

Machine learning-based prediction of out-of-hospital births in prehospital emergency care

Joanna Wach^{1,A–F}, Łukasz Lewandowski^{2,A,C–F}, Jakub Staniczek^{3,D–F}, Raúl Juárez-Vela^{4,D–F}, Michał Czapla^{4,5,A,C–F}

¹ Division of Specialist Care in Midwifery and Gynecology, Department of Midwifery, Faculty of Nursing and Midwifery, Wrocław Medical University, Poland

² Department of Medical Biochemistry, Wrocław Medical University, Poland

³ Department of Obstetrics and Gynecological Oncology, Medical University of Silesia, Katowice, Poland

⁴ Group of Research in Care (GRUPAC), Faculty of Health Sciences, University of La Rioja, Logroño, Spain

⁵ Department of Emergency Medical Service, Wrocław Medical University, Poland

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;

D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online)

Adv Clin Exp Med. 2026

Address for correspondence

Joanna Wach

E-mail: joanna.wach@umw.edu.pl

Funding sources

This research was funded by the Polish Ministry of Science and Higher Education under the statutory grant of Wrocław Medical University (SUBZ.L030.25.052).

Conflict of interest

None declared

Received on January 18, 2026

Reviewed on February 15, 2026

Accepted on March 6, 2026

Published online on March 25, 2026

Cite as

Wach J, Lewandowski Ł, Staniczek J, Juárez-Vela R, Czapla M. Machine learning-based prediction of out-of-hospital births in prehospital emergency care [published online as ahead of print on March 25, 2026]. *Adv Clin Exp Med*. 2026. doi:10.17219/acem/218972

DOI

10.17219/acem/218972

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported (CC BY 3.0) (<https://creativecommons.org/licenses/by/3.0/>)

Abstract

Background. Unplanned out-of-hospital births constitute rare but high-risk obstetric emergencies managed by emergency medical services (EMS). Rapid assessment of labor progression in prehospital settings is challenging due to limited diagnostic resources and time pressure, increasing the risk of adverse maternal and neonatal outcomes. Machine learning (ML) may support early risk stratification using routinely collected prehospital data.

Objectives. To develop and validate supervised ML models for predicting prehospital birth and to evaluate whether these models reflect clinically intuitive obstetric reasoning.

Materials and methods. This retrospective observational study analyzed 3,002 EMS-attended labor cases in Poland (August 2021–January 2022). The outcome was birth occurring before hospital arrival. Candidate predictors included maternal characteristics, obstetric history, stage of labor, vital signs, and intrapartum findings. Penalized logistic regression (elastic net), random forest (RF), support vector classifier with radial basis function kernel (SVC-RBF), Gaussian naïve Bayes (GNB), and k-nearest neighbors (kNN) models were trained using stratified fivefold cross-validation. Model performance was evaluated using discrimination metrics (area under the receiver operating characteristic curve (ROC-AUC) and precision-recall AUC (PR-AUC)) and calibration metrics (Brier score and logarithmic loss (log loss)). Nested cross-validation was applied to reduce overfitting. Model interpretability was assessed using standardized coefficients, permutation importance, and Shapley Additive Explanations (SHAP) values.

Results. Penalized logistic regression demonstrated robust performance (ROC-AUC: 0.97 ± 0.01 ; PR-AUC: 0.81 ± 0.04 ; Brier score: 0.036 ± 0.015). Random forest and SVC-RBF models achieved comparable discrimination (ROC-AUC up to 0.97), whereas kNN performed less well (ROC-AUC = 0.84). The 2nd stage of labor was the dominant predictor ($\beta = 1.39$), followed by amniotic fluid status ($\beta = -0.44$). Sensitivity analysis excluding the stage of labor reduced model performance but retained moderate discrimination (ROC-AUC ≈ 0.76), indicating that additional clinical variables contributed to prediction.

Conclusions. Machine learning models demonstrated high internal predictive performance for prehospital birth using routinely available EMS data and reproduced clinically intuitive decision patterns. Such tools may support, but not replace, prehospital obstetric decision-making.

Key words: machine learning, emergency medical services, obstetrics, paramedics, out-of-hospital birth

Highlights

- Machine learning (ML) models accurately predict prehospital birth in EMS settings, achieving high performance (ROC-AUC up to 0.97).
- Penalized logistic regression demonstrated robust discrimination and calibration using routinely collected pre-hospital obstetric data.
- Key predictors include stage of labor and amniotic fluid status, reflecting clinically intuitive decision-making patterns.
- ML-based tools may support early risk stratification in out-of-hospital births, improving maternal and neonatal outcomes in emergency care.

Background

Unplanned out-of-hospital births represent rare but clinically demanding obstetric emergencies encountered by emergency medical services (EMS). Although they account for a small proportion of prehospital callouts, these events are characterized by time pressure, limited diagnostic resources, and an increased risk of adverse maternal and neonatal outcomes, including postpartum hemorrhage, neonatal hypothermia, and the need for immediate resuscitative interventions.^{1–4} In 2023, over 272,000 live births were recorded in Poland, representing a decrease of nearly 33,000 compared with the previous year. This declining trend in birth numbers has significant implications for the functioning of the maternity care system. In some regions, it has led to the closure of obstetric wards, which may in turn contribute to an increase in out-of-hospital births.⁵ Prehospital clinicians are therefore required to rapidly assess labor progression and determine whether safe transport to hospital is feasible or whether delivery is likely to occur before arrival, often under conditions of substantial uncertainty and variable clinical experience.^{2,3} In recent years, there has been a substantial increase in the number of scientific publications addressing the application of artificial intelligence (AI) in medicine. Artificial intelligence-based technologies demonstrate considerable potential to transform and optimize diagnostic, prognostic, and decision-making processes across multiple areas of healthcare.^{6,7}

From a clinical perspective, obstetrics and prehospital emergency care represent settings in which decision-making is often time-critical and must be performed with limited diagnostic resources. Unplanned out-of-hospital births, although relatively rare, constitute high-risk events that require rapid assessment of labor progression and immediate evaluation of maternal and neonatal safety, often by clinicians without direct access to specialist obstetric support.^{3,8,9}

The scientific literature increasingly reports the use of AI to support clinical decision-making during pregnancy. These applications include, i.a., the analysis of fetal images obtained using magnetic resonance imaging

(MRI) with AI-based algorithms, prediction of preterm birth based on electrohysterographic (EHG) signals, and assessment of the risk of fetal compromise during labor.^{10–12} Beyond specific clinical applications, increasing attention has been directed toward the processes by which such AI-based tools are developed. In particular, the importance of interdisciplinary collaboration between clinicians – including obstetricians and midwives – and data science specialists has been emphasized to ensure the clinical relevance and interpretability of predictive models.^{9,13,14}

Artificial intelligence represents a promising tool for addressing complex problems related to risk prediction and clinical assessment. By integrating multiple clinical and demographic variables, AI-based models enable the identification of risk factors associated with out-of-hospital childbirth and may therefore improve predictions of delivery occurring in prehospital settings.¹⁵ In clinical obstetrics and emergency medicine, supervised machine learning (ML) approaches are particularly relevant because they allow the prediction of predefined, clinically meaningful outcomes based on routinely collected patient data, thereby supporting – but not replacing – clinical judgement.^{7,16}

Out-of-hospital births attended by EMS are rare but clinically demanding events.⁴ Their sudden onset, limited availability of resources, and the need for rapid clinical decision-making make them a significant challenge in emergency medicine. Therefore, identifying and analyzing predictive factors associated with unplanned out-of-hospital births may play an important role in improving the quality of care provided to both the mother and the newborn in prehospital settings.^{8,17,18}

Objectives

The objective of this study was to develop ML models to predict out-of-hospital births and to evaluate whether these models capture clinically meaningful patterns based on variables routinely assessed during prehospital obstetric care.

