A Comparative Study via Explainable AI Methods of Preterm Birth Prediction Model Trained on Ultrasound Images with Slight Changes

Yeong-Eun Jeon[1*], Ga-Hyun Son[2*], Ho-Jung Kim[1], Jae-Jun Lee[3] and Dong-Ok Won[1,4]

***Abstract*— Preterm birth (PTB), the leading cause of infant death, is mainly predicted by cervical length (CL). But using CL alone, even medical experts have difficulty predicting PTB. Therefore, we aimed to identify new predictors of PTB except CL using eXplainable Artificial Intelligence (XAI). Moreover, we conducted a comparative study to investigate the impact of the model's learning pattern and features according to the differences in expert guidance and XAI methods on the same ultrasound images. As a result, we figured out that these influences may be a hurdle to finding new prediction factors for preterm birth.**

I. INTRODUCTION

Each year, 15 million infants are born prematurely, and in 2015, nearly 1 million infants died from complications of preterm birth (PTB)[1]. PTB is delivered before 37 weeks of pregnancy, and some studies are being conducted to prevent it[2], [3]. To predict PTB on ultrasound images, cervical length (CL) is used as a representative measure. The CL is the length from the internal to the external os of the uterus, and a shortened CL is associated with an increased risk of PTB[4]. However, PTB is not caused only by a short CL, clinicians have struggled to predict it. Therefore, we aimed to discover new factors that can predict PTB other than CL on ultrasound images based on AI technology in this study. To find new PTB predictors, we trained a deep neural network model using ultrasound images and used eXplainable Artificial Intelligence (XAI) to interpret the model's feature map.

II. METHOD

A. Deep neural network model for Classification

ResNet[5] and DenseNet[6] were used as classification models. They provided the benefit of preventing image features during network training because they had connectivity between layers. We also selected EfficientNet[7], which has the squeeze-and-excitation (SE) block. It extracts only the important information from each channel and figures out how significant each channel is in relation to the others so that the model can choose the most crucial features.

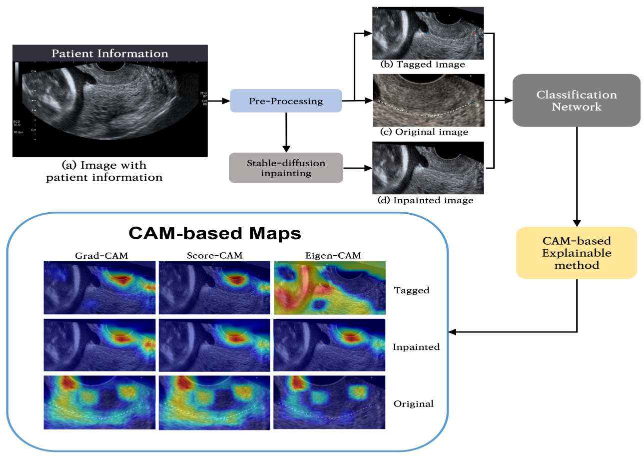

Fig. 1. The proposed framework with two parts: 1) feature extraction with a deep neural network, 2) the interpretation part using an XAI method.

B. Class Activation Map(CAM)-based methods

We selected explainable approaches based on Class Activation MAP (CAM), which are good at localization on a heatmap. The first method is Gradient Class Activation Map (GradCAM)[8]. It uses the idea of a gradient to figure out the weight of model. Secondly, ScoreCAM[9] creates a heatmap, with the importance determined by the difference between the baseline and the feature of each channel. The last method is EigenCAM[10], which uses the first eigen vector obtained through learned weight matrix. It can generate a heatmap without changing the network structure or calculating a noisy gradient.

C. Dataset and Data-Preprocessing

1196 ultrasound images (214 preterm and 982 full-term) were collected at Hallym University Kangnam Sacred Heart Hospital (IRB No. 2018-01-026). To consider the influence of medical expert guidance on ultrasound images, we applied image processing to data. Accordingly, the first ultrasound image indicated the cervix's location with red and blue points (Tagged), the second showed a marking line without points (Original), and the third image used generative AI to remove all indications (Inpainted). The training and evaluation dataset was divided at a ratio of 8:2 each. Also, we conducted data augmentation to solve the issue of data imbalance.

*These authors contributed equally to this work.

[1]Yeong-Eun Jeon, Ho-Jung Kim and Dong-Ok Won are with the Department of Artificial Intelligence Convergence, Hallym University, Chuncheon, Republic of Korea {M22056, hojungkim, dongok.won}@hallym.ac.kr

[2]Ga-Hyun Son is with the Department of Obstetrics and Gynecology, Hallym University Kangnam Sacred Heart Hospital, Kangnam, Republic of Korea gahyuns@hallym.or.kr

[3]Jae-Jun Lee is with the Department of Anesthesiology and Pain Medicine, Hallym University Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea iloveu59@hallym.or.kr

[4]Dong-Ok Won is with the College of Medicine, Hallym University, Chuncheon, Republic of Korea

