Original Article
Yonsei Med J 2022 Jul;63(7):692-700
https://doi.org/10.3349/ymj.2022.63.7.692
Yonsei Medical Journal
YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Machine Learning Model for Classifying the Results of Fetal Cardiotocography Conducted in High-Risk Pregnancies

Tae Jun Park[1*], Hye Jin Chang[2*], Byung Jin Choi[1], Jung Ah Jung[2], Seongwoo Kang[2], Seokyoung Yoon[2], Miran Kim[2,3], and Dukyong Yoon[4,5,6]

Departments of [1]Biomedical Informatics and [2]Obstetrics and Gynecology, Ajou University School of Medicine, Suwon;
[3]COSMOSWHALE Inc., Ansan;
[4]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul;
[5]Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin;
[6]BUD.on Inc., Jeonju, Korea.

Purpose: Fetal well-being is usually assessed via fetal heart rate (FHR) monitoring during the antepartum period. However, the interpretation of FHR is a complex and subjective process with low reliability. This study developed a machine learning model that can classify fetal cardiotocography results as normal or abnormal.

Materials and Methods: In total, 17492 fetal cardiotocography results were obtained from Ajou University Hospital and 100 fetal cardiotocography results from Czech Technical University and University Hospital in Brno. Board-certified physicians then reviewed the fetal cardiotocography results and labeled 1456 of them as gold-standard; these results were used to train and validate the model. The remaining results were used to validate the clinical effectiveness of the model with the actual outcome.

Results: In a test dataset, our model achieved an area under the receiver operating characteristic curve (AUROC) of 0.89 and area under the precision-recall curve (AUPRC) of 0.73 in an internal validation dataset. An average AUROC of 0.73 and average AUPRC of 0.40 were achieved in the external validation dataset. Fetus abnormality score, as calculated from the continuous fetal cardiotocography results, was significantly associated with actual clinical outcomes [intrauterine growth restriction: odds ratio, 3.626 (p=0.031); Apgar score 1 min: odds ratio, 9.523 (p<0.001), Apgar score 5 min: odds ratio, 11.49 (p=0.001), and fetal distress: odds ratio, 23.09 (p<0.001)].

Conclusion: The machine learning model developed in this study showed precision in classifying FHR signals. This suggests that the model can be applied to medical devices as a screening tool for monitoring fetal status.

Key Words: Cardiotocography, high-risk-pregnancy, machine learning

Received: November 17, 2021 **Revised:** March 24, 2022
Accepted: April 4, 2022

Co-corresponding authors: Miran Kim, MD, PhD, Department of Obstetrics and Gynecology, Ajou University School of Medicine, 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea.
Tel: 82-31-219-5300, Fax: 82-31-219-5250, E-mail: kmr5300@ajou.ac.kr and Dukyong Yoon, MD, PhD, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
Tel: 82-31-5189-8450, Fax: 82-31-5189-8566, E-mail: dukyong.yoon@yonsei.ac.kr

*Tae Jun Park and Hye Jin Chang contributed equally to this work.
•MK is the founder and an employee of COSMOSWHALE Inc. DY is the founder and an employee of BUD.on Inc. The other authors declare no conflicts of interests.

INTRODUCTION

Fetal cardiotocography is widely used to monitor fetal status during the intrapartum period.[1] Specifically, it continuously monitors fetal heart rate (FHR) and uterine contraction through ultrasound transducers that capture both FHR, as transmitted via the maternal abdomen, and the pressure intensity of uterine contractions. Although more than 50 years have passed since the first clinical usage of fetal cardiotocography, it remains valuable as an indirect evaluation approach, especially as direct approaches, such as the fetal scalp electrode,[2] are considered too invasive. This is important, as many issues require monitoring. For example, fetuses with intrauterine growth restriction (IUGR) are at high risk of distress. Clinicians should

therefore pay attention to fetal conditions during antenatal care that require close monitoring of FHR.[3] Moreover, maturation of the autonomic nervous system is a key controller in fetal cardiovascular regulation,[4] with reports showing a positive association between FHR variability and gestational age.[5] As IUGR is a pathologic condition in which fetuses do not grow as expected, based on their gestational age, being able to screen for this issue using fetal cardiotocography would be useful.

Even though cardiotocography is a safe evaluation method, it requires interpretation by obstetrics experts, which limits its potential for widespread continuous interpretation or self-monitoring by pregnant woman while at home. However, it is critical to perform fetal cardiotocography in high-risk pregnancies, particularly to assess fetal well-being and determine necessary clinical interventions. Some high-risk pregnancies may even warrant three or more cardiotocography sessions each day to ensure appropriate observation, thus requiring hospitalization. However, limited medical resources make it difficult to do this for every pregnancy.

To overcome the current limitations of the usability of cardiotocography, previous studies have attempted to implement computationally automatic interpretations.[6,7] However, studies on interpreting fetal cardiotocography results were based on hand-crafted features, which are only useful for checking local acceleration and deceleration.[8,9] Moreover, even though models developed in existing studies used Physionet open data [Czech Technical University (CTU)-University Hospital in Brno (UHB), FECGSYNDB],[10,11] they did not provide gold-standard interpretations, nor were there sufficient amounts of data for training a machine learning model to extract useful information from a raw cardiotocography. In summary, existing models trained using CUT-UHB data are limited in terms of reliability, reproducibility, and generalizability. Thus, in this study, we aimed to develop a machine learning model that could be used to assess pregnancy hazards (IUGR, fetal distress, and Apgar score) based on fetal cardiotocography results.

