost algorithm effectively identified parturients at elevated risk, and the accompanying visual tool offers evidence-based support for stratified management of intrapartum fever in clinical practice.

Keywords: epidural analgesia for labor, fever, intrapartum, labor analgesia, machine learning, predictive model

Introduction

Epidural labor analgesia (epidural analgesia; EA) is a widely utilized analgesic technique in obstetric practice. However, its benefits are partially offset by a significant complication—epidural-related maternal fever (ERMF). In this study, ERMF specifically refers to maternal fever (oral temperature $\geq 38.0^{\circ}\text{C}$) associated with epidural analgesia during labor, after excluding other causes such as confirmed infection. The incidence of epidural-related maternal fever (ERMF) during intrapartum EA has been reported to range from 15% to 34.1%,¹ with some studies noting occurrences in about 20% of cases and up to 33%. The occurrence of intrapartum fever has been associated with increased use of antibiotics,

prolonged labor duration, and elevated risks of fetal intrauterine distress and neonatal asphyxia.² Additionally, ERMF has been linked to adverse neonatal outcomes, including early-onset sepsis and cerebral palsy.³ Current preventive strategies such as regulating delivery room temperature, minimizing the frequency of vaginal examinations, and administering low-concentration local anesthetics, have demonstrated limited effectiveness and do not offer individualized risk assessment. Consequently, the identification and clinical management of ERMF have become central areas of investigation. Recent studies have begun to elucidate the multifactorial nature of ERMF and explore advanced predictive models, with the aim of shifting toward more personalized approaches.

New Evidence on Pathophysiological Mechanisms and Risk Factors

The development of ERMF is influenced by a complex interplay of factors. The pathophysiology of ERMF is recognized as a non-infectious process, primarily driven by a sterile inflammatory response and altered central thermoregulation.⁴⁻⁶ The “sterile inflammation hypothesis” is supported by a pilot study which found that parturients with ERMF exhibited a 1.5-fold increase in interleukin-6 levels, suggesting a systemic inflammatory response. Intriguingly, this study also observed decreasing levels of circulating free mitochondrial DNA (mtDNA) during labor in the ERMF group, proposing mtDNA as a potential trigger for this sterile inflammatory process.⁷ The “Altered Thermoregulation Hypothesis” suggests that the neuraxial blockade itself disrupts central thermoregulatory set points, leading to an imbalance between heat production and loss. Notably, a retrospective study found no significant difference in the incidence of maternal fever between patients receiving continuous spinal analgesia and those receiving epidural analgesia, despite substantially lower local anesthetic doses in the former group. This implies that the fever may be more related to the central neural blockade itself rather than solely to systemic drug effects.⁸

This pathophysiological complexity underscores a pressing clinical need for proactive, individualized risk assessment. This pathophysiological complexity underscores a pressing clinical need for proactive, individualized risk assessment. Current preventive measures are non-specific and of limited efficacy. Crucially, risk is dynamic, evolving with factors like the duration of rupture of membranes and the cumulative dose of local anesthetics. A key multivariate logistic regression analysis identified that first-time mothers, using epidural anesthesia pain during labor, premature rupture of membranes, second stage of labor lasting over 120 minutes, grade III meconium staining, and fetal weight exceeding 4000g are independent risk factors for ERMF.⁹ Other corroborated factors include premature rupture of membranes, gestational diabetes, and a higher number of vaginal examinations.¹⁰ Furthermore, technical aspects of analgesia are being explored. For instance, research is ongoing to evaluate whether techniques like Programmed Intermittent Epidural Bolus compared to traditional continuous infusion might influence fever risk by reducing the total local anesthetic consumption.¹¹

Limitations of Existing Predictive Models and the Opportunity for Machine Learning

The pursuit of accurate prediction for ERMF has seen various approaches, each with limitations. Although nomogram-based approaches have been developed to estimate the risk of intrapartum fever, their validation has been hampered by inadequate data collection.⁹ While the perfusion index (PI) was initially proposed as a potential indicator for fever prediction, a complete predictive model has not yet been established.¹² Many existing predictive models are constrained by limited data scope and lack the capacity for dynamic monitoring, reducing their utility in real-time clinical contexts. They fail to integrate dynamic information about risk factors—such as the cumulative dose of local anesthetics or the prolonged duration of analgesia—as these factors evolve over time. In contrast, modern Machine Learning techniques present a compelling opportunity to overcome these hurdles. The choice of ML is particularly apt for this research problem for several reasons:

High-Dimensional Data Integration: ML algorithms excel at integrating and finding complex, non-linear patterns within high-dimensional, multi-source data, including demographic characteristics, obstetric history, real-time vital signs, analgesic parameters, and laboratory markers.

Temporal Dynamics and Real-Time Prediction: Unlike static models, specific ML models (eg, Recurrent Neural Networks or hybrid architectures like CNN-LSTM) are inherently suited for analyzing time-series data. This study employs machine learning methods to address the ERMF prediction problem, offering the following unique advantages:

Predictive Performance and Explainability: ML models can achieve superior predictive accuracy. Furthermore, the integration of Explainable AI methods, such as SHAP,^{13,14} can illuminate the model's decision-making process by identifying the key factors driving a high-risk prediction for an individual patient, thereby enhancing clinical trust and adoption. This study employs a predictive modeling framework, with its core objective being the development of a machine learning model—a computer algorithm capable of automatically learning patterns from data and making predictions. The framework aims to identify predictors (ie, risk factors) associated with ERMF occurrence and construct a predictive tool that can accurately assess an individual's risk of developing the condition. It is important to emphasize that this method reveals statistical associations between variables, not causal relationships. The present study aimed to construct a machine learning-based predictive model for ERMF by analyzing and integrating a wide range of relevant clinical variables. The goal is to facilitate early identification of patients at high risk for intrapartum fever, thereby supporting timely intervention and potentially improving maternal and neonatal outcomes.