Material and methods

Study design and setting

This study employed a retrospective observational design based on routinely collected prehospital EMS data. The analysis included all EMS-attended childbirth events occurring outside hospital settings in Poland between August 2021 and January 2022. Clinical and operational information was extracted from standardized medical rescue procedure records completed by EMS personnel as part of routine care documentation. Cases were identified using International Classification of Diseases, Tenth Revision (ICD-10) diagnostic codes corresponding to childbirth and labor, including preterm labor and delivery (O60), precipitate labor (O62.3), and full-term uncomplicated delivery (O80). The study population consisted of EMS-attended obstetric or labor-related callouts. The primary outcome was whether delivery occurred before hospital arrival.

This study was conducted in accordance with the Declaration of Helsinki. The research protocol was approved by the Independent Bioethics Committee of Wrocław Medical University (Poland; decision No. KB-206/2023N). The requirement for informed consent was waived by the Committee due to the retrospective nature of the study and the use of fully anonymized data, in accordance with applicable national regulations.

Study population and data

A total of 5,097 EMS records were reviewed. Of these, 2,095 cases (41%) were excluded because patients were in the 3rd or 4th stage of labor upon EMS arrival. The remaining 3,002 cases (59%) involved women in the 1st or 2nd stage of labor and were included in the final analysis. Women aged 16 years or older who contacted EMS due to the onset of labor and were in the 1st or 2nd stage of labor at the time of EMS intervention were included. Exclusion criteria included age under 16 years, being in the 3rd or 4th stage of labor upon EMS arrival, and incomplete or missing EMS documentation.

Factors associated with prehospital deliveries attended by EMS teams were examined. Collected data included the location and reason for EMS activation, vital signs (heart rate, blood pressure, oxygen saturation, and blood glucose level), and maternal health conditions such as gestational diabetes, gestational hypertension, COVID-19 infection, and other comorbidities, including thrombosis, thyroid disorders, depression, and epilepsy. Obstetric history was also considered, including the number of pregnancies and deliveries, gestational age, the course of pregnancy, and access to prenatal care. Labor-related factors were assessed, including uterine contractions, rupture of membranes, stage of labor, and pregnancy complications such as the risk of preterm delivery, cervical insufficiency, fetal growth restriction, and oligohydramnios. Intrapartum complications

– including hemorrhage, umbilical cord prolapse, retained placenta, and eclampsia – were also analyzed.

The study was conducted in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.¹⁹

Statistical analyses

The aim of the analysis was to evaluate whether ML models trained on routinely collected prehospital obstetric data can predict the occurrence of a prehospital birth and reflect clinically intuitive assessment of labor progression. The binary outcome was the presence or absence of a prehospital birth, as defined in the source registry. Candidate predictors were specified a priori and included maternal characteristics, obstetric history, prehospital vital signs and point-of-care measurements, clinical presentation, and labor-related variables routinely assessed by EMS. Detailed definitions of all predictors are provided in the Supplementary data.

The dataset was split into training (approx. 80%) and test (approx. 20%) sets using stratified sampling by outcome. Preprocessing steps – including imputation of continuous variables, encoding of categorical variables, and scaling – were fitted exclusively on the training data and subsequently applied to the test set to prevent information leakage. Records with missing categorical predictors were excluded after the split to ensure unambiguous encoding. Several supervised learning algorithms were evaluated, including penalized logistic regression, random forest (RF) classifier, support vector classifier with radial basis function kernel (SVC-RBF), Gaussian naïve Bayes (GNB), and k-nearest neighbors (kNN). Hyperparameter tuning was performed within the training data using cross-validation. Model performance was assessed using the area under the receiver operating characteristic curve (ROC-AUC), precision–recall AUC (PR-AUC), Brier score, logarithmic loss (log loss), and calibration measures. Nested cross-validation was additionally performed to evaluate potential optimism related to model selection. Model interpretability was assessed using permutation-based feature importance and Shapley Additive Explanations (SHAP) analysis. Sensitivity analyses excluding stage of labor were conducted to evaluate model dependence on temporally proximal predictors. A detailed description of the modeling pipeline, preprocessing steps, and interpretability analyses is provided in the Supplementary data.

Results

Study population characteristic

Table 1 summarizes the baseline (pre-imputation) characteristics of the study population stratified by birth outcome. The analysis included 3,002 observations, of which 2,735 were classified as no prehospital birth and 267 as prehospital birth.

Table 1. Characteristics of the study population before imputation

Variable	Level	Overall	No prehospital birth	Prehospital birth
n	–	3,002	2,735	267
Age [years]	–	29.00 [23.00, 34.00]	28.00 [23.00, 34.00]	30.00 [25.00, 34.00]
	missing, n (%)	53 (1.8)	41 (1.5)	12 (4.5)
Respiratory rate	–	16.00 [14.00, 18.00]	16.00 [14.00, 18.00]	16.00 [14.00, 18.00]
	missing, n (%)	400 (13.3)	345 (12.6)	55 (20.6)
Saturation [%]	–	98.00 [98.00, 99.00]	98.00 [98.00, 99.00]	98.00 [97.00, 99.00]
	missing, n (%)	331 (11.0)	290 (10.6)	41 (15.4)
Systolic blood pressure [mm Hg]	–	130.00 [120.00, 140.00]	130.00 [120.00, 140.00]	130.00 [120.00, 140.00]
	missing, n (%)	670 (22.3)	597 (21.8)	73 (27.3)
Diastolic blood pressure [mm Hg]	–	80.00 [70.00, 85.00]	80.00 [70.00, 85.00]	80.00 [75.75, 84.00]
	missing, n (%)	679 (22.6)	604 (22.1)	75 (28.1)
Heart rate [bpm]	–	90.00 [82.00, 100.00]	90.00 [81.00, 100.00]	92.00 [85.00, 103.50]
	missing, n (%)	365 (12.2)	324 (11.8)	41 (15.4)
Glucose [mg/dL]	–	102.00 [90.00, 117.00]	101.00 [90.00, 116.00]	104.00 [87.75, 126.00]
	missing, n (%)	2,415 (80.4)	2,202 (80.5)	213 (79.8)
Gestational week	–	39.00 [38.00, 40.00]	39.00 [37.00, 40.00]	39.00 [38.00, 40.00]
	missing, n (%)	50 (1.7)	43 (1.6)	7 (2.6)
Number of pregnancies	–	2.00 [1.00, 4.00]	2.00 [1.00, 3.00]	3.00 [2.00, 4.00]
	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
Number of labors	–	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	2.00 [2.00, 3.00]
	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
Public location	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,809 (93.6)	2,564 (93.7)	245 (91.8)
	yes	193 (6.4)	171 (6.3)	22 (8.2)
Multipara	missing, n (%)	30 (1.0)	27 (1.0)	3 (1.1)
	no	873 (29.4)	820 (30.3)	53 (20.1)
	yes	2,099 (70.6)	1,888 (69.7)	211 (79.9)
Complications during pregnancy	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,468 (82.2)	2,246 (82.1)	222 (83.1)
	yes	534 (17.8)	489 (17.9)	45 (16.9)
Medical care during pregnancy	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	197 (6.6)	172 (6.3)	25 (9.4)
	yes	2,805 (93.4)	2,563 (93.7)	242 (90.6)
Stage of labor	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	I	2,655 (88.4)	2,643 (96.6)	12 (4.5)
	II	347 (11.6)	92 (3.4)	255 (95.5)
Bleeding	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,741 (91.3)	2,491 (91.1)	250 (93.6)
	yes	261 (8.7)	244 (8.9)	17 (6.4)
Any fetal movement problem	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,941 (98.0)	2,677 (97.9)	264 (98.9)
	yes	61 (2.0)	58 (2.1)	3 (1.1)
Amniotic fluid status	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	rupture of membranes	1,525 (50.8)	1,271 (46.5)	254 (95.1)
	preserved	1,477 (49.2)	1,464 (53.5)	13 (4.9)
Comorbidities present	missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,655 (88.4)	2,416 (88.3)	239 (89.5)
	yes	347 (11.6)	319 (11.7)	28 (10.5)

Table 1. Characteristics of the study population before imputation – cont.