MATERIALS AND METHODS

This study received approval from the Institutional Review Board of Ajou University Hospital (IRB No. AJIRB-MED-MDB-20-152). The requirement for informed consent was waived because this study retrospectively used anonymized data.

Data source and preparation

This study investigated fetal cardiotocography data obtained from Ajou University Hospital Obstetrics and Gynecology from January 2017 to December 2019 (Fig. 1). More specifically, 17492 fetal cardiotocography results were available. This study used 1456 of these results from 665 mothers, which were interpreted by three obstetricians to create a qualified dataset.

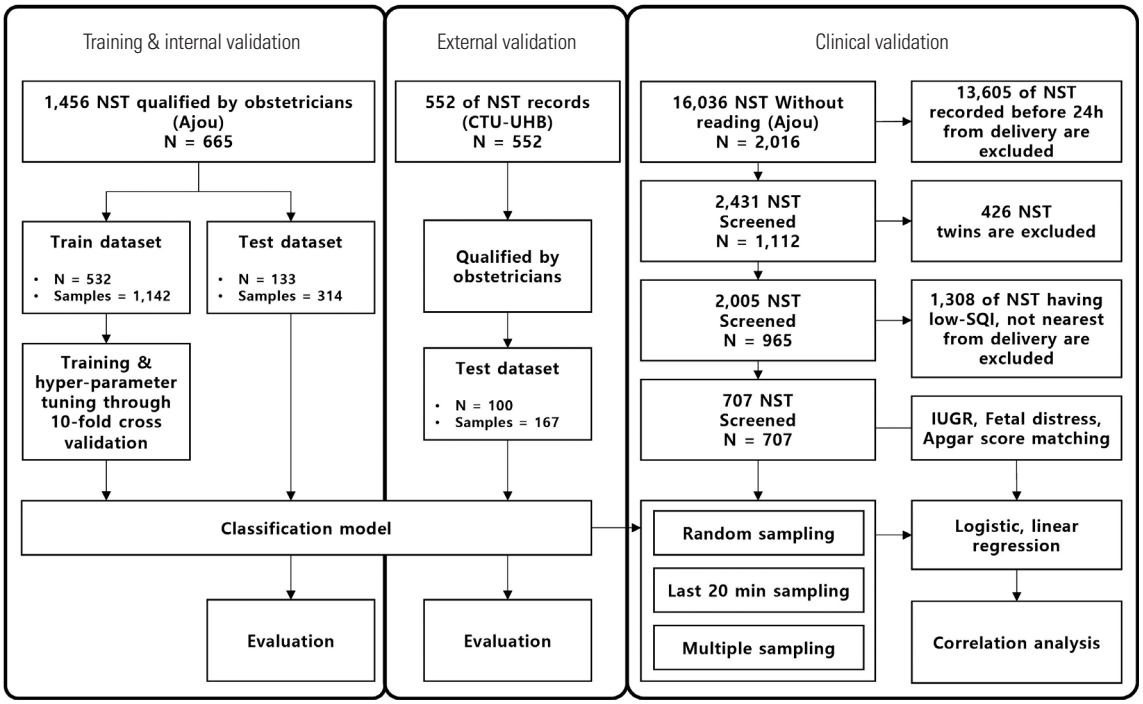

Fig. 1. Study design. Datasets were gathered and read by a team of obstetricians. Readings were recorded to evaluate 1456 fetal cardiotocography (qualified fetal cardiotocography, left side of the flowchart), while the other fetal cardiotocography results without the reading were used to evaluate the model in clinical situations (clinical validation dataset, right side of the flowchart). For the qualified fetal cardiotocography dataset, the classification model was trained to find abnormal data. Clinical validation datasets were created to represent three clinical situations according to the time window selected from the fetal cardiotocography results. NST, non stress test; CTU-UHB, Czech Technical University-University Hospital in Brno; SQI, signal quality index; IUGR, intrauterine growth restriction.

The obstetricians included three board-certified obstetricians, a 10-year senior obstetrician, and a 20-year senior obstetrician (qualified in fetal cardiotocography). The dataset was divided into three, and each segment was read by one obstetrician. After that, the two senior experts reviewed each data segment together. If the review result differed from the previous reading, the reading was revised after consultation with the two senior doctors. Each obstetrician selected and extracted records with durations of 24 minutes from the fetal cardiotocography results, so that the model could learn the data in the same circumstances under which the test commonly progresses.

For external validation of the model, fetal cardiotocography data were extracted from the CTU-UHB database. The database comprises 552 cardiotocography recordings that were carefully selected from 9164 recordings collected between 2010 and 2012 at the CTU in Prague and the UHB. Among them, we selected up to 100th from the front. These samples were then interpreted by the same obstetricians.

Fetal cardiotocography readings contain the following six shapes: reactive, non-reactive, early deceleration, variable deceleration, late deceleration, and prolonged deceleration (Supplementary Fig. 1, only online).[12] To overcome data imbalances, all readings except "Reactive" and "Mild variable deceleration" were labeled as abnormal. The model was then trained to classify abnormal from normal fetal cardiotocography readings.