Materials and Methods

General Data

Clinical data were prospectively collected from 422 parturients who received epidural labor analgesia at Zunyi First People's Hospital (Third Affiliated Hospital of Zunyi Medical College) between September 2022 and August 2023. Inclusion criteria were: (1) term singleton pregnancy with cephalic presentation, (2) American Society of Anesthesiologists (ASA) physical status classification I or II, and (3) voluntary request for epidural labor analgesia. Exclusion criteria were: (1) body temperature exceeding 38 °C prior to initiation of epidural analgesia, (2) presence of infectious conditions such as thoracic or abdominal cavity infections, urinary tract infections, respiratory tract infections, or other fever-associated conditions, (3) high-risk pregnancies, (4) administration of non-steroidal anti-inflammatory drugs, (5) cognitive or functional impairments, language or communication disorders, intellectual disabilities, or epilepsy, and (6) allergy to anesthetic agents. Additional exclusion criteria included epidural puncture failure, cesarean section due to any indication, stillbirth, or fetal malformations. This study followed the Declaration of Helsinki (as revised in 2013) and was approved by the hospital's ethics committee (Ethics Approval Number: 2022–147). Written informed consent was obtained from all participants.

Analgesia and Measurement Methods

Maternal vital signs were routinely monitored. All epidural puncture procedures at this institute are performed by anesthesiologists with over five years of clinical experience, following standardized operating procedures:

Preoperative Preparation: Position the patient in the lateral decubitus position and select the L2-L3 or L3-L4 intervertebral space as the puncture site.

Preoperative Preparation: Position the patient in the lateral decubitus position and select the L2-L3 or L3-L4 intervertebral space as the puncture site.

Catheter Fixation: Secure the catheter properly. After aspirating to confirm the absence of blood or cerebrospinal fluid, inject 3 mL of 1.5% lidocaine with epinephrine as a test dose.

Analgesia Initiation and Maintenance: After confirming safety, administer 8–15 mL of a loading dose consisting of 0.08%–0.1% ropivacaine combined with 0.4–0.5 µg/mL sufentanil. Subsequently, connect a programmable infusion pump to maintain analgesia using a combination of programmed intermittent epidural pulse infusion and patient-controlled analgesia (PCA) mode. Epidural infusion was discontinued following completion of the third stage of labor, and the catheter was removed after perineal suturing. Oral intake was permitted during labor. Maternal body temperature was continuously monitored throughout the duration of epidural analgesia.

Primary Outcome Definition

Intrapartum fever was defined as a single oral temperature measurement of ≥ 38 °C prior to delivery. The reported incidence of fever in patients receiving epidural labor analgesia ranges from 2.4% to 36.7%.¹

Data Collection

Based on clinical expertise and previous literature, 55 variables considered relevant to the development of the ERMF prediction model were selected. The variables encompassed the following domains: General information: Age, body mass index (BMI), baseline body temperature, number of pregnancies, deliveries, miscarriages, pregnancy preservations, gestational age, and fetal weight); Comorbidities: Diabetes mellitus, hypertension, hypothyroidism, anemia, hypoproteinemia, vaginitis, cervical erosion, and human papillomavirus (HPV) infection; Laboratory parameters: Hemoglobin level, white blood cell count, neutrophil count, neutrophil count and percentage, lymphocyte percentage, alanine aminotransferase, aspartate aminotransferase, albumin, prealbumin, serum creatinine, potassium, and sodium; Anesthesia-related data: Visual analog scale (VAS) score for analgesia, opioid usage and type, electronic infusion pump pulse dose, duration of labor analgesia, and dosage of ropivacaine; Obstetric data: Delivery room temperature, cervical dilation at the initiation of analgesia, number of vaginal examinations, use of oxytocin, antibiotics, pethidine, or magnesium sulfate during labor, presence of a nuchal cord, premature rupture of membranes, degree of amniotic fluid contamination, and use of forceps-assisted delivery.

Data Cleaning and Preprocessing

Missing Value Imputation and Sensitivity Analysis: Among the 55 predictor variables included in this study, the majority (approximately 98%) exhibited good data completeness, with a missing rate below 2%. To address missing data, this study employed the following strategies: categorical variables were imputed using the mode, continuous variables were imputed using the median (to mitigate the impact of skewed distributions), and missing indicator variables were created for all imputed variables to capture information about the missing patterns. To assess potential bias from imputation strategies, we conducted sensitivity analysis: five complete datasets were generated using multiple imputation. CatBoost models were trained on both the original imputed data and the multiple imputed datasets. Results showed negligible model performance differences ($AUC < 0.015$), indicating the imputation strategy used in this study has negligible impact on model training. **Standardized Processing:** This study standardized continuous variables using Z-scores. For categorical variables, algorithm-specific processing was applied: the CatBoost model employed label encoding and was declared as a categorical feature; other algorithms uniformly used one-hot encoding. To address the issue of class imbalance, this study employs the Synthetic Minority Over-sampling Technique (SMOTE),¹⁵ as illustrated in the flowchart (Figure 1). The application of SMOTE follows a rigorous process to prevent data leakage: 1) Timing of application: After completing the training/test set split (7:3), SMOTE is applied only within the training set. During five-fold cross-validation, synthetic samples are generated only in the training portion of each fold, while the validation portion retains the original distribution. The test set is never involved in this process. 2) Sampling Parameters: The target minority-majority ratio is set to 1:1, ensuring balanced positive and negative samples in the training set. This process guarantees unbiased model evaluation while enhancing the model's recognition capability for minority classes.

Machine Learning Methods

We employ multiple machine learning algorithms to build models, including Bootstrap aggregating (Bagging), extreme trees, random forests, adaptive boosting, gradient boosting, extreme gradient boosting, lightweight extreme gradient boosting, logistic regression, neural networks, categorical gradient boosting, and voting models—totaling 11 ensemble learning algorithms.

Bagging: Bagging¹⁶ enhances model accuracy and stability by constructing multiple independent models and averaging or voting on their predictions. Its core principle involves using bootstrap sampling to perform random selection with replacement from the original training data, thereby creating multiple distinct training datasets. Independent models are then trained on each dataset. The primary advantage of bagging is that it enhances a model's ability to generalize to unseen data while reducing the likelihood of the model overfitting the training data.