Variable	Level	Overall	No prehospital birth	Prehospital birth
Gestational diabetes, n (%)	missing	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,849 (94.9)	2,593 (94.8)	256 (95.9)
	yes	153 (5.1)	142 (5.2)	11 (4.1)
Gestational hypertension, n (%)	missing	0 (0.0)	0 (0.0)	0 (0.0)
	no	2,942 (98.0)	2,683 (98.1)	259 (97.0)
	yes	60 (2.0)	52 (1.9)	8 (3.0)

Quantitative variables are presented as median [Q1–Q3] and categorical variables as n (%).

Maternal age in the overall cohort had a median of 29 years (interquartile range (IQR): 23–34), with slightly higher median values observed in the prehospital birth group. Gestational age at delivery was similar across outcome strata, with a median of 39 weeks in all groups.

Vital signs – including respiratory rate, oxygen saturation, systolic and diastolic blood pressure, and heart rate – showed comparable central tendencies across outcome categories. Missing data were present for several physiological measurements, most notably blood pressure variables and heart rate, while blood glucose concentration exhibited a high proportion of missing values (approx. 80%), reflecting the selective availability of this measurement in the source data. The extent of missingness for each variable is explicitly reported overall and by outcome group.

Obstetric history variables indicated a median of 2 pregnancies and 2 deliveries in the overall cohort, with higher parity more frequently observed among cases with prehospital birth. The majority of births occurred in non-public locations, and most women were classified as multiparous. Complications during pregnancy and comorbid conditions were reported in a minority of cases, with similar distributions across outcome strata.

Regarding intrapartum characteristics, the 1st stage of labor predominated in the group without prehospital birth, whereas the 2nd stage of labor was more common among cases with prehospital birth. Bleeding events and fetal movement problems were relatively infrequent in the overall population.

Amniotic fluid status differed markedly between outcome groups, with rupture of membranes more frequently recorded among prehospital births. Gestational diabetes and gestational hypertension were uncommon in the cohort and showed comparable frequencies across outcome categories.

All variables presented in Table 1 are based on observed, non-imputed data. Missing values are reported explicitly to reflect the structure and completeness of the underlying registry. No statistical hypothesis testing was performed for between-group comparisons, as the purpose of this table is to provide a descriptive overview of the study population rather than to infer associations.

Model performance

Across fivefold cross-validation, substantial discriminatory performance was observed for several evaluated models (Table 2). Penalized logistic regression with elastic net regularization (logreg_en) demonstrated consistently high discrimination, with fold-specific ROC-AUC values ranging from 0.95 to 0.99. Measures of overall probabilistic accuracy were favorable, as reflected by low log loss and Brier scores. The number of features retained in the final model was stable across folds (n = 21), suggesting robustness of the selected predictor set. The SVC-RBF achieved similarly high ROC-AUC values (approx. 0.96–0.98) and competitive PR-AUC values. However, compared with logistic regression, the SVC-RBF model exhibited less favorable

Table 2. Performance metrics of the evaluated machine learning (ML) models

Model	AUC		PR-AUC		Brier		log loss		ACC	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
logreg_en	0.970	0.015	0.805	0.041	0.036	0.015	0.131	0.047	0.971	0.007
rf	0.972	0.010	0.826	0.035	0.026	0.003	0.113	0.020	0.971	0.006
svc_rbf	0.973	0.009	0.803	0.036	0.029	0.005	0.107	0.018	0.968	0.006
knn	0.842	0.061	0.559	0.109	0.059	0.006	0.661	0.225	0.807	0.087
gnb	0.971	0.009	0.792	0.055	0.034	0.007	0.415	0.084	0.964	0.009

Values are reported as mean ± standard deviation (SD) across five-fold cross-validation. Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC) and the area under the precision–recall curve (PR-AUC). Probabilistic calibration was evaluated using the Brier score and logarithmic loss (log loss). Classification accuracy (ACC) was computed using fold-specific decision thresholds optimized with the Youden index. Detailed fold-level performance metrics and hyperparameter configurations are provided in Supplementary Table 2.

logreg_en – penalized logistic regression with elastic net regularization; rf – random forest; svc_rbf – support vector classifier with radial basis function kernel; knn – k-nearest neighbors; gnb – Gaussian naïve Bayes.

calibration metrics (higher log loss and Brier score) and greater variability in optimal decision thresholds across folds, indicating reduced stability of probability estimates.

Gaussian naïve Bayes also showed high discrimination; however, calibration-related metrics were inferior, with wide variation in fold-specific decision thresholds, suggesting limited reliability of predicted probabilities despite good ranking performance. The kNN model demonstrated clearly inferior and unstable performance across all evaluated metrics and was therefore considered unsuitable for further interpretation.

Overall, penalized logistic regression provided the most favorable balance between discrimination, calibration, and stability and was selected as the primary model for subsequent interpretability analyses. Full fold-level results for all evaluated models and hyperparameters used in tuning are provided in Supplementary Table 1.

Feature selection was embedded within the modeling framework via elastic net regularization applied to the final design matrix (including one-hot encoded categorical variables). The set of candidate variables was specified a priori; therefore, predictors were not removed from the modeling pipeline across folds. Instead, the elastic net penalty performed coefficient shrinkage and, in the final refitted model, set several coefficients exactly to 0 ($\beta = 0$), effectively removing their contribution to predictions while retaining them in the model matrix. This indicates that predictive performance was not driven by a single narrow subset of predictors but rather by a distributed set of clinically plausible cues, with some variables contributing negligibly after regularization. The full design matrix specification is provided in Supplementary Table 2.

Nested cross-validation

The candidate predictor specification was fixed a priori and remained unchanged across folds; therefore, feature stability refers to a stable design matrix definition rather than fold-specific feature inclusion or exclusion. To assess potential optimism bias related to hyperparameter tuning, nested cross-validation was performed, with an outer fivefold loop for performance estimation and an inner loop for model optimization.

Performance estimates obtained from the outer folds were highly consistent with the results of the primary cross-validation analysis. To assess potential optimism introduced by fitting preprocessing once prior to cross-validation, nested cross-validation was additionally performed; the results remained highly consistent, suggesting minimal practical impact of this simplification. In particular, penalized logistic regression maintained high discriminatory performance, with outer-fold ROC-AUC values ranging from approx. 0.95 to 0.99, and stable calibration metrics across folds. The relative ranking of the evaluated models remained unchanged, and no material inflation of performance estimates was observed.

Comparative discrimination performance across the evaluated models is illustrated by receiver operating characteristic (ROC) curves based on out-of-fold predictions (Fig. 1). Detailed fold-level nested cross-validation results and the corresponding hyperparameters are provided in Supplementary Tables 3 and 4.

Feature importance

Given the primary aim of evaluating whether machine learning models can reflect clinically intuitive reasoning in prehospital obstetric assessment, penalized logistic regression with elastic net regularization was selected as the primary interpretative model (Fig. 2). This approach provides direct interpretability in terms of effect direction and relative magnitude on the log-odds scale, while maintaining strong predictive performance comparable to that of more flexible algorithms. The displayed coefficients represent standardized regression coefficients (β), allowing comparison of the relative contribution of predictors within a multivariable penalized framework.

To complement this clinically intuitive and transparent model with a nonlinear perspective, a RF classifier was additionally examined (Fig. 3). Random forest achieved similar discriminatory performance and was used as a nonlinear robustness check to assess whether a more flexible model yields a comparable hierarchy of clinically relevant predictors. Feature importance derived from the RF model was used to evaluate whether a more flexible ensemble method identifies a similar hierarchy of clinically relevant predictors, despite allowing for nonlinear effects and interactions.

