Research on Improving Fetal Health Prediction Model Using Optimal Fetal Feature Selection Technique

SeokHyun Moon*, Jongbin Lee°

ABSTRACT

The phenomena of aging society and declining in birth rate have led to an increased interest in research on predicting fetal health in the field of healthcare. Existing fetal health classification models are actively being studied to improve the models' overall predictive performance. However, predicting fetal health requires a significant number of features and calculations, making rapid and accurate predictions challenging. In this paper, we proposed a method to enhance the accuracy of the fetal health classification models as well as to reduce computational time by selecting the appropriate number of features in the predictive system. Firstly, random resampling method was used to mitigate overfitting caused by data imbalance and normalized the dataset through robust scaling to handle data outliers. Secondly, sequential backward selection of feature algorithm was used to select minimal required features and K-fold cross-validation to ensure the models' accuracy. Using the proposed method, the experimental results achieved 97.2% accuracy, surpassing the results of latest fetal health prediction models studied.

Key Words : CTG, Fetal Health, Machine Learning, Sequential Backward Feature Selection, Robust Scaling Normalization, Imbalanced Datasets

Ⅰ. Introduction

The society has witnessed a rising concern over fetal health due to the increasing rates of congenital anomalies which impact to the persisting issue of low birth rates. Congenital anomalies refer to structural or functional anomalies in fetuses which potentially impose challenges on individuals and families[1,2]. The increase in the incidence of birth defects is associated with various factors, including inadequate fetal health management, genetic factors, and an aging population structure. According to the Korean Statistical Information Service (KOSTAT) for the year 2021, the birth rate has decreased by 10.3% comparing to 2020. This figure, recording a birth rate of 0.81 births per woman, is notably unprecedented since the KOSTAT has started compiling birth rate data in 1970[3]. The

KOSTAT has also reported that the estimated population decline has been advanced from 2032 to 2029[4]. The use of fetal health big data analysis aims to ensure that healthy childbirth and alleviate any concerns, even in the case of elder women, by identifying the presence of potential risk factors throughout the pregnancy period and offering early prevention and treatments.

In this paper, we proposed a methodology to improve the fetal health prediction using sequential backward selection (SBS) algorithm to seek the optimal number of features. By using SBS, the algorithm iteratively removes one feature at a time to identify the most relevant features for fetal health prediction. For better classifying performance, K-fold validation was used after training the dataset with SBS algorithm to make up suboptimal condition to achieve

◆ First Author : Yonsei University Bio & Living Engineering of GLC, seokhyunmoon@yonsei.ac.kr, 정회원
° Corresponding Author : Yonsei University Applied Information Engineering of GLC, belle.lee@yonsei.ac.kr, 정회원
   논문번호 : 202308-040-C-RN.R1, Received August 10, 2023; Revised October 24, 2023; Accepted November 7, 2023

the best possible health prediction outcomes. The proposed methodology has shown improvements in all areas of performance metrics including accuracy, compared to the latest research that selected features based on exploiting correlation coefficient-based feature reduction technique.

II. References

2.1 Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is a variation of the gradient boosting method as well as a tree-based learning algorithm known for its distinctive features: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB)[5]. Specifically, LightGBM grows tree leaf-wise, where trees are split to maximize data loss instead of aiming for a balanced tree structure. As a result, the tree levels become deeper and asymmetrical, as depicted in figure 1, but it minimizes the prediction error loss more effectively than other boosting algorithms[6]. LightGBM offers advantages in constructing fast and powerful prediction models for large-scale datasets. By leveraging the characteristics of gradient boosting algorithms, LightGBM aims to minimize errors and maximize prediction performance.
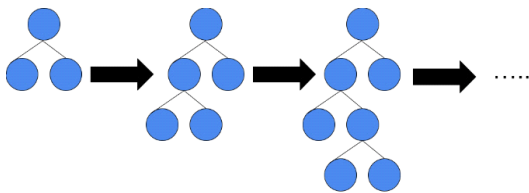

Fig. 1. Leaf-wise tree growth.

2.2 Extra Tree Classifier

The Extra Trees Classifier is one of the ensemble techniques in a machine learning model that utilizes the Extremely Randomized Trees (ExtraTrees) algorithm to perform classification tasks. Extra Trees Classifier exhibits high performance on various datasets by randomly selecting features and performing random splits, enabling faster learning compared to Random Forests[7]. Moreover, it possesses the property of preventing overfitting,

allowing for good performance with relatively fewer hyperparameter tunings.