Extreme Trees:¹⁷ Also known as Extra Random Trees, they are a variant of decision trees. Unlike conventional decision trees, Extreme Trees perform splits at each decision node using a random subset of features rather than selecting

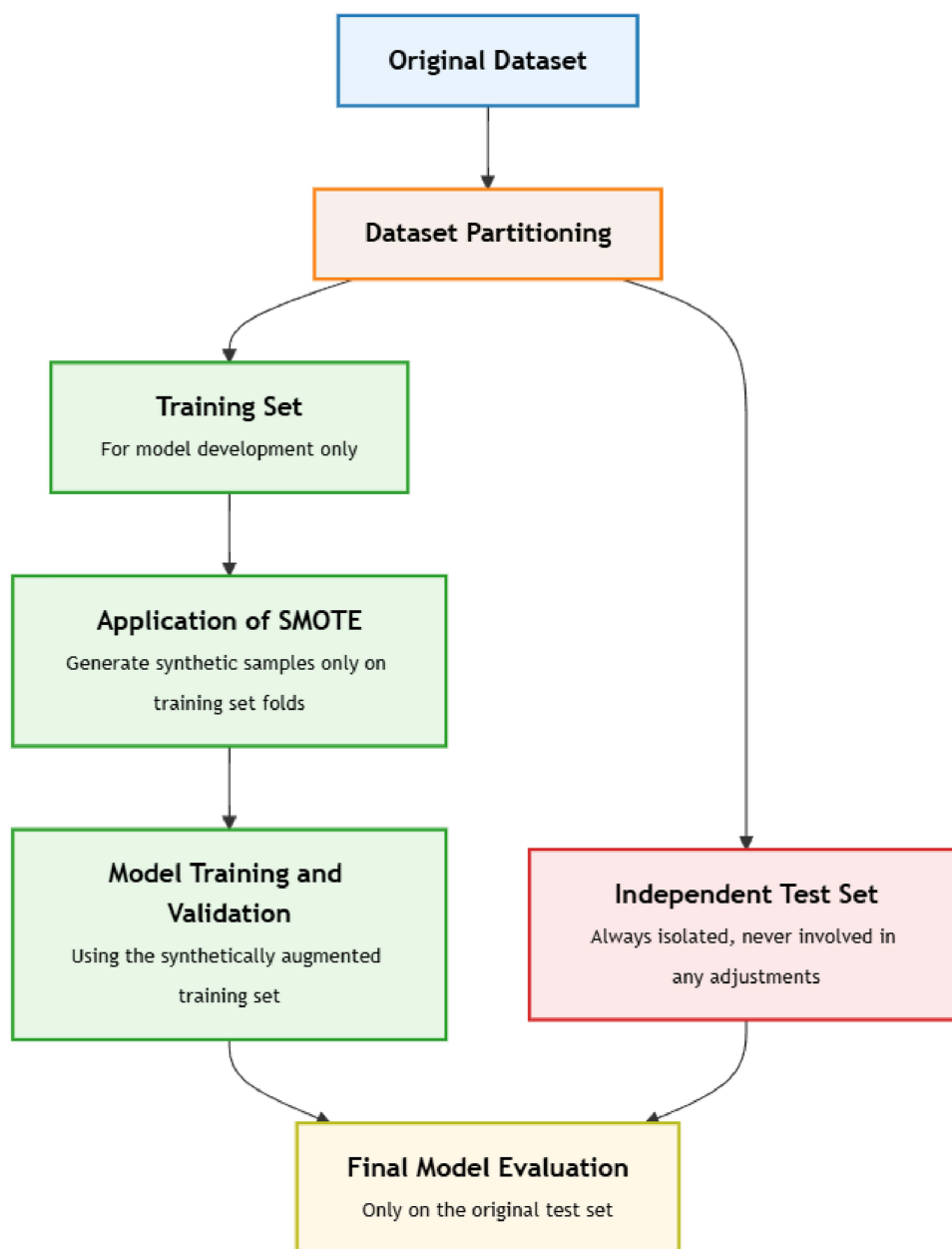


Figure 1 Model construction flowchart.

the optimal feature for splitting. They offer fast training speeds, perform well on high-dimensional data and large-scale datasets, and exhibit strong generalization capabilities and robustness.

Random Forest (RF): RF is an ensemble learning method composed of multiple decision trees. It constructs each tree using bootstrap sampling from the training data and splits nodes based on a random subset of features. Key characteristics of RF include:¹⁸ 1) **Randomness:** It introduces two forms of randomness—bootstrap sampling for training data and random feature selection at each node—enhancing model diversity and reducing feature correlation. 2) **Compatibility with High-Dimensional Data:** RF efficiently handles datasets with numerous features by considering only a subset of features for each tree. 3) **Strong Generalization Ability:** By aggregating predictions from multiple trees, it mitigates overfitting and improves generalization. 4) **Interpretability:** RF provides feature importance rankings, offering insights into the contributions of different features to predictions.

Adaptive Boosting (AdaBoost): AdaBoost is an iterative machine learning algorithm widely used in addressing classification and regression problems. The advantages of AdaBoost¹⁹ include: 1) High performance: AdaBoost constructs a strong classifier by combining multiple weak classifiers, thereby enhancing the overall model's performance. 2) Adaptability: AdaBoost progressively improves model performance by adjusting sample weights to focus on misclassified samples. 3) Interpretability: AdaBoost provides the weight of each weak classifier, aiding our understanding of each classifier's contribution to the final prediction.²⁰ All analyses were conducted using Python version 3.9.13 in the Visual Studio Code environment. Model performance was evaluated by calculating the receiver operating characteristic (ROC) curve and the area under the precision-recall curve (AUPRC), enabling comprehensive evaluation across the entire recall spectrum.

Model Optimization and Hyperparameter Tuning

This study employs Bayesian optimization as a hyperparameter search strategy, intelligently exploring the hyperparameter space by constructing a probabilistic surrogate model—a more efficient approach than traditional grid search. The optimization process is based on five-fold nested cross-validation: the inner loop evaluates hyperparameter performance through cross-validation to guide the optimization, while the outer loop assesses the final model's generalization capability. The optimization objective maximizes the AUC-ROC on the validation set while balancing accuracy and recall to ensure the model possesses robust clinical discrimination capability.

Model Validation and Cross-Validation Strategy

The complete dataset comprising 422 cases was randomly divided into a training set ($n = 337$) and a test set ($n = 85$) using an 8:2 ratio. The training set was used to construct the predictive models, and the test set was used for performance evaluation. Within the training set, perform five-fold cross-validation: Divide the training set evenly into five subsets. Train the model using four subsets in rotation, while using one subset for validation. Repeat this process five times to ensure each subset serves as the validation set once. The final model performance estimate is the average of the five validation runs.