We trained and validated the machine learning-based classification model using this qualified fetal cardiotocography data. If a single fetal cardiotocography reading was longer than 24 minutes, multiple data points were used at 24-minute intervals to ensure that they did not overlap. Moreover, we ensured that data from the same fetus were not included in both the training and testing datasets. This enabled us to obtain a number of samples to train the model to classify abnormal waveform data. As a result, the model was trained using 1142 fetal cardiotocography results (205 abnormal and 937 normal) from 80% of the mothers; it was then evaluated internally, based on 314 fetal cardiotocography results (53 abnormal and 261 normal) from 20% of the mothers. The external validation set was constructed with 167 fetal cardiotocography results (93 abnormal and 74 normal).

Next, we verified whether the abnormal probability score computed by the model was clinically meaningful. We built a dataset by matching the 707 non-read fetal cardiotocography results with clinical information (e.g., Apgar score, IUGR, and fetal distress). In conjunction with the clinical information, the fetal cardiotocography results of non-twins and of those recorded within 24 hours of childbirth were selected for preprocessing. Thereafter, the model was tested in three clinical settings: immediately before childbirth, at a randomized time point, and all time before childbirth.

To convert the PDF images in which all of the fetal cardiotocography results were saved into forms suitable for analysis, the pixels representing waveforms were separated from the background using Red Green Blue (RGB) channel differences. Next, the Hough transform algorithm and pixel-wise value transform method were used to calculate heart rate according to the relative position of the separated pixels (Fig. 2, Supplementary Material 1, and Supplementary Fig. 2, only online).[13] The code for this procedure is available at https://github.com/CMI-Laboratory/Fetal_cardiotocography_extract_code. We screened the fetal cardiotocography results based on several criteria: first, to train the model with well-preserved data, we excluded those with continuous blanks lasting more than 5 minutes and those of which 50% in total were blank. The others were interpolated via linear methods. Second, because we frequently encountered noise at the start and end of the test, the data were horizontally cropped into 20 minute intervals.

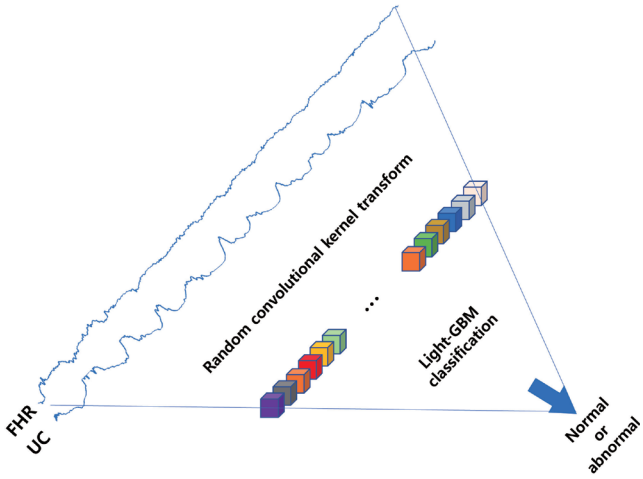

Fig. 2. Classification model structure. As an input of the classification model, 2-channel waveform data were created based on the fetal heart rate (FHR) wave and uterine contraction wave (UC). Randomly initialised convolutional kernels were applied to the input and transformed to two features that were used to classify the fetus status by lightGBM classification machine learning model.

Developing the fetal cardiotocography interpretation model

We used the random convolutional kernel transform to extract features from FHR and uterine contractions recorded in the fetal cardiotocography results and to insert features into the models to determine whether the fetal cardiotocography was abnormal (Fig. 3). The random convolutional kernel transformation method uses convolutional kernels as a means of transformation.[14] In deep learning studies using waveform data, convolutional neural networks typically focus on simple representations, such as shapes, frequencies, and data variations.[15,16] However, training convolutional neural networks models is challenging, because a large amount of data is required.[17] As such, there is currently a lack of research on fetal status being monitored by waveform data.[18-20] We considered that random kernels could be used to extract features more efficiently, with less data. In this study, 10000 random kernels were created and

https://doi.org/10.3349/ymj.2022.63.7.692

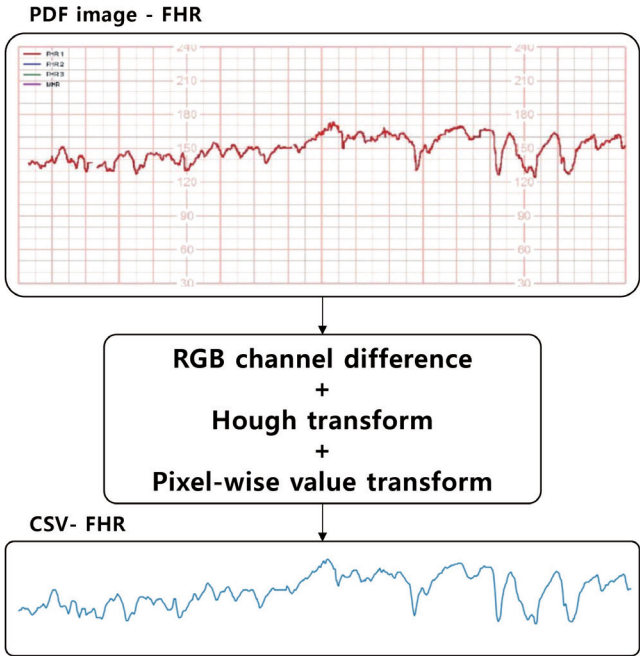

Fig. 3. Data pre-processing flow for transforming images to CSV format. Data pre-processing consisted of three image process techniques. Waveform pixels were extracted by using RGB channel differences. Then, the edges of the graph were identified through the Hough transform algorithm results. The value of the waveform pixels was calculated using the relative position of the pixel compared to the edge. Finally, data were saved to csv format. FHR, fetal heart rate; RGB, Red Green Blue; CSV, comma-separated values.

used by randomly combining dilation, length, padding, weight, and bias. A global max pooling value and positive predictive value were generated in one kernel, resulting in a classification with 20000 features (Supplementary Table 1, only online). Qualified datasets were divided into training sets and test sets to train and validate the model, with training sets divided into ten folders for cross validation to modify hyperparameters. We validated the model performance by soft-voting the test set of fetal cardiotocography results for a 10-fold trained model.