Across both modeling approaches, a highly consistent pattern of predictor relevance was observed. In both penalized logistic regression (Fig. 2) and RF feature importance analysis (Fig. 3), the 2nd stage of labor emerged as the dominant predictor. This dominance is expected because stage of labor captures temporal proximity to delivery, which is intrinsically linked to whether birth occurs before hospital arrival. This finding reflects the clinically intuitive notion that advanced labor progression is the primary determinant of whether delivery occurs prior to hospital transport. Amniotic fluid status was consistently identified as the 2nd most influential predictor across models, further supporting its established role in obstetric assessment of labor dynamics. Additional variables – including maternal vital signs, gestational age, blood glucose level, and obstetric history – contributed smaller incremental information. This pattern indicates that model performance was not driven by a single variable alone but rather by the combined influence of multiple clinically meaningful cues, each contributing modestly to risk stratification beyond stage of labor.

Candidate predictors were consistently included across cross-validation folds; however, elastic-net shrinkage in the refitted logistic regression model set multiple coefficients to exactly zero, resulting in sparse SHAP attributions consistent with the final model structure

(as discussed later in the text). Similarly, RF feature importance showed a gradual decline beyond the most influential predictors, without abrupt cut-offs, further supporting the robustness of the identified feature hierarchy. Permutation-based feature importance was computed for

all evaluated models and summarized as mean \pm standard deviation (SD) across cross-validation folds. Results consistently identified the 2nd stage of labor and rupture of membranes as the most influential predictors across modeling approaches (Supplementary Table 5). Importantly,

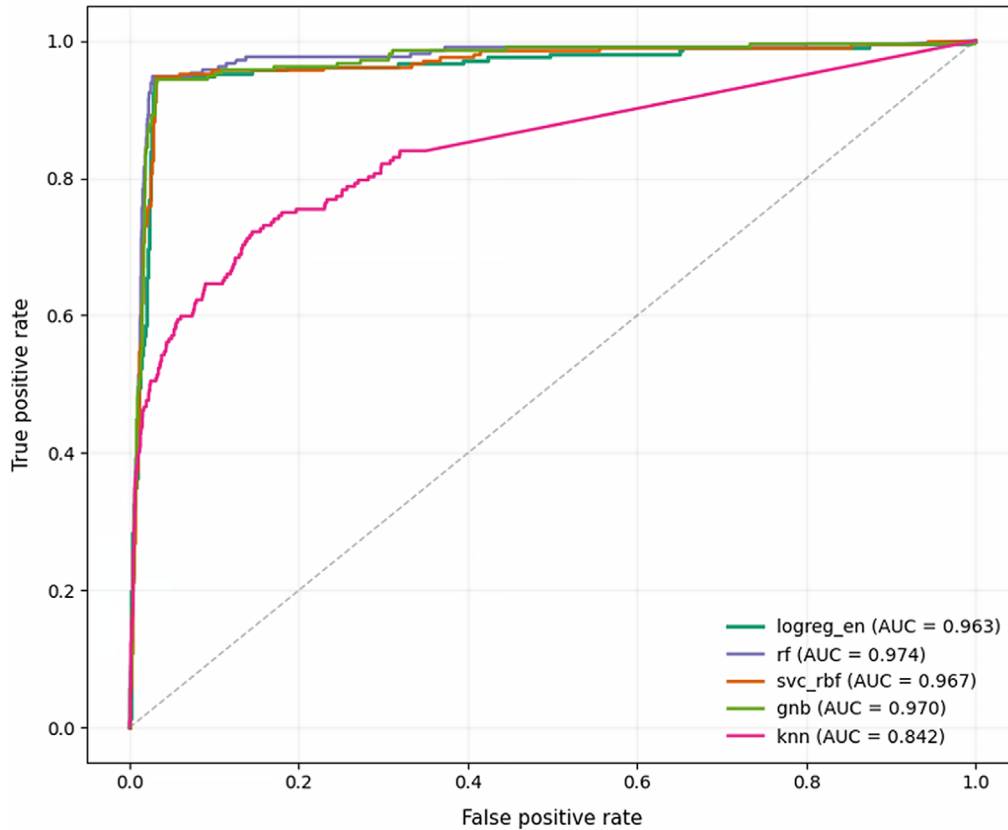


Fig. 1. Receiver operating characteristic (ROC) curves for evaluated machine learning (ML) models (out-of-fold predictions). The ROC curves illustrate the relationship between the true positive rate (sensitivity) and the false positive rate (1 – specificity) for each model based on out-of-fold predictions from cross-validation. The area under the ROC curve (AUC), shown in the legend, quantifies discriminative performance; values closer to 1.0 indicate superior discrimination. The diagonal dashed line represents a classifier with no discriminative ability (AUC = 0.5)

logreg_en – logistic regression with elastic net; rf – random forest; svc_rbf – support vector classification with radial basis function kernel; gnb – Gaussian naïve Bayes; knn – k-nearest neighbors.

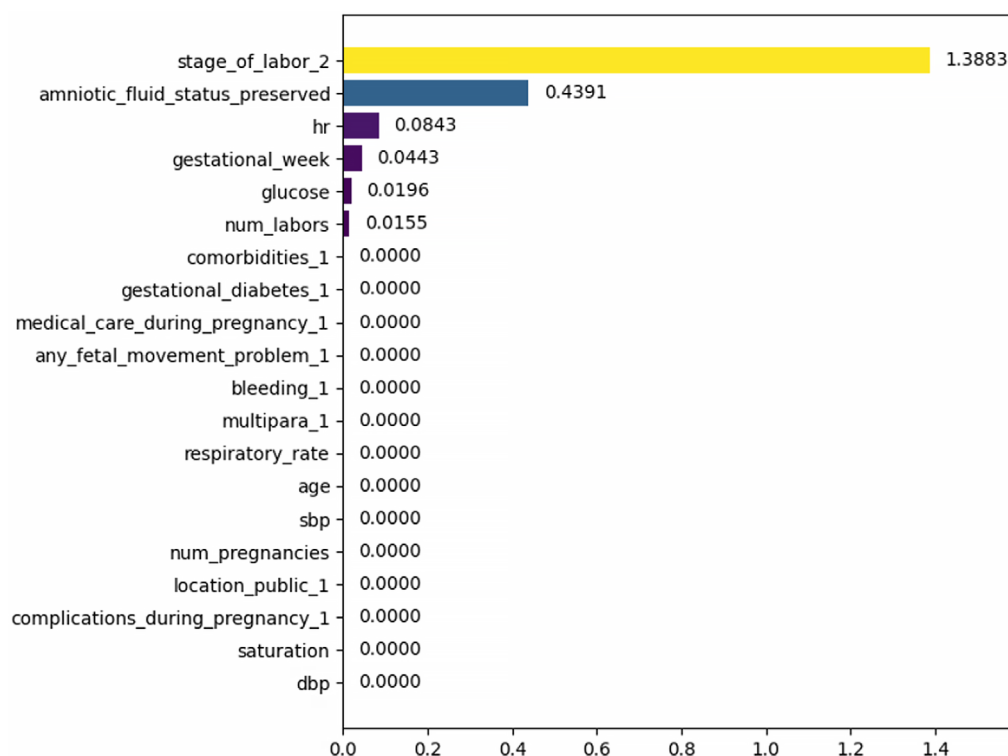


Fig. 2. Standardized regression coefficients (β) from the elastic net penalized logistic regression model. Coefficients indicate the direction and relative strength of association between predictors and the probability of prehospital birth. Positive values indicate increased likelihood, whereas negative values indicate decreased likelihood