SHAP Method for Feature Importance

Feature importance was evaluated using SHapley Additive exPlanations (SHAP) version 0.41.0, and visualizations were generated with the Local Interpretable Model-Agnostic Explanations (LIME) library implemented in Python. The magnitude of each feature's contribution to the model output was represented by the length of the corresponding bar, and the direction of the contribution was indicated by bar color.

Statistical Analysis

Statistical analyses were conducted using SPSS version 29.0. Measurement data were expressed as mean \pm standard deviation ($\bar{x} \pm s$), and between-group comparisons were performed using independent samples *t*-tests. Categorical data were expressed as frequencies (percentages), and between-group comparisons were analyzed using the chi-squared (χ^2) test. A *p*-value less than 0.05 was considered statistically significant.

Results

Comparison of Baseline Characteristics Between Groups

A total of 500 parturients were initially enrolled in the study. After applying exclusion criteria—75 cases requiring cesarean section, 2 cases of stillbirth, and 1 case involving neonatal malformation—422 participants were included in the final analysis. Among these, 119 developed intrapartum fever, corresponding to an incidence of 28.1%, while 303 did not develop intrapartum fever. A total of 55 variables were included in the comparative analysis. Statistically significant differences ($p < 0.05$) were observed between the fever and non-fever groups in the following variables: BMI, baseline body temperature, presence of hypoproteinemia, use of oxytocin, antibiotics, and pethidine, premature rupture of membranes, number of vaginal examinations, duration of first and second stages of labor, total labor duration, degree of amniotic fluid contamination, gravidity, parity, duration of analgesia, and total analgesic dosage. A detailed summary of these comparisons is provided in [Table 1](#).

Table 1 Comparison of Baseline Characteristics Between the Two Groups

Factor	Fever Group (n=119)	Control Group (n=303)	Statistic	P value
Age (y)	28(25, 30)	28(26, 30)	1.003	0.316
BMI (kg/m ²)	27.23(24.80, 30.00)	25.80(24.44, 27.50)	3.942	<0.001*
Baseline temp (°C)	36.5(36.3, 36.7)	36.4(36.2, 36.6)	2.564	0.008*
WBC (g/L)	8.9(7.5, 10.4)	8.8(7.5, 10.2)	0.536	0.592
Neutrophils (g/L)	6.6(5.7, 8.0)	6.6(5.5, 7.8)	0.794	0.427
Neutrophil %	76.3(72.3, 79.9)	75.6(72.2, 78.9)	1.019	0.308
Lymphocyte %	17.3(13.9, 20.1)	17.3(14.7, 20.8)	0.722	0.471
AST (U/L)	20(16, 25)	20(17, 26)	1.056	0.291
ALT (U/L)	10(8, 14)	11(8, 14)	1.008	0.313
Albumin (g/L)	32.8(30.5, 35.4)	32.6(30.2, 34.5)	1.562	0.118
Opioids [n(%)]			4.193	0.241
None (0)	42 (35.3)	91 (30.0)		
Fentanyl (1)	6 (5.0)	11 (3.6)		
Hydromorphone (2)	4 (3.4)	25 (8.3)		
Forceps delivery [n(%)]	6 (5.0)	4 (1.3)	5.116	0.034*
Fetal weight (g)	3300 (3050, 3500)	3200 (3000, 3400)	2.406	0.016*
Gravidity	1 (1, 2)	2 (1, 3)	2.258	0.024*
Parity	0 (0, 0)	0 (0, 1)	3.539	<0.05*
History of tocolysis	0 (0, 0)	0 (0, 0)	4.037	0.682
Analgesia duration (min)	400 (250, 625)	300 (195, 440)	4.289	<0.05*
Analgesia dose (mL)	40 (30, 64)	56 (34, 88)	4.037	<0.05*
Hypoproteinemia [n(%)]	81 (68.1)	240 (79.2)	5.825	0.016*
Oxytocin [n(%)]	89 (74.8)	149 (49.2)	22.798	<0.001*
Antibiotics [n(%)]	18 (15.1)	13 (4.3)	14.739	<0.001*
Pethidine [n(%)]	25 (21.0)	7 (2.3)	42.627	<0.001*
Nuchal cord [n(%)]	33 (27.7)	88 (29.0)	0.072	0.789
Premature rupture [n(%)]	32 (26.9)	53 (17.5)	4.693	0.030*
Fetal deceleration [n(%)]	68 (57.1)	144 (47.5)	3.162	0.075
Catheterization [n(%)]	65 (54.6)	137 (45.2)	3.030	0.082
Cervical dilation (cm)	2 (2, 2)	2 (2, 2)	1.522	0.128
Vaginal exams (count)	7 (5, 10)	6 (4, 7)	5.681	<0.05*
First stage (min)	555 (360, 760)	450 (315, 610)	3.359	<0.001*
Second stage (min)	44 (27, 67)	31 (17, 48)	4.681	<0.001*
Total labor (min)	615 (420, 870)	495 (350, 655)	4.361	<0.05*
Gestational age(d)	277 (274, 282)	277 (272, 280)	1.328	0.184
Amniotic fluid index [n(%)]			40.041	<0.05*
Normal (0)	74 (62.2)	251 (82.8)		
I degree (1)	10 (8.4)	30 (9.9)		
II degree (2)	15 (12.6)	15 (5.0)		
III degree (3)	20 (16.8)	7 (2.3)		

Note: *P<0.05.

Feature Variable Selection

Feature selection was conducted using Lasso regression analysis. Twenty variables were retained as the most informative predictors for model construction. These included aspartate aminotransferase, alanine aminotransferase, albumin, pre-albumin, BMI, baseline body temperature, delivery room temperature, number of vaginal examinations, use of oxytocin, antibiotics, pethidine, and magnesium sulfate, premature rupture of membranes, second stage of labor, degree of amniotic fluid contamination, cervical erosion, HPV infection, fetal weight, duration of analgesia, and analgesic dosage.