2.3 Sequential Backward Selection Algorithm

Sequential Backward Selection (SBS) is one of the feature selection algorithms, where the algorithm repeatedly trains the model by removing one feature at a time from the entire feature set after initially training the model with all features[8]. In the SBS algorithm, each feature is removed one by one based on the performance difference in accuracy before and after the removal, aiming to obtain the model with the best performance at that point. However, the overall performance may vary depending on the order of feature removal which could lead to suboptimal result and less reliability. Thus, additional validation process is needed to achieve higher reliability and optimal performance.

2.4 Dataset

In this study, we used childbirth monitoring device data provided by the University of California Irvine (UCI) [9]. The dataset consists of 22 features extracted from Cardiotocograms (CTG) examinations, including fetal heart rate, fetal movement, uterine contractions, abnormal short-term variability, and the average value of short-term variability, as shown in table 1. The sample size in the dataset is 2126 and are classified into three categories: normal, suspicious, and pathological cases of fetuses (1, 2, 3) as shown in table 2. The categorical data in CTG dataset are in numerical values but data normalization is required for better predicting performance due to the ranges in data across features.

Table 1. Features of CTG dataset.

1	'baseline value'
2	'accelerations'
3	'fetal movement'
4	'uterine contractions'
5	'light decelerations'
6	'severe decelerations'
7	'prolongued decelerations'
8	'abnormal short term variability'
9	'mean value of short term variability'

10	'percentage of time with abnormal long term variability'
11	'mean value of long term variability'
12	'histogram width'
13	'histogram min'
14	'histogram max'
15	'histogram number of peaks'
16	'histogram number of zeroes'
17	'histogram mode'
18	'histogram mean'
19	'histogram median'
20	'histogram variance'
21	'histogram tendency'
22	'fetal health'

Table 2. Distribution of fetal health classification classes.

Normal	Suspicious	Pathological	Total
1655	295	176	2126

Ⅲ. Proposed Fetal Health Prediction Model

3.1 Data Imbalance

As shown in figure 2, the number of fetuses classified as normal is 1655, suspicious cases are 295, and pathological cases are 175 which shows a significant data imbalance in the dataset [10]. The data imbalance among different classes can lead to overfitting on the training dataset, resulting in high accuracy with only training sets but poor predictive performance on test sets. Thus, the aim is to fix the data imbalance problem among the datasets presented in table 2 to achieve more balanced and reliable classification results.

In this paper, random resampling technique was applied to improve the imbalanced distribution of datasets[11]. Random resampling is a method that generates a new version of the training dataset with a different class distribution from the existing imbalanced classes to reduce the difference in sample numbers between majority and minority classes. Within random resampling, there are two types of technique to resolve data imbalance: under-sampling and over-sampling. As in figure 3, under-sampling
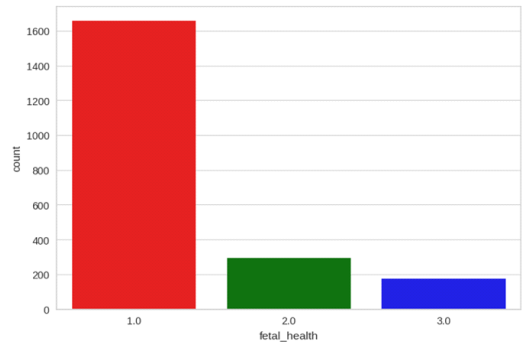

Fig. 2. Distribution of fetal health classification classes
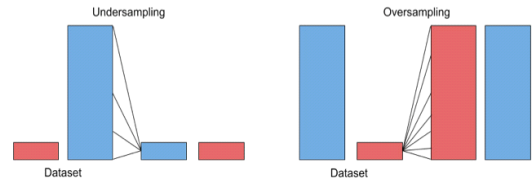

Fig. 3. Under-sampling and over-sampling.

involves removing samples from the majority class at a predefined interval to balance the class distribution and over-sampling increases the number of samples in the minority class by randomly duplicating data from the majority class at a predefined interval or by creating duplicate instances of the minority class data[12].