For classification, we used lightGBM methods, which are based on a gradient boosting machine that assembles multiple decision trees in a boosting manner to extract higher and more general classification performance. Among the gradient boosting machine-based models, we determined that there were advantages in computing speed and hyperparameter tuning speed using the lightGBM method. We set the learning rate of the model to 0.05, the number of leaves to 31, and the feature fraction to 0.9. We chose three evaluation indexes [area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and F1 score] to assess model performance.

The external validation test dataset was created by randomly selecting data by matching the ratio of normal and abnormal waveforms to the environment in which the model was trained (18%). After performing 10 random samplings with replacement, the model's performance was observed on 10 test sets. As with internal validation, the performance of the model is expressed as AUROC, AUPRC, and F1 scores, while the means and standard deviations in ten datasets are presented as results.

Clinical validation

Next, we tested whether the model's results were clinically meaningful when the model was simulated in an environment that mimics various clinical environments, using clinical validation. Using fetal cardiotocography, obstetricians can avoid unnecessary interventions in childbirth and any associated complications for mothers who are ready to give birth.[21] In the same context, we analyzed fetal cardiotocography data that were recorded within 24 hours of delivery to examine the model's ability to score the FHR of a fetus to be delivered imminently.

We used both fetal cardiotocography results and clinical data (Apgar score, IUGR, fetal distress) during birth (Fig. 4). We separated this into three datasets, in which the fetal cardiotocography progressed at random times or just prior to delivery or for an extended period of time. In other words, this included a random 20 minutes, the closest 20 minutes to delivery,[22] and all non-overlapping time windows, which obstetricians find challenging to read in clinical practice; these were selected to calculate the abnormal probability score.

We conducted multivariable logistic regression and multivariable linear regression to reveal any significant associations between the abnormal probability score evaluated by the model and real-world clinical information. Multivariable linear regressions were performed to observe the association between raw Apgar score and abnormal probability score. We also evaluated the association between the abnormal probability score and other variables (IUGR, fetal distress, and higher level of Apgar scores) via multivariable logistic regression. Apgar scores of 1 and 5 minutes for neonates were categorized as 1 if less than 7[23] and 0 otherwise. When using the dataset containing multiple time windows, we used the means of the abnormal probability scores calculated from each pregnancy. To adjust for age of the pregnancy and intrauterine pregnancy (IUP), these were included in multivariable regression models (Supplementary Material 2, only online). For the result of the multivariable logistic regression, we obtained the odds ratio by calculating the correlation coefficient of the abnormal probability score exponentially.

Next, we measured the performances of the model predicting clinical outcomes. In each situation (last, random, multiple), a calculated abnormal probability score was used to predict each of the clinical outcomes; we represented these outcomes as AUROC, AUPRC, and F1 scores.

Software

All analyses were conducted using Python (version 3.7, Python Software Foundation, Fredericksburg, VA, USA), the Python package pytorch 1.6.0, scikit learn 0.23.2, and optuna 2.3.0. The Python packages Matplotlib 3.2.2 and seaborn 0.10.0 were also

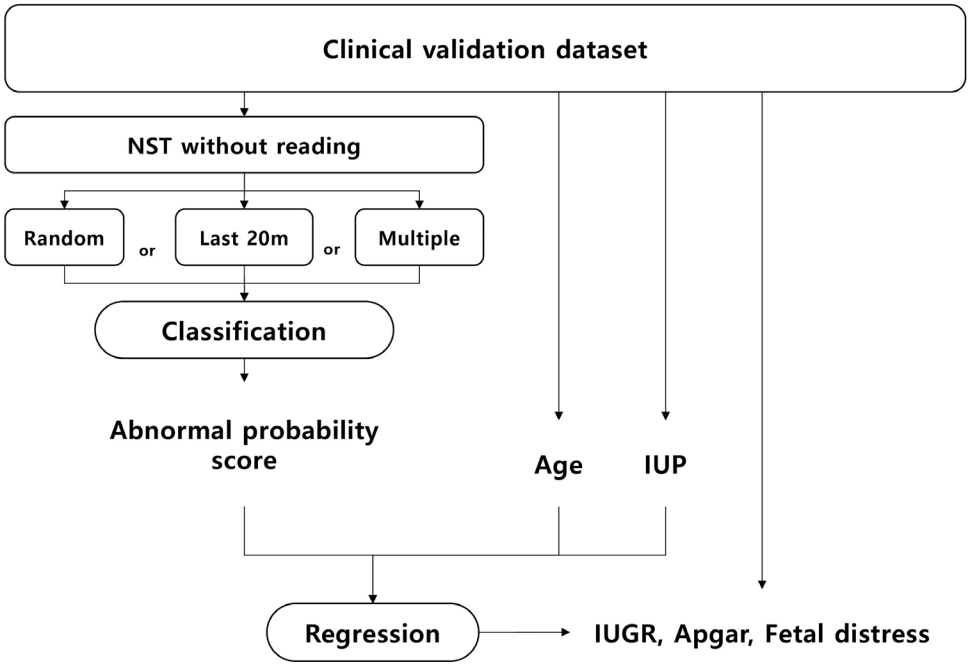

Fig. 4. Description of the clinical validation process. The clinical validation dataset consisted of fetal cardiotocography without reads combined with maternal demographic information (age, IUP) and postpartum neonatal status (IUGR, Apgar, Fetal distress). Based on delivery, the closest 20 minutes, random 20 minutes, and all non-overlapping time windows were selected to calculate the abnormal probability score. The calculated abnormal probability score, maternal age, and IUP were used as independent variables. Regression models were fitted on IUGR, Apgar, and fetal distress as dependent variables. NST, non stress test; IUP, intrauterine pregnancy; IUGR, intrauterine growth restriction.