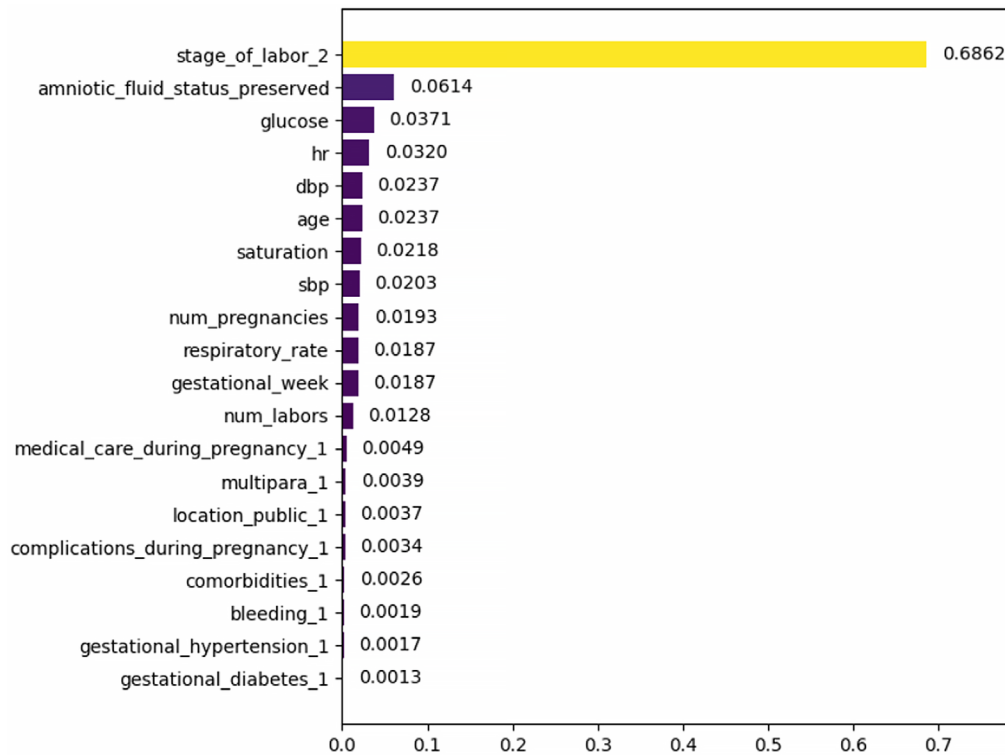


Fig. 3. Feature importance scores from the random forest classifier (RF), indicating the relative contribution of predictors to the classification of prehospital birth

the reported coefficients and feature importance measures should be interpreted comparatively rather than causally, as they reflect associations within multivariable models trained for prediction rather than for causal inference.

Sensitivity analysis excluding stage of labor

To evaluate the extent to which model performance was driven by advanced labor status, a sensitivity analysis was conducted in which the variable stage of labor was excluded from model training and evaluation (i.e., all one-hot encoded indicators derived from this variable were removed). As expected, overall discriminatory performance decreased across all evaluated algorithms. Penalized logistic regression, RF, and GNB nevertheless retained moderate discriminatory ability (ROC-AUC approx. 0.75–0.78), indicating that predictive performance was not solely dependent on this single dominant predictor but was instead distributed across multiple physiologically and clinically coherent features.

Comparison of model performance between the full and restricted specifications demonstrated consistent absolute reductions in ROC-AUC following exclusion of stage of labor, with decreases of approx. 0.19–0.21 for penalized logistic regression, RF, and GNB (Supplementary Table 6). In contrast, the SVC-RBF exhibited a substantially larger performance decline ($\Delta\text{AUC} \approx -0.35$), suggesting a stronger dependence on advanced labor status. Despite the observed reduction in discrimination, the relative ranking of models remained broadly consistent with the primary

analysis, and no evidence of model collapse or marked deterioration in calibration was observed. Together, these findings support the robustness of the modeling framework and indicate that clinically relevant information beyond stage of labor contributes meaningfully to prediction.

Permutation-based importance shift analyses were performed for the 2 primary interpretative models: penalized logistic regression and RF (Supplementary Table 7). In both models, removal of the stage of labor variable resulted in a marked reallocation of importance toward multiple clinically related predictors rather than a collapse of the model structure. This consistent shift across linear and nonlinear modeling frameworks indicates that the predictive signal associated with advanced labor status is not unique but instead reflects an aggregation of physiologically and clinically coherent cues.

SHAP-based explanation of model predictions

Shapley Additive Explanations analyses were performed for the 2 final interpretative models (elastic-net penalized logistic regression and RF) refitted on the full training set, using a fixed subsample of 200 training observations and an independent background sample of 200 observations.

In the penalized logistic regression model, SHAP attributions were sparse and aligned with elastic-net regularization: multiple predictors exhibited exactly 0 coefficients ($\beta = 0$) and consequently had SHAP contributions equal to 0 within numerical tolerance under the applied linear SHAP formulation. The dominant

predictor was the 2nd stage of labor (stage_of_labor_2), showing the largest absolute SHAP contribution (mean |SHAP| ≈ 0.266) and the largest standardized coefficient ($\beta \approx 1.39$). Amniotic fluid status was consistently the 2nd most influential factor (mean |SHAP| ≈ 0.220; $\beta \approx -0.440$), followed by smaller contributions from maternal heart rate (mean |SHAP| ≈ 0.062; $\beta \approx 0.084$), gestational age (mean |SHAP| ≈ 0.039; $\beta \approx 0.044$), blood glucose (mean |SHAP| ≈ 0.026; $\beta \approx 0.020$), and parity (number of prior labors) (mean |SHAP| ≈ 0.010; $\beta \approx 0.015$) (Table 3).

In the RF model, SHAP contributions showed a similar hierarchy, with the largest absolute attribution again observed for stage_of_labor_2 (mean |SHAP| ≈ 0.098) and amniotic_fluid_status_1 (mean |SHAP| ≈ 0.027), while the remaining predictors contributed smaller incremental information. Overall, SHAP results provided a decomposition of the fitted model output that was fully consistent with the imposed regularization structure and the primary feature-importance findings, and were interpreted descriptively as explanations of model predictions rather than as causal effects (Table 4).

Discussion

The use of ML algorithms in obstetrics and gynecology has expanded rapidly in recent years, primarily in the context of supporting diagnostic and prognostic decision-making. Previous studies have demonstrated the utility of ML-based models in predicting preterm birth and in the early identification of intrauterine fetal hypoxia through automated analysis of cardiotocography (CTG) recordings.^{20,21} Additional research has explored the use of ML approaches for predicting rare but severe outcomes, such as stillbirth,²² as well as for optimizing clinical decision-making related to cesarean section indication, with potential implications for resource allocation in highly specialized obstetric units.²³ Collectively, these findings suggest that ML techniques can effectively integrate complex clinical signals to support time-sensitive obstetric decision-making when appropriately aligned with clinical workflows.

This study extends prior work on ML-based decision support in obstetrics to the prehospital emergency care setting, a context that has been comparatively underrepresented

Table 3. Shapley Additive Explanations (SHAP) analysis results for the penalized logistic regression model

Model	Feature	n	shap_mean	shap_sd	shap_abs_mean	shap_abs_sd	coef	Zero reason	coef_abs
Logistic regression	age [years]	200	0	0	0	0	0	zero coef (elastic net)	0
	respiratory rate	200	0	0	0	0	0	zero coef (elastic net)	0
	saturation [%]	200	0	0	0	0	0	zero coef (elastic net)	0
	systolic blood pressure [mm Hg]	200	0	0	0	0	0	zero coef (elastic net)	0
	diastolic blood pressure [mm Hg]	200	0	0	0	0	0	zero coef (elastic net)	0
	heart rate [bpm]	200	-0.017	0.079	0.062	0.052	0.084	nonzero	0.084
	glucose [mg/dL]	200	-0.014	0.045	0.026	0.039	0.020	nonzero	0.020
	gestational week	200	-0.003	0.059	0.039	0.045	0.044	nonzero	0.044
	number of pregnancies	200	0	0	0	0	0	zero coef (elastic net)	0
	number of labors	200	<0.001	0.013	0.010	0.009	0.016	nonzero	0.016
	public location	200	0	0	0	0	0	zero coef (elastic net)	0
	multipara	200	0	0	0	0	0	zero coef (elastic net)	0
	complications during pregnancy: yes	200	0	0	0	0	0	zero coef (elastic net)	0
	medical care during pregnancy: yes	200	0	0	0	0	0	zero coef (elastic net)	0
	stage of labor: 2	200	-0.007	0.427	0.266	0.332	1.388	nonzero	1.388
	bleeding	200	0	0	0	0	0	zero coef (elastic net)	0
	any fetal movement problem	200	0	0	0	0	0	zero coef (elastic net)	0
	amniotic fluid preserved	200	-0.002	0.220	0.220	<0.001	-0.439	nonzero	0.439
	comorbidities present	200	0	0	0	0	0	zero coef (elastic net)	0
	gestational diabetes	200	0	0	0	0	0	zero coef (elastic net)	0
gestational hypertension	200	0	0	0	0	0	zero coef (elastic net)	0	