In this paper, the distribution of the normal class was under-sampled by an amount equal to the sum of suspicious and pathological classes. Specifically, the samples from the normal fetus class were randomly selected and removed, resulting a total of 471 samples in normal fetus class. The suspicious and pathological fetus classes had fewer samples compared to the normal fetus class. Thus, over-sampling was performed to generate samples equal to the number of samples in the normal fetus class. As shown in table 3, all classes' datasets were balanced to have exactly 471 samples each to overcome the issue of overfitting that could arise due to the data imbalance.

Table 3. Fetal health class distribution after random resampling.

Normal	Suspicious	Pathological	Total
471	471	471	1413

3.2 Data Normalization

In this paper, after fixing the data imbalance, 80% of the dataset was divided to training dataset and the rest of 20% to test dataset. However, due to the large ranges of values in the samples across the features, the data was normalized using the robust scaler technique to prevent performance degradation [13]. The robust scaler uses the median and Interquartile Range (IQR) of the data to normalize each data point by subtracting the median from each data point and dividing it by the IQR as shown in equation (1). Additionally, robust scaler is less affected by outlier values compared to other normalization techniques, making it an optimal normalization method for datasets with many outliers. Thus, CTG dataset was scaled with robust scaler to preserve the overall shape and characteristics of the data which accurately reflected and preserved the distribution of the original dataset.

$$X_{scale} = \frac{x - x_{med}}{IQR} \ , \ IQR = Q_1 - Q_3 \qquad (1)$$

where x, x_med, and Qi stands for data point, median of the dataset, i-th quartile respectively.

3.3 Analysis of Sequential Backward Selection with K-fold validation

In this paper, we employed SBS feature selection algorithm with LightGBM model to select the necessary numbers of features for fetal health prediction. The SBS algorithm iteratively removes one feature at a time to obtain the optimal subset of features. The results from training the LightGBM model, as shown in figure 4, indicate that the accuracy of the model increases when the number of features is reduced from approximately 21 to 10. However, at 10 features, the accuracy decreases until 8 features and increases again to a level close to 10 features at 7 features. When reducing the number of features to 5 or below, there is a sharp average decline of about 2.5% in accuracy, leading to the exclusion of 5 or below features from the feature selection process.

Considering accuracy alone, the optimal number of features appear to be 10, 7, and 6. At 10 features, the accuracy is 98.59%, showing a marginal
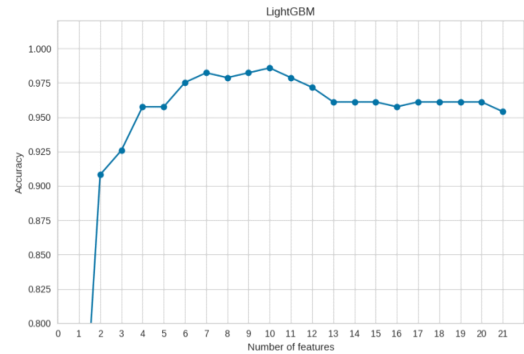

Fig. 4. Prediction result of accuracy vs. number of features

difference of 0.36% compared to 7 features. However, the processing time in the feature with 10 was 13.1801 seconds, which is approximately 2.57 times longer than 5.124 seconds in the 7 features. The model with 7 features is more computationally efficient with only at least accuracy loss . Thus, we can say the optimal number of features to 7 for computational time and concrete accuracy.

In this paper, we used K-fold validation to overcome the issue of SBS algorithm not selecting the overall optimal feature set based on the order of feature removal. K-fold validation divides the data into K subsets, or folds, and iteratively trains and evaluates the model multiple times to average the performance. K-fold validation was employed where the dataset was divided to 10 different folds, enabling the LightGBM model to be trained and evaluated 10

Table 4. Accuracy and Execution Time with Different Number of Features

No of features	Accuracy	Total Execution time (s)
1	0.6572	0.2470
2	0.9081	0.6165
3	0.9258	1.1363
4	0.9576	1.8488
5	0.9576	2.7591
6	0.9753	3.8310
7	0.9823	5.1240
8	0.9788	6.6458
9	0.9823	8.3943
10	0.9859	13.1801

Table 5. Accuracy of K-fold validation with LightGBM

No of folds	Accuracy
Fold 1	0.92
Fold 2	0.96
Fold 3	0.92
Fold 4	0.92
Fold 5	0.92
Fold 6	0.92
Fold 7	0.96
Fold 8	0.97
Fold 9	0.96
Fold 10	0.96
Average	0.94

times. In each iteration, one fold was used as the validation set, while the remaining nine folds were used as the training set to train the model. The accuracy obtained from the 10 iterations was then averaged to calculate the final performance, as shown in table 5.