used to visualise the data and results.

RESULTS

Model performance

The qualified fetal cardiotocography dataset for model development was constructed using data from 665 mothers. The average IUP was 35.2 weeks, while the average age was 34.4 years (Table 1). To verify the internal validity of the developed model, its performance was verified separately using the test set. In the test set for the qualified fetal cardiotocography dataset, the model performed as follows: AUROC=0.89, AUPRC=0.73, and F1 score=0.59 (Fig. 5A and B).

The average IUP and the average age in the CTU-UHB dataset were 39.9 weeks and 28.8 years, respectively (Table 1). For this dataset, the model performed as follows: average AUROC (SD)=0.73 (0.023), average AUPRC (SD)=0.40 (0.037), average F1 score (SD)=0.44 (0.022) (Fig. 5C and D).

Table 1. Model Development Cohort

	Ajou Univ. (n=536)	CTU-UHB (n=100)
Samples	1456	167
Normal	1198	74
Reactive	717	52
Mild variable deceleration	481	22
Abnormal	258	93
Non-reactive	119	7
Early deceleration	36	53
Moderate variable deceleration	32	21
Severe variable deceleration	18	4
Late deceleration	29	8
Prolonged deceleration	24	0
Age (yr)	34.4±4.7	28.8±4.2
IUP (weeks)	35.2±3.2	39.9±1.3

CTU-UHB, Czech Technical University-University Hospital in Brno; IUP, intrauterine pregnancy.
Data are presented as mean±standard deviation or n.

Clinical outcome validation

In this step, we used the dataset including the data of 707 mothers to accomplish clinical validation. The mean age of the mothers in the clinical validation cohort was 34.6 years, with 52.9% undergoing emergency deliveries. Further, 66% were delivered by cesarean section. Around 10% of the mothers had diabetes or hypertension. The 1-minute average Apgar was 7.2, while the 5-minute average Apgar was 8.5 (Table 2).

Except for IUGR, all clinical outcomes had significant positive relationships with the abnormal probability score obtained from every situation (Table 3). One minute or 5 minute average Apgar scores lower than 7 were positively associated with the abnormal probability score of our model in all situations. Meanwhile, the associations with raw Apgar scores (both 1 and 5 minutes) were also significant in all situations. Because lower Apgar scores represent clinical situations that require in-