Values are reported as mean ± standard deviation (SD) of SHAP values across a fixed random subsample of training observations (n = 200), together with the mean ±SD of absolute SHAP values. SHAP values quantify the additive contribution of each feature to the model output for the positive class under the applied preprocessing pipeline. Coefficients (β) are reported only for the elastic-net penalized logistic regression model and correspond to standardized effects used by the fitted model; $\beta = 0$ indicates complete shrinkage under regularization. For the random forest model, regression coefficients are not defined (N/A). SHAP analyses are provided for interpretability and should be interpreted descriptively rather than causally. N/A – not applicable; n – number of training observations used in the subsample; shap_mean – mean SHAP value; shap_sd – standard deviation of SHAP values; shap_abs_mean – mean absolute SHAP value; shap_abs_sd – standard deviation of absolute SHAP values; coef – standardized regression coefficient (β); coef_abs – absolute value of the standardized regression coefficient (β).

Table 4. Shapley Additive Explanations (SHAP) analysis results for the random forest classifier

Model	Feature	n	shap_mean	shap_sd	shap_abs_mean	shap_abs_sd	coef	Zero reason	coef_abs
Random forest classifier	age [years]	200	-0.001	0.003	0.002	0.003	-	N/A	-
	respiratory rate	200	-0.001	0.004	0.002	0.003	-	N/A	-
	saturation [%]	200	<0.001	0.005	0.002	0.005	-	N/A	-
	systolic blood pressure [mm Hg]	200	-0.001	0.003	0.002	0.002	-	N/A	-
	diastolic blood pressure [mm Hg]	200	<0.001	0.006	0.002	0.005	-	N/A	-
	heart rate [bpm]	200	-0.001	0.008	0.005	0.007	-	N/A	-
	glucose [mg/dL]	200	-0.001	0.005	0.003	0.005	-	N/A	-
	gestational_week	200	<0.001	0.005	0.003	0.004	-	N/A	-
	number of pregnancies	200	<0.001	0.004	0.002	0.004	-	N/A	-
	number of labors	200	-0.001	0.003	0.001	0.002	-	N/A	-
	public location	200	-0.001	0.002	0.001	0.002	-	N/A	-
	multipara	200	-0.001	0.001	<0.001	0.001	-	N/A	-
	complications during pregnancy	200	-0.001	0.002	<0.001	0.001	-	N/A	-
	medical care during pregnancy	200	<0.001	0.003	<0.001	0.003	-	N/A	-
	stage of labor	200	-0.002	0.159	0.098	0.125	-	N/A	-
	bleeding	200	-0.001	<0.001	<0.001	<0.001	-	N/A	-
	any fetal movement problem	200	<0.001	<0.001	<0.001	<0.001	-	N/A	-
	amniotic fluid preserved	200	-0.001	0.035	0.027	0.023	-	N/A	-
	comorbidities present	200	<0.001	0.002	<0.001	0.002	-	N/A	-
gestational diabetes	200	<0.001	0.001	<0.001	0.001	-	N/A	-	
gestational hypertension	200	<0.001	<0.001	<0.001	<0.001	-	N/A	-	

Values are reported as mean \pm standard deviation (SD) of SHAP values across a fixed random subsample of training observations ($n = 200$), together with the mean \pm SD of absolute SHAP values. SHAP values quantify the additive contribution of each feature to the model output for the positive class under the applied preprocessing pipeline. Coefficients (β) are reported only for the elastic-net penalized logistic regression model and correspond to standardized effects used by the fitted model; $\beta = 0$ indicates complete shrinkage under regularization. For the random forest model, regression coefficients are not defined (N/A). SHAP analyses are provided for interpretability and should be interpreted descriptively rather than causally. N/A – not applicable; n – number of training observations used in the subsample; shap_mean – mean SHAP value; shap_sd – standard deviation of SHAP values; shap_abs_mean – mean absolute SHAP value; shap_abs_sd – standard deviation of absolute SHAP values; coef – standardized regression coefficient (β); coef_abs – absolute value of the standardized regression coefficient (β).

in existing research. Emergency medical services operate under substantial time pressure and with markedly limited diagnostic resources compared with hospital-based care. In this environment, clinical priorities focus on rapid patient stabilization and timely transport, and care is typically provided by general medical teams without direct access to obstetric or gynecologic specialists. These constraints increase uncertainty during labor assessment and may elevate the risk of suboptimal triage or transport decisions.²⁴

The high discriminatory performance observed in the evaluated models, including the RF classifier (ROC-AUC up to 0.98), suggests that ML-based approaches may provide useful decision support in prehospital obstetric care. Importantly, such systems are not intended to replace the clinical judgment of paramedics but rather to complement it by structuring and integrating routinely available clinical information in time-critical and cognitively demanding situations. Prior studies in emergency and acute care settings have similarly emphasized the potential role of algorithmic decision-support tools in reducing uncertainty and supporting triage under conditions of stress

and limited diagnostic resources.²⁴ As expected, variables directly related to the physiology and progression of labor – particularly stage of labor and rupture of membranes – received the highest relative importance across both the RF and SVC-RBF models. In prehospital settings, the identification of active labor or ruptured membranes represents a critical inflection point in obstetric triage, as it strongly influences decisions regarding transport vs on-scene delivery. This clinical relevance has been highlighted in previous observational studies of prehospital and emergency obstetric care, including the work of Eisenbrey et al.²⁵

Variables directly related to the physiological progression of labor – most notably stage of labor and rupture of membranes – were assigned the highest relative importance across both the RF and SVC-RBF models. In prehospital settings, the identification of active labor or ruptured membranes represents a critical inflection point in obstetric triage, as it strongly influences decisions regarding transport vs on-scene delivery, a finding consistent with prior observational studies in emergency obstetric care.^{25,26} Importantly, although the clinical association between these features and

imminent delivery is well established, their prominence in SHAP-based explanations indicates that the models preferentially relied on specific, directly observable indicators of labor progression rather than on more general physiological measures such as vital signs. This pattern suggests alignment between model behavior and routine clinical reasoning, without implying causal interpretation. One notable finding of this study was the high ranking of blood glucose levels in feature-importance analyses, exceeding that of heart rate and respiratory rate. Although blood glucose is routinely assessed in emergency medical services primarily in the context of diabetic conditions, the models identified it as an informative predictor of imminent delivery. This observation is biologically plausible, as maternal glucose concentrations have been shown to increase with advancing labor and to peak during the 2nd stage of labor, reflecting metabolic stress, physical exertion, and catecholamine-mediated mobilization of energy substrates.^{27,28}

In prehospital emergency settings, where detailed obstetric examination may be limited, blood glucose measured with a standard glucometer represents an easily obtainable and objective variable that contributed meaningful predictive information.²⁹ In contrast, general physiological parameters such as heart rate and respiratory rate, while essential for overall patient assessment, showed lower relative importance for predicting imminent delivery.^{30,31} As demonstrated in this study, the stage of labor remained the single most informative predictor; however, a central advantage of ML models lies in their capacity to integrate multiple complementary inputs. The combination of labor-related features with readily available physiological measures, including blood glucose and heart rate, enabled more informative risk stratification than reliance on any single parameter alone.²⁸

Another aspect examined in this study was the contribution of basic vital signs, including blood pressure, heart rate, and respiratory rate. From a physiological perspective, these parameters are influenced by labor-related pain, stress, and physical exertion, which increase metabolic demand and are commonly accompanied by tachycardia and tachypnea. Childbirth can therefore be considered a transient hypermetabolic state requiring rapid cardiovascular and respiratory adaptation.³²

A study by Söhnchen et al. demonstrated that maternal heart rate during labor may reach levels comparable to those observed during physical exertion, particularly during the pushing phase, reflecting the substantial cardiovascular load associated with childbirth.³³ Despite this well-established physiological response, heart rate and related vital signs received relatively low predictive weight in the ML models. This finding likely reflects their limited diagnostic specificity in prehospital emergency settings, as tachycardia in parturient patients is a multifactorial phenomenon that may arise from advanced labor, pain, anxiety, dehydration, or other nonspecific stressors.