IV. Experimental Results

4.1 Proposed Framework of the Model

In this paper, the proposed fetal health prediction model's framework is demonstrated as depicted in figure 5. From the fetal health dataset, the data were pre-processed through random sampling to fix data imbalance and normalized by robust scaling method. Then 80% of the data was divided to training and the rest of 20% to test set. From the training set, sequential backward selection feature algorithm was used to select the important number of features to ensure best accuracy for machine learning models. The K-Fold cross validation was used for training and testing the classifier performance. Overall, the machine learning model was tested through test set and evaluated the model's performance through confusion matrix.

4.2 Confusion Matrix and Model Evaluation

The confusion matrix presents the comparison between the model's predicted results and the actual class labels, categorizing them into True Positive
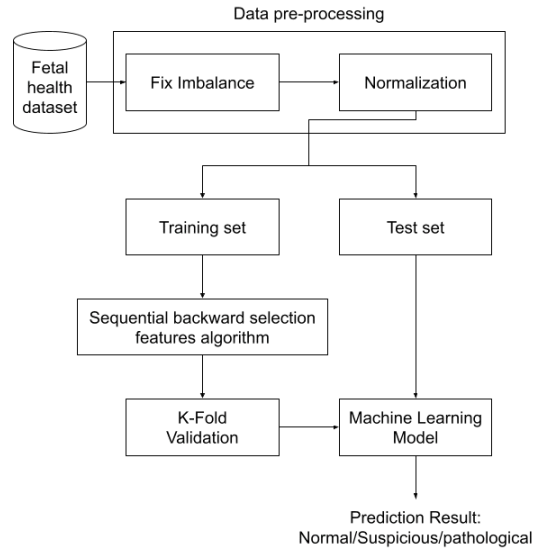

Fig. 5 Proposed framework of the fetal health model

(TP), True Negative (TN), False Positive (FP), and False Negative (FN), as shown in figure 6. It is observed that the number of correctly classified samples for the normal, suspicious, and pathological classes is 89, 92, and 94, respectively, while a total of 8 samples were misclassified.

Using this confusion matrix, the model's performance can be evaluated and compared using metrics such as accuracy, precision, recall, and F1 Score. Among these classification evaluation metrics, accuracy is calculated by dividing the number of correctly classified samples by the total number of samples, as shown in equation (2). Recall, on the other hand, measures the proportion of correctly predicted positive observations among all positive observations in the actual class, as calculated in equation (3). Precision is the ratio of correctly predicted positive
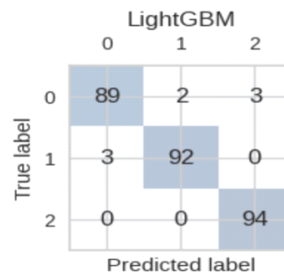

Fig. 6. Confusion matrix of LightGBM.

observations to the total predicted positive observations, as shown in equation (4). Lastly, the F1 Score is the weighted average of precision and recall, calculated according to equation (5).

By utilizing these classification evaluation metrics, the model's performance can be evaluated and analyzed, providing valuable insights into its classification performance for each model.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1 = 2 \ \times \frac{Recall \times Precision}{Recall+Precision} \quad (5)$$

4.3 Experimental Results

This paper compared and analyzed the results of fetal health prediction models using LightGBM with SBS proposed and without. At comparison evaluation in table 6, four classification metrics such as accuracy, precision, recall, and F1 Score, were used to evaluate the effectiveness of these models.

Based on the results in table 6, Proposed model can achieve better performance with an accuracy of 97.2%, recall of 0.972%, precision of 0.972% and F1 of 97.2% comparing to conventional model in predicting fetal health.

Table 6. Comparison of evaluation metrics with models.

Model	Accuracy	Recall	Precision	F1
LightGBM with SBS	0.972	0.972	0.972	0.972
LightGBM without SBS	0.958	0.958	0.959	0.958

Table 7. Comparison of evaluation metrics

Model	Accuracy	Recall	Precision	F1
LightGBM with SBS	0.972	0.972	0.972	0.972
Extre Tree	0.967	0.967	0.970	0.968

Furthermore, when comparing the results of this study with the research on the extra tree model [14], all four classification evaluation metrics have been improved as shown in table 7.