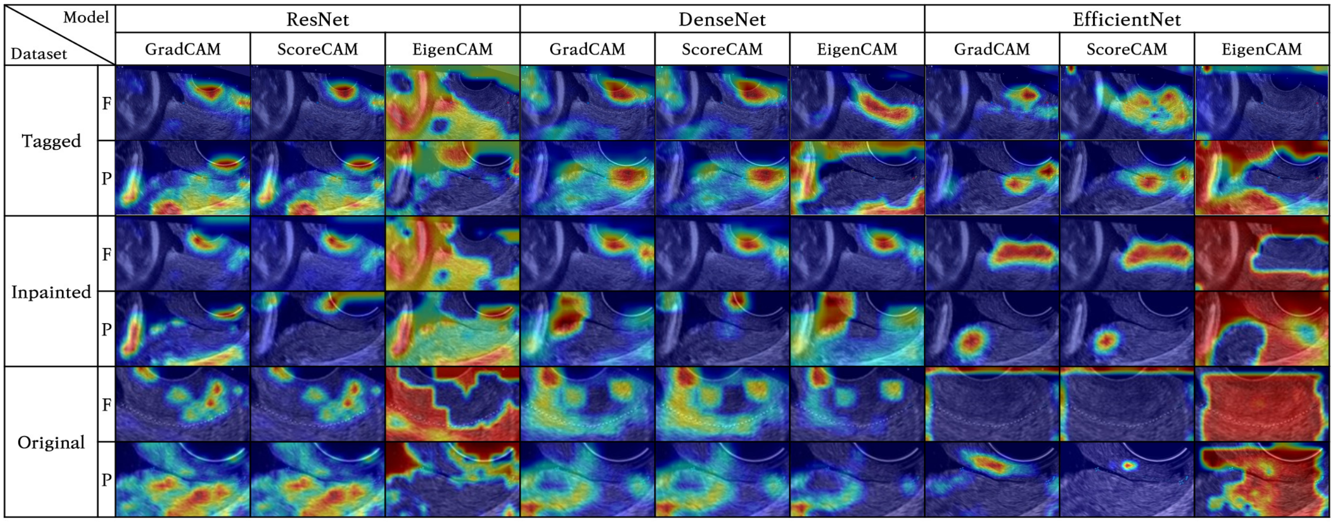

Fig. 2. This shows result of heatmap for Full-Term(F) and Preterm(P) classes on the same ultrasound image according to XAI methods for each classification model: ResNet(Right), DenseNet(Middle), EfficientNet(Left).

III. RESULT AND DISCUSSION

Fig. 2. shows heatmaps for Full-term and Preterm classes on the same ultrasound images according to the XAI method for each classification model. For most of the data, the heatmap results of GradCAM and ScoreCAM showed similar. However, compared to other methods, the heatmap of EigenCAM was different. In particular, the EigenCAM result of the EfficientNet tended to differ from the GradCAM and ScoreCAM. Also, in full-term, the upper part was highlighted more than the position of the cervical length (CL). In the preterm, in the tagged image, it was highlighted right above the CL or at the external os, and it was highlighted on the left part of the image in the inpainted dataset. In the original, it was found that the area was clearly different for each model. These showed that the learning pattern and interpretation of the model were changed according to the guidance given by the clinician during the learning stage. This indicated that the information included in the image may vary depending on the clinician's hand-designed marker so the model results might also appear differently. This might make it more difficult to discover new preterm predictors.

IV. CONCLUSIONS

In this paper, we compared the results of the trained model using eXplainable Artificial Intelligence (XAI) for discovering new preterm predictors. These showed that data with clinician guidance can impact the model's ability to train and the model cannot find the new features. Therefore, when selecting data to be used for training, it is crucial to examine image aspects that can be cheated by the model in medical field. Also, we need to choose an explainable method that can show reasonable analysis results on medical data.

ACKNOWLEDGMENT

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT (MSIT), the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (No. 1711139109, KMDF_PR_20210527_0005), partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A5A8019303), and partly supported by the Bio&Medical Technology Development Program of the NRF funded by the Korean government (MSIT) (No. RS-2023-00223501).

REFERENCES

[1] J. Perin et al, Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals, The Lancet Child & Adolescent Health, vol. 6, no. 2, 2022, pp. 106–115.
[2] G. -H. Son, J. -J. Lee, Y. -M. Kim and K. -Y. Lee, The role of antimicrobial peptides in preterm birth, Int. journal of molecular sciences, vol. 22, no. 16, 2021, pp. 8905.
[3] Y. Takayama et al, Soft wearable tactile sensor for softness evaluation through internal examination, in 2019 IEEE Int. Conf. Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 3210-3215.
[4] J. Owen et al, Mid-trimester endovaginal sonography in women at high risk for spontaneous preterm birth, Jama, vol. 286, no.11, pp. 1340–1348
[5] K. He et al, Deep residual learning for image recognition, in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp.770-778.
[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, Densely Connected Convolutional Networks, arXiv e-prints, Aug. 2016.
[7] M. Tan and Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in Int. Conf. Machine Learning, 2019, pp. 6105-6114.
[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in Proc. IEEE Int. Conf. Computer Vision, 2017, pp. 618-626.
[9] H. Wang et al, Scorecam: Score-weighted visual explanations for convolutional neural networks, in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition workshops, 2020, pp. 24-25.
[10] M. B. Muhammad and M. Yeasin, Eigen-cam: Class activation map using principal components, in 2020 Int. Joint Conf. Neural Networks (IJCNN), IEEE, 2020, pp. 1-7.