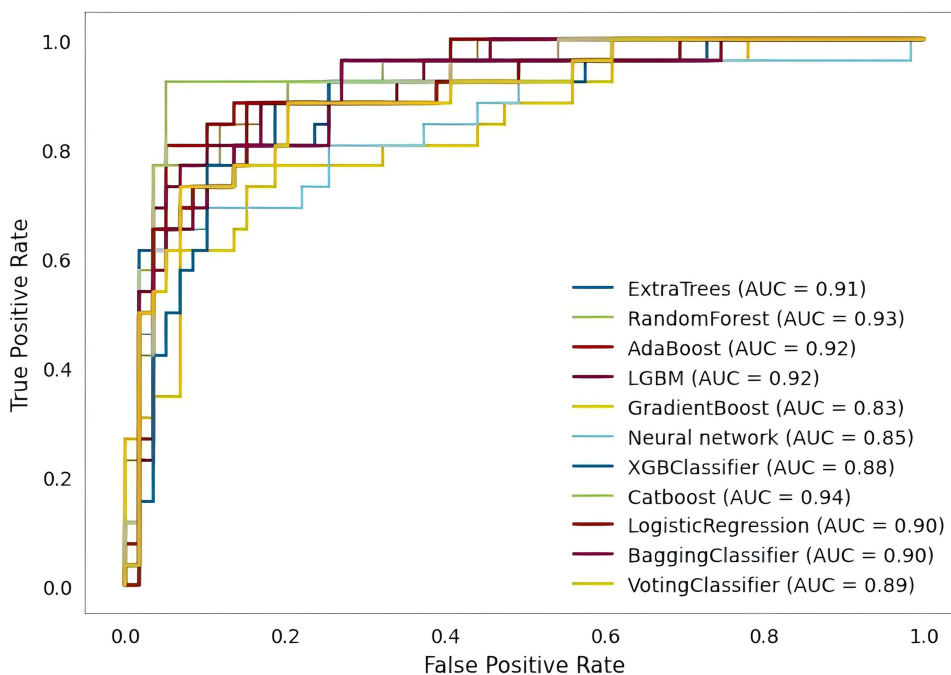


Figure 2 The receiver operating characteristic (ROC) curves of the 11 predictive models for epidural-related maternal fever (ERMF).

Predictive Model Construction and Performance Evaluation

The dataset was randomly divided into a training set comprising 80% of the cases (n = 337; fever group, n = 93; non-fever group, n = 244) and a test set comprising 20% of the cases (n = 85; fever group, n = 26; non-fever group, n = 59). The CatBoost algorithm demonstrated the highest predictive performance among the models evaluated. The ROC curve was used to assess model discrimination. The AUC for the CatBoost model in the validation cohort was 0.94,^{9,21} with a statistically significant result ($p < 0.001$) (Figure 2). The model also achieved an F1 score of 0.84, accuracy of 0.91, precision of 0.88, and recall of 0.81. These results indicated that the CatBoost model provided the best performance for predicting ERMF. Detailed performance metrics for all evaluated models are presented in Table 2.

Model Interpretability Analysis

To enhance clinical utility, a web-based visualization interface was developed based on the CatBoost model. This platform incorporates 20 predictive indicators, including BMI, degree of amniotic fluid contamination, number of vaginal examinations, premature rupture of membranes, and duration of second stage of labor. Clinicians can input patient-

Table 2 Performance Evaluation of the 11 Predictive Models for ERMF

	Accuracy	AUC	Confidence Interval	Recall	Precision	F1
Extra Trees	0.86	0.91	0.75~0.94	0.73	0.79	0.76
RandomForest	0.41	0.93	0.68~0.97	1.00	0.34	0.51
AdaBoost	0.35	0.92	0.68~0.94	1.00	0.32	0.49
LGBM	0.64	0.92	0.49~0.84	1.00	0.46	0.63
GradientBoost	0.41	0.83	0.59~0.93	1.00	0.34	0.51
Neural network	0.82	0.85	0.65~0.93	0.65	0.85	0.74
XGB	0.45	0.88	0.49~0.88	1.00	0.36	0.53
Catboost	0.91	0.94	0.70~0.96	0.81	0.88	0.84
Logistic Regression	0.85	0.90	0.71~0.91	0.73	0.76	0.75
Bagging	0.84	0.90	0.64~0.92	0.73	0.73	0.73
Voting	0.51	0.89	0.71~0.92	1.00	0.38	0.55