Ⅴ. Conclusions

In the paper, proposed framework of the fetal health prediction can improve computational time and classification accuracy by selecting optimal number of features in SBS algorithm with K-fold validation. In the proposal, the data obtained from Cardiotocograms (CTG) examinations was used. At first stage as data pre-processing, we applied random resampling to resolve the data imbalance problem in three classes. Specifically, Under and over-sampling for the normal fetus class and for the suspicious and pathological fetus classes can prevent overfitting and data loss problem caused by imbalanced data distribution. Secondly, the data was normalized using the robust scaler to ensure accurate results and to preserve the overall structure and characteristics of the data while performing scaling. Lastly, SBS algorithm and K-folder cross validation were proposed to enhance the accuracy and reliability by selecting the optimal number of features through removing one by one procedure of SBS method. It has been simulated with five different ML models, and the results indicated that the LightGBM model showed the best performance according to the confusion metrics and model evaluation.

This proposed research holds the potential to detect early signs of fetal health risks during pregnancy and provide timely information for necessary treatments, thereby reducing the occurrence of congenital abnormalities and fetal mortality rates.

References

[1] D. C. Kim, "Effect of the child benefit on the increase of the total fertility rate," *The J. Korean Public Policy*, vol. 20, no. 2, pp. 3-25, 2018.

[2] J. H. Oh, "A study on prediction for reflecting variation of fertility rate by province under

ultra-low fertility in Korea," *The Korean J. Applied Statistics*, vol. 34, no. 1, pp. 75-98, 2021.

[3] KOSTAT, "*2022 Birth Statistics*," 2023.

[4] J. Reefhuis and M. A. Honein, "Maternal age and non-chromosomal birth defects, Atlanta-1968-2000: Teenager or thirty-something, who is at risk?" *Birth Defects Research. A Clinical and Molecular Teratology*, vol. 70, no. 9, pp. 572-579, 2004.

[5] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," *Advances in NIPS*, vol. 30, pp. 3146-3154, 2017.

[6] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms," *Mathematics*, vol. 8, no. 5, pp. 765, 2020.

[7] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, pp. 3-42, 2006.

[8] J. Xie and W. Xie, "Several feature selection algorithms based on the discernibility of a feature subset and support vector machines," *Chin. J. Comput.*, vol. 37, no. 8, pp. 1704-1718, 2014.

[9] D. Campos and J. Bernardes, "*Cardiotocography*," UCI Machine Learning Repository, 2010. (https://doi.org/10.24432/C51S4N.)

[10] J. M. Johnson and T. M. Khoshgoftaar, "The effects of data sampling with deep learning and highly imbalanced big data," *Inf. Syst. Frontiers*, vol. 22, no. 5, pp. 1113-1131, 2020.

[11] P. Branco, L. Torgo, and R. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1-50, 2017.

[12] H. He and Y. Ma, "*Imbalanced Learning: Foundations, algorithms, and applications*," Wiley-IEEE Press, 2013.

[13] D. Levy, Y. Carmon, J. C. Duchi, and A. Sidford, "Large-Scale methods for distributionally robust optimization," *Advances in NIPS*, vol. 33, pp. 8847-8860, 2020.

[14] C. K. Kim, J. H. Kim, and G. W. Kim, "Fetal health prediction model exploiting correlation coefficient based feature reduction technique," *Korea Software Congress 2022*, vol. 2022, no. 12, pp. 1489-1491, Jeju Island, Korea, Dec. 2022.

SeokHyun Moon

Sep. 2019~Current : B.S. degree, Yonsei University

<Research Interests> Machine Learning for Healthcare, Disease Diagnosis, Deep Learning,

[ORCID:0009-0001-7546-4374]

Jongbin Lee

Dec. 2005~May 2019 : Director, Intel Co.

Mar. 2020~Jul. 2023 : Professor, Sangmyung University

Mar. 2023~Current : Professor, Yonsei University & CEO, FTI Co.

<Research Interests> ICT technology, wireless communications, Data Analysis, Machine Learning

[ORCID:0000-0002-6144-2406]