https://doi.org/10.3349/ymj.2022.63.7.692

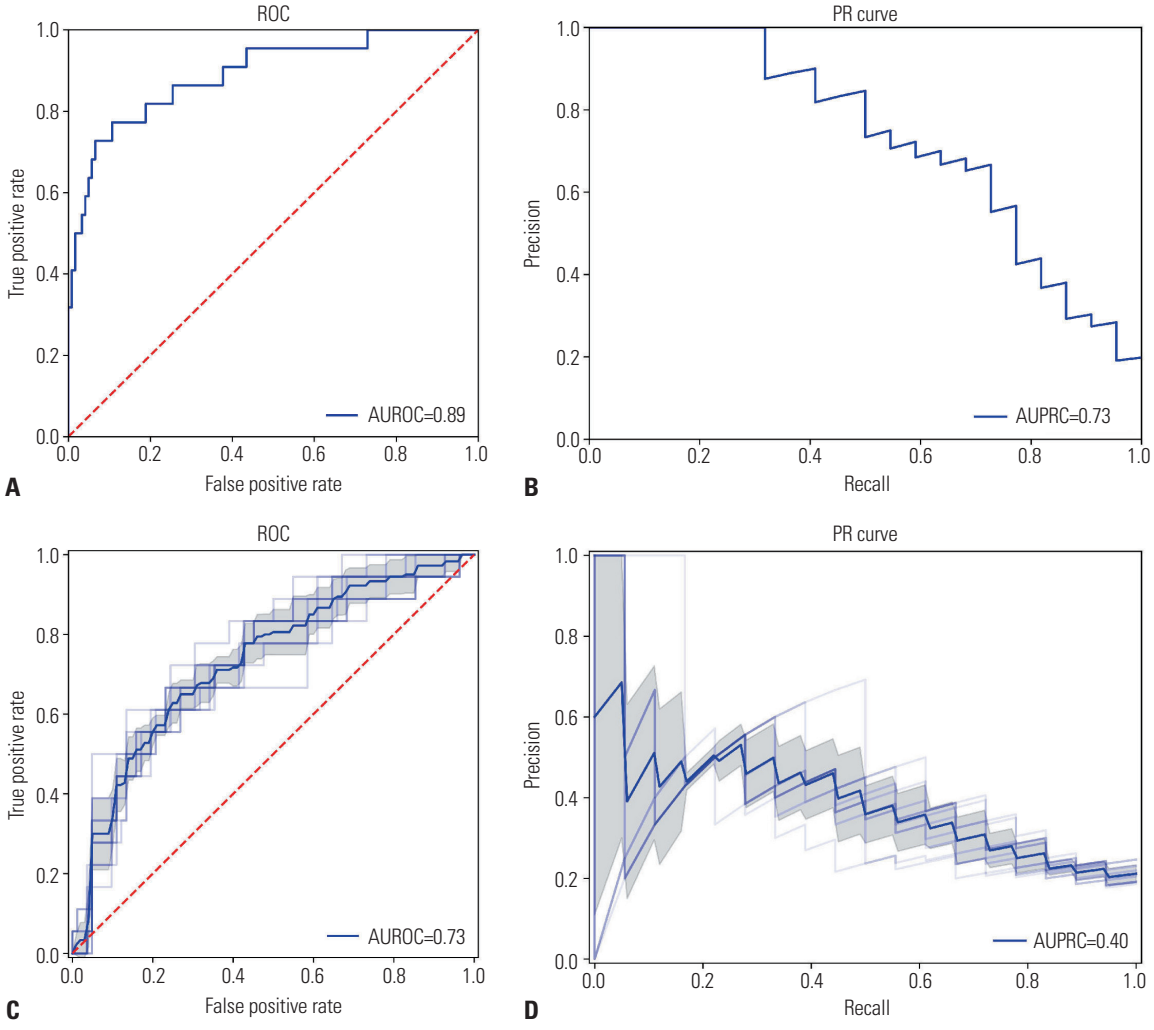

Fig. 5. The receiver operating characteristic (ROC; left) and precision-recall (PR; right) curves of the internal (A and B) and external validation dataset (C and D). The results were calculated through soft-voting models (10-fold each). The value of the area under the curve is shown at the right-bottom side of the graph. AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve.

creased attention, the coefficient value of our model's abnormal probability score was negative.

In particular, the means of the abnormal probability scores from the continuously observed situation (multiple window), odds ratio [IUGR: 3.626 (p=0.031), fetal distress: 23.09 (p<0.001), and Apgar1min (<7): 9.523 (p<0.001)] were larger than the odds ratio in both the randomly observed and last 20-minute situations.

DISCUSSION

In this study, we constructed a classification model based on machine learning methods that could continuously aid obstetricians in assessing pregnancy hazards (IUGR, fetal distress, and Apgar) based on fetal cardiotocography results. A regression analysis revealed that the abnormal probability score that the classification model derived from fetal cardiotocography results had the largest odds ratio value for clinical information in the continuously observed situation. Further, abnormal probability scores were significantly associated with all clinical outputs, with the exception of IUGR in two situations (random sampling and last 20 minutes sampling).

This model can help clinicians assess fetus status by quickly considering the condition of the newborn. In this context, some women with high-risk pregnancies require cardiotocography applications three or more times each day to ensure sufficient observation. This constitutes a major challenge for obstetricians who must read and interpret the fetal cardiotocography results of numerous patients. This not only relieves physical burdens and time constraints for obstetricians, but also enables higher quality care for mothers and their babies.

The multivariable logistic regression confirmed that the risk of fetal distress increased by 10.2 times, as the abnormal probability score increased, in random sampling. Further, the risk of 1- and 5-minute Apgar scores falling below 7 was 3.2 times

and 7.8 times higher, respectively. In the last sampling, the risk of fetal distress increased by 7.6 times, while the risk of the 1- and 5-minute Apgar scores falling below 7 also increased by 3.4 and 7.9 times, respectively.

In the multiple sampling situation, the correlation was much stronger for all items than in the previous two cases. The risk of IUGR was 3.6 times higher, whereas fetal distress was 23 times higher than for a normal fetus. The risk of 1- and 5-minute Apgar scores being lower than 7 was 9.5 times and 11.5 times higher, respectively. The multivariable linear regression revealed negative correlations for the abnormal probability scores and Apgar scores in all situations. The negative correlation with the Apgar score was greatest in the multiple sampling situation, but lowest in the random sampling situation.

Apgar score is a universal and immediate way of assessing newborn well-being, and is affected by a variety of factors, including neonatal physical health and responsiveness to external resuscitations.[24] Apgar score is a valuable method for predicting neonatal adverse outcomes: if the 5-minute Apgar score is below 7, then the neonate should be closely monitored, every five minutes, until the Apgar score is restored to the normal range (7–10).[25] Our algorithm model estimated an abnormal probability score and could warn obstetricians to expect a poor Apgar score. A randomized controlled study by Kamala, et al.[26] indicated that continuous fetal monitoring via Doppler detected abnormal FHR earlier than intermittent Doppler exam. Another systematic study showed that it was difficult to use the intermittent auscultation technique to predict periodic FHR changes, such as deceleration and saltatory patterns, which may be early signs of fetal distress.[27] We also found a negative correlation between abnormal probability score and Apgar score: there were greater correlations in multiple sampling than in random single sampling. These findings are consistent with those of previous studies.

To train the classification machine learning model, we compiled a dataset based on fetal cardiotocography reading data from Ajou University. However, some readings were dominantly collected (fetal cardiotocography read as "reactive"), whereas others were very few in number, including early deceleration, late deceleration, various deceleration, prolonged deceleration, and non-reactive. Due to the relatively low numbers of the latter types, we labeled all readings, except reactive, as abnormal.