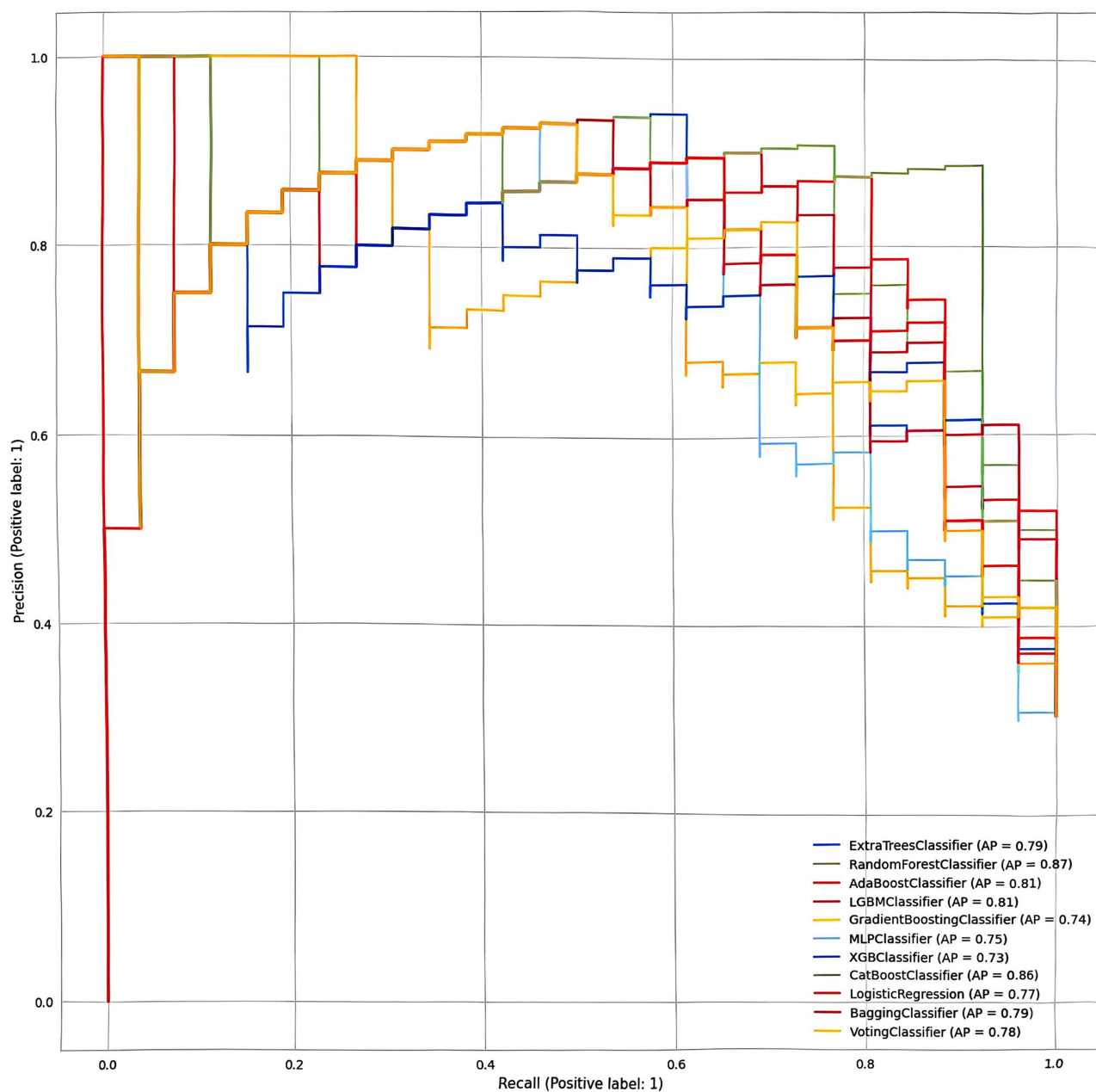


Figure 3 Local interpretability of the CatBoost model for epidural-related maternal fever (ERMF) risk prediction.

specific data into the interface, which then generates a real-time probability estimate of intrapartum fever. The platform also provides local feature interpretation, which displays the contribution of each variable to the predicted risk of intrapartum fever for individual cases. This functionality is designed to assist clinicians in identifying parturients at high-risk, thereby facilitating timely, personalized interventions aimed at reducing the incidence of maternal fever and improving associated neonatal outcomes (Figure 3).

Discussion

Construction of the ERMF Predictive Model: Scientific Rigor and Clinical Relevance

ERMF represents a multifactorial and complex physiological condition involving an interplay of pharmacological influences, physiological alterations, immune activation, and potential infection. Although its underlying mechanisms

remain incompletely defined, existing evidence has linked ERMF with adverse maternal and neonatal outcomes. Consequently, efforts to reduce the incidence of ERMF continues to represent a critical focus in obstetric anesthesia. The application of machine learning enables comprehensive integration of these complex factors, allowing the identification of latent relationships that may not be readily discernible through conventional analytical methods. Significant predictive variables identified by the model should be regarded as strong statistical associations with the outcome measure, but such associations do not directly equate to causal effects. These associations may be influenced by potential confounding factors, limitations in data collection, or complex interactions among variables.

In the present study, the machine learning framework allowed for the development of a predictive model capable of quantifying the relative contribution of multiple variables, such as anesthetic type, labor stage, or baseline maternal temperature, to the risk of intrapartum fever. This predictive model enables real-time estimation of ERMF risk during labor, serving as a clinical decision-support tool that aids in the early identification of parturients at elevated risk and the formulation of targeted interventions.

Given the considerable interindividual variability in physiological status, comorbidities, and anesthetic responses among parturients, a machine learning-based approach offers individualized risk prediction, enabling tailored clinical management strategies that move beyond uniform protocols. Such personalized assessment supports the delivery of more effective and safer care. Early identification of high-risk parturients also facilitates timely preventive measures that may reduce the incidence of postpartum fever-related complications, decrease hospital stay durations, avoid unnecessary emergency interventions, optimize resource utilization, and enhance the overall efficiency of hospital operations.

Analysis of Risk Factors Identified by the ERMF Predictive Model

BMI and Degree of Amniotic Fluid Contamination as Independent Risk Factors for ERMF

Previous findings have indicated that BMI in parturients functions as an independent risk factor for ERMF, with an increasing BMI associated with a heightened risk of intrapartum fever.²² The present findings are consistent with this association. Elevated BMI may lead to greater metabolic heat production, while an increased thickness of subcutaneous adipose tissue can impair heat dissipation, leading to progressive accumulation of heat and subsequent elevation in core body temperature.

These outcomes indicate that elevated BMI is strongly associated with intrapartum fever, warranting increased clinical attention toward this population. In addition, the degree of amniotic fluid contamination was found to be positively correlated with the occurrence of fever, with more severe contamination linked to a higher likelihood of intrapartum fever among parturients.^{21,23} It is hypothesized that severe contamination permits bacterial, viral, or other microbial infiltration of the uterus and fetus via the amniotic fluid, thereby triggering the release of inflammatory mediators, such as cytokines and inflammatory proteins, which in turn initiate an inflammatory response.

Premature Rupture of Membranes, Duration of Labor, and Frequency of Vaginal Examinations as Contributory Factors

The fetal membrane serves as a protective barrier critical for maintaining a sterile intrauterine environment. Normally, rupture occurs once cervical dilation is near completion. When rupture precedes the onset of labor, it is classified as a premature rupture of membranes (PROM). Subsequent vaginal examinations following PROM may result in mechanical disruption of the vaginal mucosa, thereby compromising the natural protective barrier and increasing the risk of intrauterine infection, including chorioamnionitis and fever. In the present analysis, PROM and a frequency of more than six vaginal examinations were both associated with a higher likelihood of intrapartum fever, findings that align with previous research.^{24,25} Furthermore, evidence supports that prolonged duration of the first and second stages of labor was significantly associated with maternal fever during labor. Extended labor duration typically necessitates more frequent vaginal examinations which may increase infection risk and exacerbate inflammatory responses.²² These findings underscore the need for obstetricians to weigh the diagnostic utility of repeated vaginal examinations against their potential to contribute to maternal discomfort and risk of infection.