In the analyzed models, variables such as blood pressure, heart rate, and respiratory rate exhibited characteristics

of high-variability features, limiting their ability to reliably distinguish between imminent delivery and earlier stages of labor. Consequently, although these vital signs remain essential for monitoring maternal safety and identifying potential complications, their contribution to predicting sudden out-of-hospital birth was comparatively limited. In contrast, more stable and labor-specific features, such as stage of labor and amniotic fluid status, provided greater discriminatory information for prediction, as they are less influenced by nonspecific stress responses.

The clinical utility of ML models in obstetrics is increasingly dependent on the integration of diverse physiological and demographic variables. In our study, the model demonstrated high performance despite the absence of maternal body mass index (BMI) and ethnicity data. However, as emphasized in recent literature, pre-pregnancy BMI is a critical determinant of labor progression and the risk of emergency interventions.³⁴ Furthermore, the algorithmic fairness debate highlights that clinical tools developed using ethnically homogeneous populations may exhibit performance gaps when applied to more diverse cohorts.³⁵ In the context of prehospital care, where rapid decision-making is essential, incorporating these variables could further enhance the model's sensitivity in predicting precipitous labor across different patient profiles.³⁶

Although AI-based models show significant potential, their implementation must consider the risk of overreliance, particularly among less experienced clinicians. Uncritical reliance on algorithmic suggestions – often referred to as automation bias – may lead to clinical errors if model outputs are not integrated with a comprehensive clinical assessment of the patient. Furthermore, the introduction of such tools into prehospital emergency care raises important medicolegal questions regarding liability in the event of adverse outcomes.³⁷ Whether legal conflict arises from following an erroneous AI recommendation or from ignoring a correct one remains a complex challenge for future regulatory frameworks. Therefore, these tools should be clearly defined as clinical decision support systems (CDSS) that provide additional information rather than definitive instructions, ensuring that the ultimate clinical and legal responsibility remains with the healthcare professional. Recent studies emphasize that the deployment of AI in high-stakes environments such as EMS must address the risk of automation bias, particularly in contexts where clinicians may rely excessively on algorithmic outputs. This issue is especially relevant in obstetrics, where medicolegal liability remains a major concern for practitioners.^{38,39}

The results of our study indicate substantial potential for the application of ML models in prehospital obstetric care. It should be emphasized that the proposed approach is not intended to replace the clinical decision-making autonomy of healthcare professionals, but rather to function as a CDSS within the prehospital care environment. A key advantage of the model is its ability to integrate multiple

physiological signals that may be difficult to interpret under conditions of fatigue, stress, and time pressure. Blood glucose levels emerged as an important marker associated with labor progression, as they are physiologically related to metabolic demand, physical exertion, and hormonal responses during labor. However, because blood glucose measurements were selectively recorded and exhibited a high proportion of missing values, their apparent importance may partly reflect measurement patterns. Therefore, these findings should be interpreted as predictive rather than causal. Further research conducted on larger prospective cohorts is necessary to fully validate the model before its potential implementation in EMS system.

Limitations of the study

This study has several limitations. First, its retrospective observational design and reliance on routinely collected EMS documentation may be subject to incomplete or inconsistent data recording, which could influence model performance. Second, the study was conducted within a single national emergency medical system, potentially limiting generalizability to other healthcare settings with different organizational structures, staffing models, or prehospital obstetric protocols. Third, the high predictive performance observed in this study partly reflects the inclusion of variables temporally close to the outcome, such as stage of labor. Although sensitivity analyses demonstrated that meaningful predictive information persisted after exclusion of these variables, performance estimates should be interpreted in the context of this temporal proximity. Fourth, the dataset did not include information on maternal BMI or gestational weight gain, as these parameters are generally not recorded in prehospital documentation. Fifth, the study population consisted predominantly of individuals of European ancestry, which may limit the generalizability of the model to more diverse populations. Finally, the models were developed and validated using internal cross-validation; external validation in independent and prospective cohorts is required before clinical implementation can be considered.

Conclusions

Machine learning models trained on routinely collected prehospital obstetric data demonstrated high discriminatory performance for predicting out-of-hospital birth events. Importantly, model explanations indicated that predictions were driven primarily by clinically intuitive and readily observable indicators of labor progression, suggesting alignment between model behavior and established prehospital obstetric assessment. These findings suggest that ML-based approaches may support prehospital clinical decision-making by integrating multiple complementary clinical cues under time pressure, without

replacing clinical judgment. Variables such as blood glucose contributed additional predictive information beyond general vital signs, highlighting the potential value of incorporating easily obtainable physiological measures into risk stratification. External validation in larger prospective cohorts is required before clinical implementation can be considered.

Supplementary data

The supplementary materials are available at <https://doi.org/10.5281/zenodo.18959031>. The package contains the following files:

Supplementary Table 1. Fold-level performance metrics and hyperparameter configurations.

Supplementary Table 2. Features retained in the model across all cross-validation folds.

Supplementary Table 3. Fold-level performance metrics obtained from the outer loop of nested cross-validation.

Supplementary Table 4. Fold-specific hyperparameter configurations selected during nested cross-validation.

Supplementary Table 5. Permutation-based feature importance across evaluated ML models.

Supplementary Table 6. Sensitivity analysis excluding stage of labor (2nd stage): performance of restricted models and comparison with full models.

Supplementary Table 7. Permutation-based feature importance shift between full and restricted models excluding stage of labor (stage 2).

Data Availability Statement

The datasets supporting the findings of the current study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.18959181>.

Consent for publication of personal information


Not applicable.


Use of AI and AI-assisted technologies


Generative AI was used to assist in translating the manuscript into English, utilizing OpenAI's ChatGPT 5.2. The authors take full responsibility for this use.

ORCID iDs

Joanna Wach  <https://orcid.org/0009-0003-6544-7225>

Łukasz Lewandowski  <https://orcid.org/0000-0002-7624-631X>

Jakub Staniczek  <https://orcid.org/0000-0002-0740-7167>

Raúl Juárez-Vela  <https://orcid.org/0000-0003-3597-2048>

Michał Czapla  <https://orcid.org/0000-0002-4245-5420>

References

1. Quattrocchi P. Policies and practices on out-of-hospital birth: A review of qualitative studies in the time of coronavirus. *Curr Sex Health Rep.* 2022;15(1):36–48. doi:10.1007/s11930-022-00354-7