Although we trained the fetal cardiotocography classification model using severely imbalanced data, its performance (AUPRC=0.73) was sufficient for extracting the discriminative features of the abnormal waveform data.[28] Because smaller models resist overfitting to the dataset,[29] our models' usage of random kernels for the feature extractor more easily detected abnormal features in the imbalanced dataset.

This study had some limitations in terms of data collection. Although various methods were applied to transform the data into a usable format, it did not include the original Hz at which the fetal cardiotocography signals were collected by the device. Furthermore, when testing the external validation of our model, we noted a decrease in performance, compared with the inter-

Table 2. Baseline Characteristics of the Clinical Validation Cohort (n=707)

Variable	Overall
Age (yr)	34.6±4.8
IUP (weeks)	35.9±3.2
Year	
2017	277 (39.2)
2018	234 (33.1)
2019	196 (27.7)
Emergency	
No	333 (47.1)
Yes	374 (52.9)
Delivery	
C/S	466 (65.9)
NSVD	150 (21.2)
SVD	90 (12.7)
VBAC	1 (0.1)
Infant sex	
Female	361 (51.1)
Male	346 (48.9)
Weight (g)	2642.4±768.6
BPD (cm)	8.7±0.9
Apgar 1 min	7.2±1.6
Apgar 5 min	8.5±1.4
Fetal-growth disorders	
No	552 (78.1)
Yes	155 (21.9)
Fetal distress	
No	632 (89.4)
Yes	75 (10.6)

IUP, intrauterine pregnancy; C/S, cesarean section; NSVD, normal spontaneous vaginal delivery; SVD, spontaneous vaginal delivery; VBAC, vaginal birth after cesarean; BPD, biparietal diameter.
Data are presented as mean±standard deviation or n (%).

Table 3. Results of Multivariable Logistic Regression and Multivariable Linear Regression

Clinical situation	Logistic regression, odds ratio (*p* value)				Linear regression, coefficient (*p* value)	
	IUGR	Reported fetal distress	Apgar 1 min (<7)	Apgar 5 min (<7)	Apgar 1 min	Apgar 5 min
Random	2.114 (0.05)	10.25 (<0.01)	3.239 (<0.01)	7.824 (<0.01)	-0.800 (<0.01)	-0.464 (<0.05)
Last	1.554 (0.25)	7.596 (<0.01)	3.386 (<0.01)	7.907 (<0.01)	-0.989 (<0.01)	-0.646 (<0.01)
Multiple	3.626 (<0.05)	23.09 (<0.01)	9.523 (<0.01)	11.49 (<0.01)	-1.53 (<0.01)	-0.951 (<0.01)

IUGR, intrauterine growth restriction.

https://doi.org/10.3349/ymj.2022.63.7.692

nal validation performance. These differences are related to different medical centers using different cardiotocography, image types, and data collection algorithms (Supplementary Fig. 3, only online). This may affect the convolutional kernel-based feature extraction step, and solving these problems remains a challenge. For this reason, our image pre-processing program cannot currently be applied at other institutes. We plan to update the data collection algorithm used in this study to work at other institutes, then assess model performance based on fetal cardiotocography results from an external hospital.

The model should also be tested in a practical clinical setting, particularly to determine whether it adequately supports obstetricians in reading fetal cardiotocography results, thereby reducing fatigue. This will establish the relevant physical and mental benefits while demonstrating its ability to improve outcomes for patients.

In conclusion, this study developed a machine learning model that can classify fetal conditions based on fetal cardiotocography results. This model can help clinicians to assess pregnancies more quickly in situations that require continuous observation. We expect the model to function as a decision support system for long-term fetal cardiotocography applied in a hospital setting, as well as a sufficient fetal cardiotocography reporting system that pregnant women can use at home.

ACKNOWLEDGEMENTS

This research was supported by R&BD Program through the INNOPOLIS, funded by the Ministry of Science and ICT (2020-IT-RD-0140). This study was also supported by the Korea Medical Device Development Fund grant, which is funded by the Korean government (Ministry of Science and ICT; Ministry of Trade, Industry and Energy; Ministry of Health & Welfare; and Ministry of Food and Drug Safety) (Project Number: 1711138152, KMDF_PR_20200901_0095).

AUTHOR CONTRIBUTIONS

Conceptualization: Miran Kim and Dukyong Yoon. Data curation: Hye Jin Chang, Jung Ah Jung, Seongwoo Kang, and Seokyoung Yoon. Formal analysis: Tae Jun Park and Byung Jin Choi. Funding acquisition: Miran Kim. Investigation: Hye Jin Chang. Methodology: Tae Jun Park and Byung Jin Choi. Project administration: Miran Kim. Resources: Dukyong Yoon. Software: Tae Jun Park and Byung Jin Choi. Supervision: Miran Kim. Validation: Tae Jun Park. Visualization: Tae Jun Park. Writing—original draft: Tae Jun Park and Seokyoung Yoon. Writing—review & editing: Hye Jin Chang, Miran Kim, and Dukyong Yoon. Approval of final manuscript: all authors.