Association Between Oxytocin, Pethidine Use, and Intrapartum Fever

Oxytocin, while effective in promoting uterine contractions and reducing the duration of labor, was found to be associated with an increased incidence of maternal fever during labor, in line with results reported by Ren et al. This effect may be attributable to the oxytocin-induced release of prostaglandin E₂ and inflammatory mediator F_{2α}. Prostaglandin E₂ acts directly on the hypothalamic thermoregulatory center, elevating the set point via cellular signaling pathways and consequently inducing fever.²⁶ Furthermore, uterine contractions stimulated by oxytocin may lead to transient uterine ischemia, promoting the release of inflammatory mediators that contribute to systemic inflammation and fever development. Accordingly, limiting non-essential use of oxytocin during labor may be warranted. Pethidine, commonly used to suppress nociceptive signaling during labor, may interfere with hypothalamic thermoregulation, thereby impairing the body's ability to regulate core temperature. Additionally, pethidine may induce respiratory center depression, which could reduce heat dissipation and further elevate the risk of intrapartum fever. Although the literature addressing the association between pethidine and epidural analgesia-related intrapartum fever remains limited, the mechanisms proposed in this study remain hypothetical and require further empirical validation.

Advantages and Clinical Utility of the CatBoost Algorithm

CatBoost is a gradient boosting algorithm designed to efficiently handle categorical features automatically, eliminating the need for manual encoding procedures.²⁷ In the context of intrapartum fever prediction, CatBoost can directly incorporate categorical variables such as mode of delivery (eg, vaginal delivery or cesarean section) and the presence of maternal comorbidities (eg, gestational diabetes, hypertension), thereby reducing the burden of data transformation and enhancing model development efficiency. When applied to large datasets, CatBoost demonstrates high training efficiency by capturing complex data relationships through categorical feature handling and optimized algorithms. In this study, the algorithm showed superior accuracy in predicting ERMF. In clinical settings, CatBoost models are capable of generating rapid risk predictions, enabling clinicians to initiate preventive measures. In addition, CatBoost demonstrates strong adaptability to practical data challenges, including noise and missing values. Despite these strengths, the interpretability of CatBoost remains limited relative to linear models, with some clinicians viewing the model outputs with skepticism due to their inherent complexity. Nonetheless, CatBoost can be integrated into hospital information systems for real-time interaction with parturient electronic medical records, allowing for timely clinical interventions that improve both the effectiveness and scientific rigor of obstetric decision-making.

Clinical Translation and Future Outlook

The ERMF prediction model developed in this study holds significant clinical translational value, with its core function being to enable dynamic risk stratification as a real-time decision support tool. The clinical implementation pathway comprises four key components:

- 1) System Integration: The model will be embedded within the hospital's electronic medical record system, automatically activating the predictive workflow at the initiation of epidural analgesia.
- 2) Real-time Prediction: The model automatically integrates static information and dynamic data, periodically calculating and updating the probability of fever risk.
- 3) Risk Intervention: Establish a three-tier management system: Low risk (<20%): Standard monitoring; Moderate Risk (20%-50%): Strengthen temperature monitoring; High risk (>50%): Initiate multidisciplinary intervention, including adjustment of analgesia protocols and neonatal team alert.
- 4) Application Value: Transitioning from reactive treatment to proactive prevention, optimizing healthcare resource allocation, and ultimately improving maternal and infant outcomes.

Finally, it is essential to reiterate the predictive nature of this model. While it provides a practical tool for clinical risk stratification, the identified predictors require further validation of their potential causal effects through the application of causal inference methods in specifically designed prospective studies or randomized trials.

Limitations

Despite the string predictive performance of the CatBoost algorithm, certain limitations must be acknowledged. First, the model's internal structure and decision-making process remain relatively complex and challenging to interpret. Second, the sample size used in this study is relatively limited, and it is a single-center study with all data sourced from the First People's Hospital of Zunyi City, Guizhou Province, China. Therefore, the predictive model we developed may be influenced by specific population characteristics, regional environmental factors, and the center's particular clinical and anesthetic protocols. Future efforts should focus on rigorously external validating this model in multicenter cohorts encompassing diverse geographic distributions and demographic characteristics to assess its generalizability and advance its widespread clinical translation. Third, although a web-based real-time application tool has been developed, it has only been validated within the research center. Future efforts should focus on rigorously external validating this model in a multicenter cohort encompassing diverse geographic distributions and demographic characteristics to assess its generalizability and advance its widespread clinical translation.

Conclusion

This study developed and validated a machine learning-based prediction model for epidural-related maternal fever using comprehensive clinical data. The CatBoost-based model demonstrated superior performance in identifying high-risk parturients (accuracy 0.91, AUC 0.94, precision 0.88, recall 0.81, F1 score 0.84), providing a practical tool for individualized intrapartum care. It is crucial to emphasize that the identified predictors represent statistical associations rather than causal relationships, and their clinical interpretation should consider potential confounding factors. While showing promising results, the model requires external validation across diverse populations and healthcare settings to verify its generalizability and scalability. The resulting model provides clinical reference value for the early identification of high-risk parturients and supports individualized management strategies in obstetric practice.

Funding

Zunyi Science and Technology Plan Project (Zunyi Science Cooperation HZ No. (2023-51)).

Disclosure

The authors have no conflict of interest.