2. Hill M, Miles A, Flanagan B, Hansen S, Mills B, Hopper L. Out-of-hospital births and the experiences of emergency ambulance clinicians and birthing parents: A scoping review of the literature. *BMJ Open*. 2025;15(5):e086967. doi:10.1136/bmjopen-2024-086967
3. Sheikhi RA, Heidari M. The challenges of delivery in pre-hospital emergency medical services ambulances in Iran: A qualitative study. *BMC Emerg Med*. 2024;24(1):156. doi:10.1186/s12873-024-01073-z
4. Strózik M, Wiciak H, Raczynski A, Smereka J. Emergency medical team interventions in Poland during out-of-hospital deliveries: A retrospective analysis. *Adv Clin Exp Med*. 2024;34(10):1731–1737. doi:10.17219/acem/184141
5. Statistics Poland. Population: Size and structure and vital statistics in Poland by territorial division in 2023. As of 31 December. Warsaw, Poland: Statistics Poland; 2024. <https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-stan-i-struktura-ludnosci-oraz-ruch-naturalny-w-przekroju-terytorialnym-w-2023-r-stan-w-dniu-31-12,6,36.html>
6. Van De Sande D, Van Genderen ME, Smit JM, et al. Developing, implementing and governing artificial intelligence in medicine: A step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform*. 2022;29(1):e100495. doi:10.1136/bmjhci-2021-100495
7. Rong G, Mendez A, Bou Assi E, Zhao B, Sawan M. Artificial intelligence in healthcare: Review and prediction case studies. *Engineering*. 2020;6(3):291–301. doi:10.1016/j.eng.2019.08.015
8. Gebhard J, Graf J, Abele H, Pauluschke-Fröhlich J. Einbindung und Umgang von Notfallsanitätern bei ungeplanten außerklinischen Geburten: Ein Online-Survey. *Gesundheitswesen*. 2024;86(1):18–27. doi:10.1055/a-2183-5837
9. Owusu FO, Addai-Manu H, Agbedinu ES, et al. Prediction of caesarean section birth using machine learning algorithms among pregnant women in a district hospital in Ghana. *BMC Pregnancy Childbirth*. 2025;25(1):690. doi:10.1186/s12884-025-07716-8
10. Tadepalli K, Das A, Meena T, Roy S. Bridging gaps in artificial intelligence adoption for maternal-fetal and obstetric care: Unveiling transformative capabilities and challenges. *Comput Methods Programs Biomed*. 2025;263:108682. doi:10.1016/j.cmpb.2025.108682
11. El Arab RA, Al Moosa OA, Albahrani Z, Alkhalil I, Somerville J, Abudadas F. Integrating artificial intelligence into perinatal care pathways: A scoping review of reviews of applications, outcomes, and equity. *Nurs Rep*. 2025;15(8):281. doi:10.3390/nursrep15080281
12. Yaseen I, Rather R. A theoretical exploration of artificial intelligence's impact on fetomaternal health from conception to delivery. *Int J Womens Health*. 2024;16:903–915. doi:10.2147/IJWH.S454127
13. Tzitoridou-Chatzopoulou M, Zournatzidou G, Kourakos M. Predicting future birth rates with the use of an adaptive machine learning algorithm: A forecasting experiment for Scotland. *Int J Environ Res Public Health*. 2024;21(7):841. doi:10.3390/ijerph21070841
14. Abdi F, Roozbeh N, Darsareh F, Mehrnoush V, Vahidi Farashah MS, Montazeri F. Developing a prognostic model for predicting preterm birth using a machine learning algorithm. *BMC Pregnancy Childbirth*. 2025;25(1):974. doi:10.1186/s12884-025-08136-4
15. Patel DJ, Chaudhari K, Acharya N, Shrivastava D, Muneeba S. Artificial intelligence in obstetrics and gynecology: Transforming care and outcomes. *Cureus*. 2024;16(7):e64725. doi:10.7759/cureus.64725
16. Sim JZT, Fong QW, Huang W, Tan CH. Machine learning in medicine: What clinicians should know. *Singapore Med J*. 2023;64(2):91–97. doi:10.11622/smedj.2021054
17. Withanarachchie V, Todd V, Dicker B, Maessen SE. Navigating emotions, communication, and pain during prehospital labour: A mixed-methods survey with emergency ambulance services. *BMC Emerg Med*. 2025;25(1):83. doi:10.1186/s12873-025-01236-6
18. Beard DT, Ladd M, Jenkins SM, Kahwaji CI. EMS prehospital deliveries. In: *StatPearls*. Treasure Island, USA: StatPearls Publishing; 2026. Bookshelf ID: NBK525996. <http://www.ncbi.nlm.nih.gov/books/NBK525996>. Accessed March 9, 2026.
19. von Elm E, Altman DG, Egger M, et al. Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806–808. doi:10.1136/bmj.39335.541782.AD
20. Włodarczyk T, Płotka S, Szczepański T, et al. Machine learning methods for preterm birth prediction: A review. *Electronics*. 2021;10(5):586. doi:10.3390/electronics10050586
21. Olayemi OC, Olasehinde OO. Machine learning prediction of fetal health status from cardiocography examination in developing healthcare contexts. *J Comput Sci Res*. 2024;6(1):43–53. doi:10.30564/jcsr.v6i1.6242
22. Gunenc O, Dogru S, Yaman FK, Ezveci H, Metin US, Acar A. The application of machine learning models to predict stillbirths. *Medicina (Kaunas)*. 2025;61(3):472. doi:10.3390/medicina61030472
23. Cheng R, Feng B, Zheng Y, et al. MvBody: Multi-view-based hybrid transformer using optical 3D body scan for explainable cesarean section prediction [posted online as preprint on November 5, 2025]. *arXiv*. 2025. doi:10.48550/ARXIV.2511.03212
24. Poranen A, Kouvonen A, Nordquist H. Human errors in emergency medical services: A qualitative analysis of contributing factors. *Scand J Trauma Resusc Emerg Med*. 2024;32(1):78. doi:10.1186/s13049-024-01253-7
25. Eisenbrey D, Dunne RB, Fales W, Torossian K, Swor R. Describing pre-hospital deliveries in the state of Michigan. *Cureus*. 2022;14(7):e26723. doi:10.7759/cureus.26723
26. Abdalla Elsheikh NE, M Osman HM, Alfaki Ahmed SA, et al. Effectiveness and implementation of obstetric triage during pregnancy and childbirth: A systematic review. *Cureus*. 2025;17(8):e89215. doi:10.7759/cureus.89215
27. Risberg A, Sjöquist M, Wedenberg K, Larsson A. Elevated glucose levels in early puerperium, and association with high cortisol levels during parturition. *Scand J Clin Lab Invest*. 2016;76(4):309–312. doi:10.3109/00365513.2016.1149881
28. Bitar G, Fishel Bartal M. Intrapartum glycemic control and clinical outcomes. *Diabetes Spectr*. 2025;38(4):400–406. doi:10.2337/dsi25-0010
29. Deng Y, Wu H, Ng NYH, et al. Association between maternal glucose levels in pregnancy and offspring's metabolism and adiposity: An 18-year birth cohort study. *Diabetologia*. 2025;68(10):2205–2216. doi:10.1007/s00125-025-06476-6
30. Erickson EN, Gotlieb N, Pereira LM, Myatt L, Mosquera-Lopez C, Jacobs PG. Predicting labor onset relative to the estimated date of delivery using smart ring physiological data. *NPJ Digit Med*. 2023; 6(1):153. doi:10.1038/s41746-023-00902-y
31. Widatalla N, Keenan E, Palaniswami M, Khandoker A. Investigating the role of maternal heart rate variability in the onset of labor. *Front Med (Lausanne)*. 2025;12:1659620. doi:10.3389/fmed.2025.1659620
32. Green LJ, Pullon R, Mackillop LH, et al. Postpartum-specific vital sign reference ranges. *Obstet Gynecol*. 2021;137(2):295–304. doi:10.1097/AOG.0000000000004239
33. Söhnchen N, Melzer K, Tejada BMD, et al. Maternal heart rate changes during labour. *Eur J Obstet Gynecol Reprod Biol*. 2011;158(2):173–178. doi:10.1016/j.ejogrb.2011.04.038
34. Edwards SE, Cohen R, Zhao Z, et al. Characterizing labor progression and duration according to maternal body mass index [published online as ahead of print on November 8, 2025]. *Am J Obstet Gynecol*. 2025. doi:10.1016/j.ajog.2025.11.003
35. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight: Reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874–882. doi:10.1056/NEJMms2004740
36. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. doi:10.1126/science.aax2342
37. Maliha G, Gerke S, Cohen IG, Parikh RB. Artificial intelligence and liability in medicine: Balancing safety and innovation. *Milbank Q*. 2021;99(3):629–647. doi:10.1111/1468-0009.12504
38. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among healthcare workers: Systematic review. *J Med Internet Res*. 2025;27:e69678–e69678. doi:10.2196/69678
39. Fischer A, Rietveld A, Teunissen P, Hoogendoorn M, Bakker P. What is the future of artificial intelligence in obstetrics? A qualitative study among healthcare professionals. *BMJ Open*. 2023;13(10):e076017. doi:10.1136/bmjopen-2023-076017