ORCID iDs

Tae Jun Park https://orcid.org/0000-0001-9053-8178
Hye Jin Chang https://orcid.org/0000-0002-1122-1269
Byung Jin Choi https://orcid.org/0000-0002-1445-5888
Jung Ah Jung https://orcid.org/0000-0003-4312-7354
Seongwoo Kang https://orcid.org/0000-0002-1540-0934
Seokyoung Yoon https://orcid.org/0000-0002-4695-2555
Miran Kim https://orcid.org/0000-0001-5553-5334
Dukyong Yoon https://orcid.org/0000-0003-1635-8376

REFERENCES

1. Ayres-de-Campos D. Electronic fetal monitoring or cardiotocography, 50 years later: what's in a name? Am J Obstet Gynecol 2018;218:545-6.
2. Kawakita T, Reddy UM, Landy HJ, Iqbal SN, Huang CC, Grantz KL. Neonatal complications associated with use of fetal scalp electrode: a retrospective study. BJOG 2016;123:1797-803.
3. O'Neill E, Thorp J. Antepartum evaluation of the fetus and fetal well being. Clin Obstet Gynecol 2012;55:722-30.
4. Hoyer D, Żebrowski J, Cysarz D, Gonçalves H, Pytlik A, Amorim-Costa C, et al. Monitoring fetal maturation-objectives, techniques and indices of autonomic function. Physiol Meas 2017;38:R61-88.
5. Gierałtowski J, Hoyer D, Schneider U, Żebrowski JJ. Formation of functional associations across time scales in the fetal autonomic control system--a multifractal analysis. Auton Neurosci 2015;190:33-9.
6. Galazios G, Tripsianis G, Tsikouras P, Koutlaki N, Liberis V. Fetal distress evaluation using and analyzing the variables of antepartum computerized cardiotocography. Arch Gynecol Obstet 2010;281:229-33.
7. Silveira C, Pereira BG, Cecatti JG, Cavalcante SR, Pereira RI. Fetal cardiotocography before and after water aerobics during pregnancy. Reprod Health 2010;7:23.
8. Ayres-de-Campos D, Spong CY, Chandraharan E. FIGO consensus guidelines on intrapartum fetal monitoring: cardiotocography. Int J Gynaecol Obstet 2015;131:13-24.
9. Yanamandra N, Chandraharan E. Saltatory and sinusoidal fetal heart rate (FHR) patterns and significance of FHR 'overshoots'. Curr Women's Health Rev 2013;9:175-82.
10. Andreotti F, Behar J, Zaunseder S, Oster J, Clifford GD. An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms. Physiol Meas 2016;37:627-48.
11. Romagnoli S, Sbrollini A, Burattini L, Marcantoni I, Morettini M, Burattini L. Annotation dataset of the cardiotocographic recordings constituting the "CTU-CHB intra-partum CTG database". Data Brief 2020;31:105690.
12. Sweha A, Hacker TW, Nuovo J. Interpretation of the electronic fetal heart rate during labor. Am Fam Physician 1999;59:2487-500.
13. Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. Commun ACM 1972;15:11-5.
14. Dempster A, Petitjean F, Webb GI. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Min Knowl Discov 2020;34:1454-95.
15. Baloglu UB, Talo M, Yildirim O, Tan RS, Acharya UR. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recognit Lett 2019;122:23-30.
16. Ullah A, Rehman SU, Tu S, Mehmood RM, Fawad, Ehatisham-Ul-Haq M. A hybrid deep CNN model for abnormal arrhythmia detection based on cardiac ECG signal. Sensors (Basel) 2021;21:951.
17. Abdoli S, Cardinal P, Koerich AL. End-to-end environmental sound classification using a 1D convolutional neural network. Expert Syst Appl 2019;136:252-63.
18. Petrozziello A, Jordanov I, Aris Papageorghiou T, Christopher Redman WG, Georgieva A. Deep learning for continuous electronic fetal monitoring in labor. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:5866-9.
19. Zhao Z, Zhang Y, Deng Y. A comprehensive feature analysis of the fetal heart rate signal for the intelligent assessment of fetal state. J

Clin Med 2018;7:223.
20. Zhong W, Liao L, Guo X, Wang G. Fetal electrocardiography extraction with residual convolutional encoder-decoder networks. Australas Phys Eng Sci Med 2019;42:1081-9.
21. Raouf S, Sheikhan F, Hassanpour S, Bani S, Torabi R, Shamsalizadeh N. Diagnostic value of non stress test in latent phase of labor and maternal and fetal outcomes. Glob J Health Sci 2014;7:177-82.
22. Spilka J, Chudáček V, Koucký M, Lhotská L, Huptych M, Janků P, et al. Using nonlinear features for fetal heart rate classification. Biomed Signal Process Control 2012;7:350-7.
23. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg 1953;32:260-7.
24. American Academy of Pediatrics Committee on Fetus And Newborn, American College of Obstetricians and Gynecologists Committee on Obstetric Practice. The Apgar score. Pediatrics 2015;136:819-22.
25. Li F, Wu T, Lei X, Zhang H, Mao M, Zhang J. The Apgar score and infant mortality. PLoS One 2013;8:e69072.
26. Kamala B, Kidanto H, Dalen I, Ngarina M, Abeid M, Perlman J, et al. Effectiveness of a novel continuous doppler (Moyo) versus intermittent doppler in intrapartum detection of abnormal foetal heart rate: a randomised controlled study in Tanzania. Int J Environ Res Public Health 2019;16:315.
27. Blix E, Maude R, Hals E, Kisa S, Karlsen E, Nohr EA, et al. Intermittent auscultation fetal monitoring during labour: a systematic scoping review to identify methods, effects, and accuracy. PLoS One 2019;14:e0219573.
28. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10:e0118432.
29. Myung IJ. The importance of complexity in model selection. J Math Psychol 2000;44:190-204.

https://doi.org/10.3349/ymj.2022.63.7.692