References

1. Segal S. Labor epidural analgesia and maternal fever. *Anesth Analg.* 2010;111(6):1467–1475. doi:10.1213/ANE.0b013e3181f713d4
2. Towers CV, Yates A, Zite N, Smith C, Chernicky L, Howard B. Incidence of fever in labor and risk of neonatal sepsis. *Am J Obstet Gynecol.* 2017;216(6):596e591–596e595. doi:10.1016/j.ajog.2017.02.022
3. Curtin WM, Katzman PJ, Florescue H, Metlay LA, Ural SH. Intrapartum fever, epidural analgesia and histologic chorioamnionitis. *J Perinatol.* 2015;35(6):396–400. doi:10.1038/jp.2014.235
4. Attard Cortis P, Bleaser T, Devroe S, Lucas N. Neuraxial labour analgesia and maternal fever: a narrative review. *Eur J Anaesthesiol.* 2025;43(1):45–52. doi:10.1097/EJA.0000000000002280
5. Wang H, Yang Z, Wei S, et al. Perinatal outcomes and risk factors for epidural analgesia-associated intrapartum maternal fever: a retrospective study. *J Matern Fetal Neonatal Med.* 2023;36(1):2179383. doi:10.1080/14767058.2023.2179383
6. Jia L, Cao H, Guo Y, et al. Evaluation of epidural analgesia use during labor and infection in full-term neonates delivered vaginally. *JAMA Network Open.* 2021;4(9):e2123757–e2123757. doi:10.1001/jamanetworkopen.2021.23757
7. Hafner C, Windpassinger M, Tretter EV, et al. Role of mitochondrial DNA level in epidural-related maternal fever: a single-centre, observational, pilot study. *BMC Pregnancy Childbirth.* 2024;24(1):341. doi:10.1186/s12884-024-06551-7
8. Seiler FA, Scavone BM, Shahul S, Arnolds DE. Maternal fever associated with continuous spinal versus epidural labor analgesia: a single-center retrospective study. *Anesth Analg.* 2022;135(6):1153–1158. doi:10.1213/ANE.0000000000005905
9. Jiang Z, Hu X, Zeng H, et al. Nomogram for perinatal prediction of intrapartum fever: a retrospective case-control study. *BMC Pregnancy Childbirth.* 2021;21(1):445. doi:10.1186/s12884-021-03891-6
10. Zhao B, Li B, Wang Q, Song X. The relationship between epidural analgesia and intrapartum maternal fever and the consequences for maternal and neonatal outcomes: a prospective observational study. *J Matern Fetal Neonatal Med.* 2022;35(25):5354–5362. doi:10.1080/14767058.2021.1879042
11. Howle R, Ragbourne S, Zolger D, Owolabi A, Onwochei D, Desai N. Influence of different volumes and frequency of programmed intermittent epidural bolus in labor on maternal and neonatal outcomes: a systematic review and network meta-analysis. *J Clin Anesth.* 2024;93:111364. doi:10.1016/j.jclinane.2023.111364

12. Sun C, Ren S, Chen C, et al. Pulse perfusion index for predicting intrapartum fever during epidural analgesia. *J Clin Anesth.* 2022;80:110852. doi:10.1016/j.jclinane.2022.110852
13. Li L, Yang XT, Zou J, et al. Predictive value of the neutrophil-to-lymphocyte ratio for epidural labor analgesia-associated intrapartum fever: a retrospective single-center study. *BMC Anesthesiol.* 2025;25(1):96. doi:10.1186/s12871-025-02972-9
14. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* 2025;79:103470. doi:10.1016/j.redox.2024.103470
15. Fan Y, Fan C, Mao P, et al. Study on perinatal-related factors of maternity and newborn in parturients with intrapartum fever in part of Eastern China: a cross-sectional study. *J Clin Lab Anal.* 2022;36(1):e24050. doi:10.1002/jcla.24050
16. Zhao C, Wu D, Huang J, et al. BoostTree and BoostForest for ensemble learning. *IEEE Trans Pattern Anal Mach Intell.* 2023;45(7):8110–8126. doi:10.1109/TPAMI.2022.3227370
17. Ghazwani M, Begum MY. Computational intelligence modeling of hyoscine drug solubility and solvent density in supercritical processing: gradient boosting, extra trees, and random forest models. *Sci Rep.* 2023;13(1):10046. doi:10.1038/s41598-023-37232-8
18. Ganaie MA, Tanveer M, Suganthan PN, Snaes V. Oblique and rotation double random forest. *Neural Netw.* 2022;153:496–517. doi:10.1016/j.neunet.2022.06.012
19. Wan S, Zhan Y, Chen S, et al. Boosting graph contrastive learning via adaptive sampling. *IEEE Trans Neural Netw Learn Syst.* 2024;35(11):15971–15983. doi:10.1109/TNNLS.2023.3291358
20. Li K, Zhou G, Zhai J, Li F, Shao M. Improved PSO_AdaBoost ensemble algorithm for imbalanced data. *Sensors.* 2019;19(6):1476.
21. Yuan X, Qiu L, Huang Y, et al. Influencing factors of intrapartum fever after epidural labor analgesia. *Rev Assoc Med Bras.* 2024;70(11):e20240565. doi:10.1590/1806-9282.20240565
22. Douma MR, Stienstra R, Middeldorp JM, Arbous MS, Dahan A. Differences in maternal temperature during labour with remifentanyl patient-controlled analgesia or epidural analgesia: a randomised controlled trial. *Int J Obstet Anesth.* 2015;24(4):313–322. doi:10.1016/j.ijoa.2015.06.003
23. Kim SY, Hong SY, Kwon DY, et al. Degree of intrapartum fever and associated factors: three group analysis of no fever, borderline and overt fever. *J Obstet Gynaecol Res.* 2021;47(3):1153–1163. doi:10.1111/jog.14651
24. Ren J, Wang T, Yang B, et al. Risk factors and safety analyses for intrapartum fever in pregnant women receiving epidural analgesia during labor. *Med Sci Monit.* 2021;27:e929283. doi:10.12659/MSM.929283
25. Burgess APH, Katz JE, Moretti M, Lakhi N. Risk factors for intrapartum fever in term gestations and associated maternal and neonatal sequelae. *Gynecol Obstet Invest.* 2017;82(5):508–516. doi:10.1159/000453611
26. Jindal S, Steer PJ, Savvidou M, et al. Risk factors for a serious adverse outcome in neonates: a retrospective cohort study of vaginal births. *BJOG.* 2023;130(12):1521–1530. doi:10.1111/1471-0528.17531
27. Hanani AA, Donmez TB, Kutlu M, Mansour M. Predicting thyroid cancer recurrence using supervised CatBoost: a SHAP-based explainable AI approach. *Medicine.* 2025;104(22):e42667. doi:10.1097/MD.0000000000042667

International Journal of Women's Health

Publish your work in this journal

The International Journal of Women's Health is an international, peer-reviewed open-access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of women's healthcare including gynecology, obstetrics, and breast cancer. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-womens-health-journal>

Dovepress
Taylor & Francis Group